How Social Contingency Shapes Early Child-Caregiver Interactions

Christopher Martin Mikkelsen Cox

Doctor of Philosophy

Department of Linguistics & Cognitive Science, Aarhus University
Department of Language & Linguistic Science, University of York

December 2023

Abstract

Child-caregiver interactions involve mutual anticipation and contingent responses where each interactant adapts their behaviour in response to social feedback from the other interactant. This dissertation focuses on two features that originate in these social feedback loops and are highly characteristic of early child-caregiver interactions: ***speech content*** (i.e., infant-directed speech, IDS) and ***temporal structure*** (i.e., turn-taking). To map out the sampling space of previous studies and to gain a better understanding of how social contingency shapes the form and structure of early child-caregiver interactions, I conduct systematic reviews, meta-analyses, and critically leverage cumulative science practices. ***Study I*** is a systematic review and Bayesian meta-analysis of the acoustic features of IDS. The study aims to investigate how the acoustic features of IDS differ across infant ages and languages and to understand these results in relation to the purported functions of IDS. ***Study II*** goes hand in hand with the meta-analysis on IDS and provides a comprehensive acoustic analysis of the prosodic and vocalic properties of Danish caregivers' spontaneous IDS and adult-directed speech (ADS). ***Study III*** is a systematic review and Bayesian meta-analysis on the developmental trajectory of infants' turn-taking abilities, which attempts to identify key moderators affecting response latencies to gain a better understanding of the cognitive mechanisms underlying children's turn-taking abilities. These three studies not only contribute to our understanding of how social contingency shapes both the speech content and temporal structure of early child-caregiver interactions, but also provide suggestions for future strands of research and theory development.

Author's Declaration

I declare that this thesis is a presentation of original work and I am the first author or joint first author in every article included. My co-authors to the papers are the following: Riccardo Fusaroli, Tamar Keren-Portnoy, Christina Bergmann, Andreas Roepstorff, Greg Bryant, Christina Dideriksen, Morten H. Christiansen, Emma Fowler, Vivian Nguyen, Otto Versyp (for specific contributions, please read the Research Degree Thesis Statement of Authorship document at the beginning of each article). This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references. All three studies in this dissertation have been published as peer-reviewed articles.

Cox, C., 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.

Cox, C., 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.

Nguyen*, V., Versyp*, O., Cox*, C., & Fusaroli, R. (2022). A systematic review and Bayesian meta-analysis of the development of turn taking in adult–child vocal interactions. *Child Development, 93*(4), 1181-1200.

* *Joint first-authorship*

Acknowledgements

This section is proudly sponsored by a brain running on strong coffee and a healthy smidgen of sheer madness here at the finish line of a long journey. Brace yourself. If nothing else, I hope that these acknowledgements have entertainment value.

One of the amazing things about doing work in developmental science is that you are constantly reminded that curiosity, exploration and play are essential components to building knowledge and learning new skills. Doing science also needs a heavy dose of this form of reckless curiosity, and throughout this process I have tried to let play and the joy of discovery shape my scientific path. Science also involves collaboration, and I have many people to thank for playing a part in this wonderful adventure.

First and foremost, a standing ovation for my main supervisor, Riccardo Fusaroli, whose attention to detail and helpfulness never cease to amaze me. Thanks for being a great supervisor, an inspirational role model, for scaffolding my development and exchanging beautiful ideas with me. I am looking forward to the coming years of learning, curiosity, and discovery.

There would not even be a beginning to this journey without a serendipitous conversation between Tamar Keren-Portnoy and Andreas Roepstorff in the middle of a desert back in 2018. This conversation ended up catapulting me back into the academic world and down this exciting path. You both deserve a huge thanks for trusting me and for all of your support.

The Interacting Minds Centre in Aarhus provided such an amazing, open academic environment for my PhD journey, where no idea was off the table and wackiness was actively encouraged.

Curiosity and play in research is perfectly embodied at this research centre, and I do not underestimate the importance the IMC has had on my thinking.

Similarly, a big thank you to our reading group on phonological development, composed, amongst others, of Marilyn Vihman, Lorraine McCune, Margherita Belia, Catherine Laing, Ali Langner, Florence Oxley, Niveen Omar and Khalid Hudhayri. These meetings have provided a wonderful space for discussions and thoughts that have had a lot of downstream effects on how I both read, review and write papers. It has also been amazing to follow along with and support fellow PhD students going through a similar process.

I have also had the privilege to travel during this process and would like to thank each and every one of the people I met along the way on these travels, including Mike Frank, Alex Cristia, Keith Johnson, Martin Zettersten, and Jessica Kosie, along with countless other inspiring researchers that make me believe that what we do has value in the real world.

I would also like to extend a big thanks to the ManyBabies Consortium. To establish such a visionary community that can reach beyond the boundaries of different countries and cultures and unite researchers who are interested in a common goal is no small feat. A special thanks goes to Martin Zettersten for always being so generous with his time and careful work – and for being a fun fellow explorer of tough questions, uncertainty and good explanations.

I would also love to thank my closest family, Mike, Helle and Peter, without whom nothing would make sense. It is hard to understand how lucky we have been as a family, but I am sure that a large part of this luck stems from the unwavering support from our family.

I would also like to thank my partner-in-crime and now fiancé, Johanne, for being the most inspiring, lovable character in my life story. Our meeting marked the beginning of this PhD journey, and this thesis is dedicated to you and to the unwritten pages that await us.

Table of Contents

2.1 Study I: A systematic review and Bayesian meta-analysis of the acoustic features of infant-directed speech

2.2 Study II: Infant-directed speech does not always involve exaggerated vowel distinctions: Evidence from Danish

2.3 Study III: A systematic review and Bayesian meta-analysis of the development of turn taking in adult– child vocal interactions

1 Introduction

Why is ‘peek-a-boo’ such an attractive game for young children to play with their caregivers? The structure of the game is simple. Caregivers briefly conceal their presence behind their hands and trigger a cascade of emotions and responses in the child when they reappear. One explanation for its charm may be that this engaging ‘conversation’ involves mutual anticipation and contingent responses where the behaviour of each participant is shaped by the reciprocal behaviour of their interactant. This form of social feedback loop requires active participation of both parties and illustrates how participants co-produce moments of socially contingent adaptation in their daily interactions and behaviour (Donnelly & Kidd, 2021; Nikolaus & Fourtassi, 2023; Provenzi et al., 2018; Tamis-LeMonda et al., 2001; Warlaumont et al., 2014).

The foundations for these rudimentary interactions start in the third trimester of pregnancy as foetal hearing develops to perceive vocal signals transmitted through the intrauterine fluid of the child-bearer (DeCasper & Fifer, 1980; Lee & Kisilevsky, 2014). At birth, infants prefer and respond most consistently to communicative signals heard in the womb (DeCasper & Fifer, 1980) and thereafter gradually start to engage in social interactions through non-linguistic means, such as eye gaze, gesture and cry signals (Parsons et al., 2012; Snow, 1977b, 1977a, 2013; Vermillet et al., 2022). At the same time, caregivers adapt multiple dimensions of their behaviour to suit the developmental needs of their immature interlocutor (Chong et al., 2003; Kosie & Lew-Williams, 2023): ***from action*** (Karmazyn-Raz & Smith, 2023; Suanda et al., 2018) ***and gesture*** (Dimitrova & Moro, 2013; Iverson et al., 1999) ***to touch*** (Lew-Williams et al., 2019; Seidl et al., 2015)***, emotion*** (Moses et al., 2001; Pintar Breen et al., 2018) ***and speech*** (Broesch & Bryant, 2018; Hilton et al., 2022).

Caregivers' socially contingent coordination of their behaviours across so many different dimensions lies at the heart of child-caregiver interactions and is a basic property of communicative systems (Dideriksen et al., 2023; Fusaroli & Tylén, 2012; Lorge & Katsos, 2019; Pickering & Garrod, 2004, 2021; Rasenberg et al., 2020, 2022; Trujillo & Holler, 2023). Taking juvenile addressees into account and adapting behaviour accordingly happens not only in humans, but also in non-human animals; zebra finches, for example, introduce songs directed to their juveniles with an introductory motif (Chen et al., 2016), bats use a specific peak frequency and timbre in their pup-directed vocalisations (Fernandez & Knörnschild, 2020), and long-tailed macaque monkeys perform infant-directed tool-use actions with more frequent pauses, regular repetitions and longer durations (Masataka et al., 2009). Such magnification of signals may in turn facilitate infants' learning of specific skills and direct infants' attention to meaningful structures (Brand et al., 2002; Ferrari, 2009; Hirai et al., 2022).

The social contingency of communicative signals – that is, the tendency to accommodate the knowledge, needs and abilities of the interlocutor in the interactive process – shapes both the speech content and temporal structure of early child-caregiver interactions. A full picture of early child-caregiver interactions involves consideration of their wide range of dimensions, from emotion and gesture to action and touch (cf., Schatz et al., 2022; Kosie & Lew-Williams, 2023; Yu & Smith, 2012). To tackle foundational issues in the field and to provide a comparison point for generalisation of patterns to other dimensions, this dissertation focuses on two features that are highly characteristic of early child-caregiver interactions:

I. ***Acoustics of Infant-Directed Speech (IDS)*** *– the set of acoustic modifications that caregivers make to their own speech when communicating with infants and young children.*

II. ***Infant Turn-taking Skills*** *– the structured exchange of conversational or social interactions where participants alternate speaking roles through coordinated timing responses.*

Both the speech content and temporal structure of early child-caregiver interactions originate in the mutual feedback loops between interactants (Elmlinger et al., 2019, 2023; Goldstein & Schwade, 2008; Laing & Bergelson, 2020; Ritwika et al., 2020; Warlaumont et al., 2014). This bidirectional process of behavioural adaptation exhibits asymmetry in that caregivers coordinate and adjust their contingent behaviours to their children more than children adjust their behaviour to caregivers (Beebe et al., 2016); however, the feedback loops still require active participation from both interactants and the developing skills, behaviour and knowledge of the child form a crucial part of the socially contingent accommodations that take place (Donnelly & Kidd, 2021; Nikolaus & Fourtassi, 2023; Provenzi et al., 2018; Tamis-LeMonda et al., 2001; Warlaumont et al., 2014). By focusing on a subset of the dimensions of early child-caregiver interactions in this way, I hope to reveal patterns and structures that can be applied to a wider range of dimensions; for example, exaggeration in the dimension of speech may also apply to that of gesture or touch (Masataka et al., 2009), and the back-and-forth temporal adjustments in turn-taking may also generalise to the domains of action or gesture (Masataka, 1992, 1995; Novack et al., 2021; Reilly & Bellugi, 1996). The focus on a subset of the features of child-caregiver interactions allows us to interact with questions of the following type:

I. To what extent do caregivers speaking different languages exhibit variability in the acoustic modifications they make in their IDS? Do caregivers change the acoustics of their IDS as the infant grows older?

II. How can we characterise the timing structures of early child-caregiver vocal interactions and their developmental trajectories during ontogeny? To what extent can we gain information about the mechanisms underlying the timing of turn-taking?

To gain a better understanding of how to approach these questions and to map out the sampling space of previous studies (Kidd et al., 2023; Kidd & Garcia, 2022b, 2022c), I critically leverage cumulative science practices by conducting systematic reviews and meta-analyses. This cumulative approach to science allows us to generalise beyond the limitations of a single study and evaluate regularities across multiple populations, conditions and contexts while recognising the limits of such generalisations (Baumgartner et al., 2023; Brand et al., 2019; Cristia et al., 2022; Fusaroli et al., 2022; Koile & Cristia, 2021; Lakens et al., 2016; Tsuji et al., 2014, 2017; Zettersten, Cox, et al., 2023; Schreiner et al., 2022). I argue and demonstrate that construing scientific progress as a cumulative and iterative process of data aggregation and theory revision allows us to identify new directions to explore in the research landscape (cf., Fusaroli et al., 2022).

This dissertation is composed of three individual studies completed and published during my doctoral studies (Cox, Bergmann, et al., 2023; Cox et al., 2022; Nguyen et al., 2022). Where relevant, the dissertation will incorporate references to ongoing work pursued in extension of topics explored in this strand of research (Cox, Templeton, & Fusaroli, 2023; Fusaroli et al., 2023; Zettersten, Cox, et al., 2023). The three journal articles in their peer-reviewed form are included in the present dissertation in **Section 2** – the only difference to the published versions of the articles is the idiosyncratic formatting and referencing preferences of each of the individual journals. All data and code are available online on the Open Science Forum or GitHub, the links for which can be found at the start of each of the papers in **Section 2**).

The structure of the remainder of this dissertation is as follows. **Section 1.1** provides an overview of the acoustics of IDS (**Section 1.1.1**), how its use varies across cultures (**Section 1.1.2**) and discusses its biophysical origins and cross-cultural adaptations (**Section 1.1.3**). **Section 1.2** provides an overview of turn-taking and the development of children's abilities in this domain (**Section 1.2.1**) and outlines the cognitive mechanisms that have been proposed to underlie children's early developing turn-taking abilities (**Section 1.2.2**). **Section 1.3** discusses fundamental limitations in the current ecosystem of science (**Section 1.3.1**) and outlines how cumulative science practices can help us shed new insights on the generalisability and boundary conditions of scientific topics of study (**Section 1.3.2**). **Section 1.4** provides a short overview of the aims of each of the studies and contextualises what they can tell us about early child-caregiver interactions.

Section 2 contains the three journal articles: a meta-analysis of the acoustic features of IDS (**Section 2.1**), an acoustic analysis of Danish IDS and ADS (**Section 2.2**), and a Bayesian meta-analysis of turn-taking development in child-caregiver vocal interactions (**Section 2.3**).

Section 3 contextualises the findings from the studies. **Section 3.1** outlines a framework that considers behavioural adaptations in child-caregiver interactions as emergent products of contributions from ontogeny, cross-cultural norms and biophysical constraints. **Section 3.2** argues for why this work adds to our knowledge about behavioural adaptations in early ontogeny (**Section 3.2.1**), across cultures (**Section 3.2.2**) and in biology (**Section 3.2.3**). **Section 3.3** uses these insights to point to future directions and argues for a greater focus on turn-by-turn adaptations (**Section 3.3.1**), incorporation of individual differences (**Section 3.3.2**), theory-driven cross-cultural studies (**Section 3.3.3**), methodological cross-pollination from non-human investigations (**Section 3.3.4**), as well as cumulative science and theory development (**Section 3.3.5**).

1.1 IDS Acoustics and Use Across Languages and Cultures

In a typical 'peek-a-boo' scenario, we might expect the caregiver to modify the acoustics of their speech to the child. Scientific studies of IDS have noted marked differences from ADS across many different languages for multiple linguistic variables (Ferguson, 1964; Golinkoff et al., 2015; Saint-Georges et al., 2013; Snow, 1977a). In this section, I provide an overview of the acoustic features and causal pathways of IDS (**Section 1.1.1**) as well as discuss the diversity in infants' linguistic input across cultures (**Section 1.1.2**), with a view to shedding light on how the biophysical origins of IDS and cultural variation shape its features during child-caregiver interactions (**Section 1.1.3**).

1.1.1 The Features & Causal Pathways of IDS

The investigation of cross-linguistic differences in the acoustics of IDS is of theoretical relevance and intimately connected with our understanding of the purpose of this particular speech style (Pye, 1986; Ratner & Pye, 1984; Ingram, 1995). The commonalities that have emerged across a wide variety of languages and cultures can be usefully divided into two distinct groups. Compared to ADS, IDS exhibits i) ***distinct prosodic features*** in the form of a higher fundamental frequency (f_o), a greater degree of f_o variability, a slower speech rate, and phrasal lengthening (Cristia & Seidl, 2014; Hilton et al., 2022; Martin et al., 2015), and ii) ***unique segmental properties***, in the form of longer and acoustically exaggerated discrete segments, such as consonants and vowels (Burnham et al., 2002; Cristia, 2010; Cristia & Seidl, 2014; Kalashnikova & Burnham, 2018; Kuhl et al., 1997; Lovcevic et al., 2020; Miyazawa et al., 2017). The tendency for caregivers to exaggerate the pitch of their IDS and slow down their speech rate finds a prosodic equivalent in sign language where caregivers' infant-directed gestures are slower and larger in scale (Masataka, 1992, 1995; Novack et al., 2021; Reilly & Bellugi, 1996). There is some cross-linguistic variability in the prosodic features of IDS, but this mainly revolves around the extent of prosodic modulation; for example, caregivers speaking British English IDS exhibit a relatively lower pitch than those speaking American English IDS (Fernald et al., 1989; Floccia et al., 2016; Shute & Wheldall, 1999), and caregivers speaking Korean or Tagalog produce greater prosodic differences between IDS and ADS compared to those speaking Tamil (Narayan & McDermott, 2016). Very few studies show no modulations in pitch or articulation rate, and if they do so, these studies stand in contrast with other studies on the same languages (Bohn, 2013; Dideriksen & Fusaroli, 2018; Narayan & McDermott, 2016).

Some of the prosodic properties of IDS exhibit dynamic changes according to infant age, which may reflect the social contingency of caregiver vocalisations; for example, many longitudinal studies on f_0 show that caregivers reduce the differences in vocal pitch between IDS and ADS during early development (Amano et al., 2006; Gergely et al., 2017; Niwano & Sugai, 2002; Stern et al., 1983), and a number of studies of articulation rate show that caregivers increase their articulation rate as the child grows older during early infancy, perhaps with a view to adapting to infants' developing abilities in processing speech (Kondaurova et al., 2013; Lee et al., 2014; Raneri, 2015).

The segmental properties of IDS also differ from that of ADS. This applies to both consonants and vowels; however, compared to vocalic differences, there are fewer studies on consonantal modifications in IDS compared to ADS (Cristia, 2010; Englund, 2005; Lee & Davis, 2010; Sundberg & Lacerda, 1999). At the heart of investigations into segmental differences between ADS and IDS lies the measure of vowel space area (Burnham et al., 2002; Cristia & Seidl, 2014; Englund, 2018; Kalashnikova et al., 2017; Kalashnikova & Burnham, 2018; Kuhl et al., 1997; Martin et al., 2015). This measure is traditionally operationalised as the area in acoustic space delineated by the average formant properties of the three 'articulatory extremes' – that is, the /i/-/a/-/u/-vowels – although newer studies are starting to compute the area encompassed by all of peripheral vowels (Marklund & Gustavsson, 2020; Rosslund et al., 2022). Adult speech perception studies show a straightforward relation between vowel space expansion and speech intelligibility (Ferguson & Kewley-Port, 2002, 2007; Bradlow et al., 1996; Whitfield & Mehta, 2019; Lam et al., 2012; Liu et al., 2005; Whitfield & Goberman, 2017). These studies argue that vowel space expansion improves speech intelligibility by ensuring distinguishable acoustic differences between vowels (Bradlow et al., 1996). Vowel space expansion has also been found in other speech registers that require intelligible speech, such as Lombard speech (i.e., speech in noise) (Castellanos et al., 1996; Tang et al., 2017) and foreigner-directed speech (Lorge & Katsos, 2019; Piazza et al., 2022; Uther et al., 2007). Pet-directed speech, on the other hand, which primarily serves an emotional function, does not usually exhibit vowel space expansion (Burnham et al., 2002), unless the pet has some perceived linguistic potential, such as a parrot (Xu et al., 2013) – although new evidence also indicates vowel space expansion for puppy-directed speech (Panneton et al., 2023). Vowel space expansion in IDS compared to ADS is highlighted as a robust feature of IDS in many languages, such as Japanese, Swedish, Russian and a number of English varieties (Cristia & Seidl, 2014; Kalashnikova et al., 2017; Kuhl et al., 1997). However, this pattern does not generalise across all languages, with languages such as Dutch, Cantonese Chinese, Norwegian and Danish showing *vowel space reduction* or no differences between IDS compared to ADS (Benders, 2013; Englund & Behne, 2005; Rattanasone et al., 2013; Rosslund et al., 2022). The majority of studies find no clear shifts in the extent of vowel space expansion in IDS during early development, and those studies that do indicate developmental changes disagree on the direction of the shift (Dodane & Al-Tamimi, 2007; Rattanasone et al., 2013).

The developmental relevance of IDS is well illustrated by its ability to attract infant attention; decades of research have investigated how infants respond to this distinctive style of speech and found that infants prefer to listen to IDS over ADS from a young age (Cooper & Aslin, 1990; Fernald & Kuhl, 1987; Pegg et al., 1992; Werker & McLeod, 1989; Zettersten, Cox, et al., 2023), even when IDS is presented in a foreign language (Byers-Heinlein et al., 2021; ManyBabies Consortium, 2020) or filtered to contain only prosodic information (Fernald & Kuhl, 1987). The attention-grabbing property of IDS goes hand in hand with studies showing its clear benefits to developmental outcomes; for example, IDS facilitates more effective neural tracking of speech material (García-Sierra et al., 2021; Kalashnikova & Burnham, 2018; Peter et al., 2016), better word segmentation abilities (Song et al., 2010; Thiessen et al., 2005; Zangl & Mills, 2007), improved discrimination between speech sounds (Kalashnikova & Carreiras, 2022; Trainor & Desjardins, 2002), and increased capacity to map novel words to their referents (Graf Estes & Hurley, 2013; Ma et al., 2011). A greater quantity of IDS also correlates with later expressive vocabulary (Hartman et al., 2017; Kalashnikova & Burnham, 2018; Porritt et al., 2014; Rosslund et al., 2022; Spinelli et al., 2017) and the complexity of vocalisations later in development (Marklund et al., 2021).

Many potential causal pathways of how IDS facilitates these developmental benefits have been proposed in theories of infant development. For example, IDS pitch and intonation may encourage and facilitate social interaction, helping language development through attention mechanisms (Gratier & Devouche, 2017; Saint-Georges et al., 2013). Or perhaps the developmental benefits of IDS may derive from the tendency for prosody to highlight certain linguistic information or pragmatic information (Bortfeld & Morgan, 2010; Ludusan et al., 2016; Soderstrom et al., 2008). Or maybe IDS prosody serves primarily to elevate emotional expression and regulate emotions, which then produces downstream effects on social cognition and language development (Trainor et al., 2000; Benders, 2013; Corbeil et al., 2013). Or maybe vowel space expansion in IDS creates a clearer speech signal that allows infants to distinguish between vowels and consonants in the speech stream more efficiently (Hartman et al., 2017; Kuhl et al., 1997; Liu et al., 2003). Or perhaps the slower rhythms of IDS may allow infants to participate in the interactive process more by giving them more time to respond (Ferjan Ramírez, 2022; Ferjan Ramírez et al., 2020). Some theories suggest that the function of IDS changes over during early development and adapts to infants' variable needs at different ages. IDS may thus initially serve primarily to draw infants' attention, modulate their temperament and express emotions, and later serve more specific linguistic and non-linguistic purposes (Eaves et al., 2016; Fernald, 1989; Peter et al., 2016; Snow, 1977b, 1977a).

The finding of vowel space expansion in some languages has played a large role in debates about the functions of IDS. Given its association with speech intelligibility in other speech styles, IDS is claimed to clarify the speech signal to aid infants' speech processing and language development (Hartman et al., 2017; Kuhl et al., 1997; Liu et al., 2003). However, fundamental limitations to vowel space measures as proxies for speech clarity requires hedging of these strong claims. One of these limitations is that studies often disregard the crucial assumption that an expansion in vowel space alone does not necessarily result in improved speech clarity. Some languages, for example, show vowel space expansion in combination with a higher degree of within-vowel variability (Martin et al., 2015; McMurray et al., 2013; Miyazawa et al., 2017; Rosslund et al., 2022). This in turn can result in *less intelligible* speech and make the task of learning sound categories harder (Benders, 2013; Cristia & Seidl, 2014; Miyazawa et al., 2017; Rosslund et al., 2022; Martin et al., 2015). In our analysis of Danish IDS in ***Study II***, we problematise the use of vowel space measures to make claims about the clarity and overlap of phonetic categories – and in turn, argue for the necessity of computing both within-vowel variability and between-vowel discriminability for accurate assessment (cf., Ludusan et al., 2021).

Another limitation of vowel space measures is that most of the studies on speech intelligibility and foreigner-directed speech have been conducted on a small subset of languages largely composed of English varieties (Whitfield & Goberman, 2014, 2017), and it remains unclear whether different phonological structures require different strategies of clarification (Benders, 2013; Englund, 2018; Rattanasone et al., 2013). In a vowel-rich language such as Danish (Basbøll, 2005; Grønnum, 1998), vowel space expansion may not be a viable strategy to increase perceptual differences among vowel categories due to the high number of vowel categories in the front part of the vowel space. In a language such as Danish, then, speech clarity may to a greater extent rely on features of articulation rate and speech segment durations, as shown in some adult perception experiments for other languages (Ferguson & Kewley-Port, 2007; Lam & Tjaden, 2013; Searl & Evitts, 2013). The jury is also still out on whether vowel space expansion in IDS is the consequence of a slower articulation rate and therefore being able to reach articulatory targets (Whitfield & Goberman, 2014, 2017), or whether it is produced as an unintended side effect of other variables, such as larynx raising (Kalashnikova et al., 2017), smiling (Englund, 2018), positive valence (Panneton et al., 2023), or elevated and variable pitch in the speech signal (McMurray et al., 2013).

What is common to all of the causal pathways reviewed here is the notion that these dynamic adaptations emerge from the social contingency of communicative signals and the need to accommodate the immaturity of the developing infant in the interactive process. Findings of acoustic changes in IDS features that suit infants' preferences (Kitamura & Burnham, 1998; Singh et al., 2002) and processing limitations (Christiansen & Chater, 2016; Saffran & Kirkham, 2018) suggest that active participation from both interactants play a crucial role in the developmental process (Goldstein & Schwade, 2008; Ko et al., 2016; Ritwika et al., 2020; Warlaumont et al., 2014). IDS holds a special status in early child development and depends crucially on caregivers' dynamic adaptation to the developmental states of their immature addressee (Fusaroli et al., 2019; Smith & Trainor, 2008). Disentangling the precise causal pathways through which IDS benefits child development is a difficult task due to the one-to-many relationships between individual acoustic features and their functions, as discussed further in **Section 3.3.2**.

1.1.2 The Diversity of Language Input across Cultures

The importance of social contingency in child-caregiver interactions also implies that the cultural context of the interaction matters; local and situated interactional dynamics to a great extent rely on non-local and trans-situational cultural routines and norms (Eliasoph & Lichterman, 2003; Fusaroli & Tylén, 2016; Steffensen & Pedersen, 2014; Tylén et al., 2010, 2013). This section considers the diversity of culture-specific types of child-caregiver interactions, with a view to gaining a fuller picture of the ways that the dynamics of the interaction can shape its speech content.

The sampling bias in social and developmental research (Christiansen et al., 2022; Kidd et al., 2023; Kidd & Garcia, 2022b, 2022c) applies no less in the context of IDS (Ochs & Schieffelin, 1984). Oversampling from particular populations severely constrains our understanding of the global variability in IDS features – and this in turn limits our ability to construct deeper theories about the features, functions and use of IDS across different cultures (Ingram, 1995; Pye, 1986; Ratner & Pye, 1984). More specifically, sampling from a narrow intersection of Western, Educated, Industrialised, Rich, and Democratic (WEIRD) cultures with specific child-rearing practices and child-caregiver interactive dynamics means that generalisations to other populations is not obvious (Nielsen et al., 2017). The research literature thus risks overestimating the importance of IDS in linguistic and cognitive development (Casillas et al., 2020; Cristia, 2023; Ingram, 1995; Ochs & Schieffelin, 1984). With these biases in mind, we can ask whether the nature and use of IDS differ across and within cultures, with a view to better understanding the diversity of language inputs that children experience during development.

A straightforward comparison of the quantity of IDS shows that the amount of input varies enormously both within and across cultures (Casillas et al., 2020; Cristia, 2023; Hart & Risley, 1992). One cultural factor is socioeconomic status, which should be construed as a heterogenous umbrella term that is composed of a set of correlated factors, such as income, parental formal education, and living situation (Pace et al., 2017; Schwab & Lew-Williams, 2016). Children growing up in North American households with a lower socio-economic status hear less IDS than children growing up in households with a higher socio-economic status (Dailey & Bergelson, 2022; Hart & Risley, 1992), and individual differences in language processing among Californian children aged 18-24 months from low socio-economic status background correlates with infant-directed input quantity (Weisleder & Fernald, 2013). Another axis of substantial variability in quantity involves that between urban and rural cultures; for example, the number of infant-directed utterances to 14-month-old children in urban communities of Mozambique is 5.7 times higher than those in rural communities (Vogt & Mastin, 2013), and there is an 11-fold difference in the quantity of input that children growing up in Chicago compared to a Yucatec Mayan village (Shneidman & Goldin-Meadow, 2012). Tsimane forager-horticulturalist cultures also spend less than one minute per daylight hour talking to children younger than four years (Cristia et al., 2019).

A substantial difference across cultures also involves who interacts with children and the extent to which other children function as caregivers (Weisner et al., 1977); for example, in many cultures, the number of vocalisations spoken by other children can be substantially higher than that of adult caregivers. This pattern has been found for a Yucatec Mayan village (Shneidman & Goldin-Meadow, 2012) as well as the indigenous Australian community of Yakanarra (Loakes et al., 2013). Adult caregivers have also been described as minor sources of language input for Samoan, Luo and Koya children (Snow & Ferguson, 1977) and to comprise a low proportion of the interactions experienced by Kenyan Kipsigis families (Harkness & Super, 2002). Comparisons between participants in Mayan and US English child-caregiver interactions reveal striking patterns; 40% of the total input heard by Mayan 13-month-old infants comes from adult caregivers (versus 92% in North America), and 65% of IDS comes from other children in the Mayan culture (versus 87% coming from primary caregivers in US English cultures) (Shneidman, 2010). Similar patterns apply for the Lesotho culture in South Africa (Loukatou et al., 2022).

Cross-cultural differences in parental strategies also influence infants' behaviour during development. Infants exhibit systematic variation in the amount of vocalisations and pointing they produce across different socioeconomic (Rowe & Goldin-Meadow, 2009; Warlaumont et al., 2014) and cultural groups (Salomo & Liszkowski, 2013). This pattern implies that cross-cultural differences in interaction dynamics change the shape of caregivers' accommodations to the needs of infants, and in turn, infants' interactive behaviour in response to caregivers' speech (Goldstein & Schwade, 2008; Ritwika et al., 2020; Warlaumont et al., 2014). This speaks in favour of the importance of interactional dynamics in shaping this speech style; parental input provides affordances for infants to start to produce more frequent babbling patterns and non-verbal communicative signals, such as pointing (Wu & Gros-Louis, 2015 and verbal vocalisations (Goldstein & Schwade, 2008; Gros-Louis, 2006; Warlaumont et al., 2014).

The substantial cross-cultural diversity in language input implies that there are many paths to language acquisition. This is further supported by observations that language development usually occurs at around the same age worldwide; for example, children acquire Inuktitut at ages comparable to middle-class North American children (Crago & Allen, 1997), and children learning Tseltal Mayan produce their first words at the same age as English-learning children (Bergelson et al., 2019; Casillas et al., 2020). These patterns may admit a crucial role for overheard speech in the developmental process, either directed to other adults or to other children (Loukatou et al., 2022). However, these patterns also provide an argument for stronger consideration that input quality in its full range of dimensions plays an important role over and above the role of quantity. Evidence from diverse societies suggests that in spite of its low prevalence in some cultures, dyadic adult-child conversations still promote lexical development more effectively than overheard speech (Shneidman, 2010; Shneidman et al., 2013; Weisleder & Fernald, 2013). By mainly focusing on the dimension of speech in the developmental process, we risk neglecting other rich social inputs, such as touch, gesture, emotion and action (Kosie & Lew-Williams, 2023; Schatz et al., 2022), which may also serve to scaffold the limited amounts of verbal language that infants receive from their caregivers.

The different structures of child-caregiver interactions reviewed above highlight the importance of viewing different cultural norms as adaptations to socio-historical contexts (Eliasoph & Lichterman, 2003; Steffensen & Pedersen, 2014; Tylén et al., 2010). For example, cultural adaptation to the environment (e.g., whether a hunting-gathering, urban-industrial or agrarian community) demands different parental strategies, and in turn shapes the speech content and temporal structure of child-caregiver interactions. Cultural differences in the amount of language socialisation a child is expected to be involved in can lead to caregivers dedicating varying amounts of time to engage children in activities like reading books, playing peek-a-boo, or other cognitive tasks (Bornstein & Putnick, 2012). The question of diversity in infants' early language experiences is key to developing theories about the role of IDS in the process of early infant development. Cross-cultural variability in the structures of child-caregiver interactions implies that the way that IDS is constructed and used is at least partially learned and culturally transmitted (Eliasoph & Lichterman, 2003; Steffensen & Pedersen, 2014;), as further explored in the next section **Section 1.1.3** and **Section 3.1**.

1.1.3 The Biophysical Origins of IDS & IDS as a Cultural Construct

The commonalities in the acoustic features of IDS across so many different languages beg the question of why caregivers would opt to modify their speech in such similar ways. Why do caregivers tend to speak in a *higher* pitch rather than a *lower* pitch when interacting with their infants? And why do caregivers naturally speak with *more* melody instead of speaking with *less* melody in their voice? These seem like trivial questions to address, but their triviality may derive from the features of IDS being shaped by a fundamental and general property of communicative signals (Bryant, 2022; Bryant & Barrett, 2007; Pisanski et al., 2022).

The forms of many types of vocal signals are shaped by their communicative functions (Bryant, 2022; Bryant & Barrett, 2007). The loud and noisy features of alarm calls may serve the function of attracting receivers' attention (Blumstein & Récapet, 2009; Owren & Rendall, 2001), and the abrupt features of prohibitives may serve to interrupt unwanted behaviour and redirect attention (Bryant, 2020, 2022). The unpleasant sound qualities of infant crying may likewise be shaped by the need to generate certain caregiver behaviours (Lingle et al., 2012; Parsons et al., 2012; Soltis, 2004; Vermillet et al., 2022), and these acoustic properties extend to distress calls in non-human species (Lingle & Riede, 2014). Similarly, the slow pitch intonations and metrical speech rhythms of infant-directed song and lullabies can soothe infants and establish emotional communion (Bryant & Barrett, 2007; Fernald, 1992; Hilton et al., 2022).

If the forms of IDS are shaped by their functions and if IDS serves similar functions across cultures, then we would expect IDS to exhibit clear acoustic regularities across languages. One of the most robust acoustic properties of IDS across cultures is a higher and more variable f_o. In non-human animals, a higher pitch in the signal denotes approachability and benignity in close contact calls and stands in stark contrast to the low-pitched and rougher sounds of aggression signals (Bryant, 2022; Pisanski et al., 2022). IDS prosody has been shown to be indistinguishable from the prosody of happy speech (Singh et al., 2002) and bears similarity with other affective speech registers like lovers' speech (Bombar & Littig Jr, 1996), infant-directed songs and lullabies (Hilton et al., 2022), and pet-directed speech (Burnham et al., 2002; Panneton et al., 2023; Xu et al., 2013). There is evidence that vowel space expansion may play a similar role to a higher pitch in this context. A detailed articulatory and acoustic study has found that vowel space expansion is likely a side effect of larynx raising (Kalashnikova et al., 2017), which shortens the vocal tract and raises the first and second formants of vowels (Barthel & Quené, 2015; Tartter, 1980). This allows speakers to sound smaller and convey friendliness and approachability (Ohala, 1980; Xu & Chuenwattanapranithi, 2007).

Because parental attention is a key resource for helpless infants to survive, infants are likely particularly interested in and reassured by speech with features suggesting that attention is directed toward them (Fernald, 1992). This implies that the melody is the message in child-caregiver interactions (Fernald, 1989); that is, the melodic nature and affective qualities of IDS function to signal approachability and to convey to infants that an adult is nearby, attending to them and keeping them safe (Fernald, 1989; Papoušek et al., 1991; Trainor et al., 2000). Caregivers' use of prosodic speech patterns in early development may thus primarily revolve around conveying a non-threatening demeanour, expressing positive emotional valence and grabbing infant attention (Kalashnikova et al., 2017). I discuss the biophysical properties underlying these behavioural adaptations further in **Section 3.2.3**.

The acoustic commonalities of IDS that we see across languages and cultures strongly imply a biophysical basis for this speech style; however, cross-cultural variation in its use and features implies that there are strong cultural constraints to keep in mind as well. There are normative pressures that guide caregivers' behaviour across cultures when it comes to IDS, as shown by the substantial variability in its prevalence across different cultures (Cristia, 2023) and clear links between parenting attitudes and certain interaction patterns (Johnston & Wong, 2002; Rowe, 2008). The influence of cultural norms on the nature of child-caregiver interactions imply that certain components of IDS are shaped by culture-specific constraints and affordances, as explored further in **Section 3.1**. Due to the sampling bias of languages and cultures in the literature

(Kidd et al., 2023; Kidd & Garcia, 2022a), there are crucial limitations in our knowledge of how IDS features change with different constellations of cultural variables. For example, cross-cultural variation in the value attached to verbal skills and educational achievement may produce differences in terms of the acoustic features of caregivers' IDS. Our theories need to accommodate cross-cultural variation in IDS acoustics as distinct adaptive design features that can result in differences across culture and context (Kidd et al., 2023; Ochs & Schieffelin, 1984).

Interim Conclusion

The acoustic features of IDS thus seem to be rooted firmly in biophysical considerations while also being influenced by cross-cultural differences in the dynamics of child-caregiver interactions. However, several unanswered questions about IDS emerge from this overview of the literature. To what extent do different acoustic features exhibit cross-linguistic variability? Do caregivers adapt the properties of their speech to infants' changing developmental needs during early development? To what extent do experimental designs and recording settings introduce biases in the results? We approach these questions in our meta-analysis of the acoustic features of IDS (***Study I***) and use insights and estimates from this meta-analysis to guide our own acoustic study of Danish IDS (***Study II***). In the remaining two sections, I provide a brief overview of the interactional structures that occur alongside IDS in the form of turn-taking (**Section 1.2**) and discuss how cumulative science can help us in the investigation of developmental phenomena (**Section 1.3**).

1.2 Turn-Taking as Interactional Infrastructure

Turn-taking is a form of interpersonal coordination that typically involves interactants avoiding overlaps and minimising the pauses between turns (Sacks, 1974). Turn-taking thus requires interactants to facilitate an efficient exchange of information by adapting their responses to each other (Levinson, 2006; Levinson, 2016). To engage in vocal communication with caregivers, young children have to adapt and respond to caregivers' cues in real time (Potter & Lew-Williams, 2019; Warlaumont et al., 2014). This complex dance of each interactant adapting their vocal responses to the other involves reciprocal behaviours that characterise many early child-caregiver interactions and require an elaborate set of communicative skills. In this section, I consider how infants' turn-taking abilities develop during early infancy (**Section 1.2.1**) and what the beginnings of this complex set of skills can reveal about mechanisms and models of turn-taking (**Section 1.2.2**).

1.2.1 Temporal Structure of Child-Caregiver Vocal Interactions

Turn-taking and language go hand in hand for adults, but long before children start to produce and understand words, they engage in reciprocal vocal communication with their caregivers. These exchanges exhibit strikingly similar timing structures to later verbal conversations (Hilbrink, 2015; Levinson, 2016) and are sometimes referred to as 'proto-conversations' (Gratier et al., 2015; Levinson, 2006). The modal response latencies of interactants in a typical adult conversation tend to be short, approximately 200 ms, which is at the absolute limits of human reaction time and too rapid to rely on planning and production in real time (Aron & Poldrack, 2006; Dingemanse & Liesenfeld, 2022; Stivers, 2009). People must therefore proactively prepare their contingent responses, detect when the likely endpoint of the other interactant's turn is, decide when to deliver their response, and predict how the other interactant will react (Magyari et al., 2014). To plan and predict with such efficiency, adult conversational partners draw upon their linguistic, interpersonal and world knowledge (Ford et al., 1996); for example, adults make predictions about the imminent content of a turn and use these predictions to estimate its duration – greater predictability facilitates more accurate timing estimations (Magyari et al., 2014, 2017; Magyari & De Ruiter, 2012). Preschool children similarly produce shorter response gaps when they can predict the end of a question (Lindsay et al., 2019). Inter-turn response times also have crucial social consequences; for example, a shared tempo between interactants reduces cognitive load and fosters a sense of connection and shared motivation (Dideriksen et al., 2020; Fusaroli et al., 2014, 2019; Fusaroli & Tylén, 2016; Konvalinka et al., 2011; Templeton et al., 2022, 2023; Tylén et al., 2013, 2013).

How do young children balance this complex juggling act of coordinating their attention, predicting turn ends and planning their responses, despite not knowing much about either language or the world? One way to get closer to answering this question is to observe how children integrate their burgeoning linguistic knowledge with their interactive skills. Infants' response latencies in early ontogeny are rapid and resemble adult response latencies before gradually slowing down at around 9 months of age and then accelerating again later (Gratier et al., 2015; Hilbrink, 2015). This nonlinear, parabolic trajectory of development may reflect the integration of infants' growing understanding and production of verbal content with their largely independent initial interactional abilities (Levinson, 2016). There is some coarse-grained evidence of bidirectional feedback between language development and the number of conversational turns initiated by both children and caregivers (Donnelly & Kidd, 2021; Gilkerson et al., 2018; Romeo et al., 2018), suggesting that more conversation begets greater levels of language knowledge, and vice versa. However, the upper bound of the influence of the parallel development of linguistic and interactional skills is unclear; older children between one and three years of age often struggle with conversational turns and frequently initiate them too late (Ervin-Tripp, 1982; Garvey & Berninger, 1981). Conversations with toddlers also exhibit frequent delays and non-contingent responses; however, these conversations tend to flow smoothly due to caregivers learning to adapt to the turn-taking behaviours of their young conversational partners (Dunn & Shatz, 1989; Ervin-Tripp, 1982).

Turn-taking in early ontogeny can thus be construed as a rapidly developing ecological niche of reciprocal interaction, which sows the seeds for infants to learn how to add linguistic meaning to their communicative repertoire over time (Donnelly & Kidd, 2021). The notion that interactional and linguistic skills develop in parallel matters for theories of language learning. For instance, when children produce preverbal vocalisations that bear a closer resemblance to language, their caregivers tend to respond more quickly – and in turn, quicker caregiver responses are linked to the emergence of language-like vocalisations in children, creating a social feedback loop (Lopez et al., 2020; Warlaumont et al., 2014). The verbal content of children's vocalisations can also elicit specific parental interactive behaviours; when 6-10-month-old infants produce consonant-vowel patterns, caregivers are more likely to respond with imitation or expansion, compared to situations where infants produce vowel-only patterns and caregivers tend to engage in more playful vocalisations (Gros-Louis, 2006). Other studies have expanded on this idea, revealing that parents take both the meaning and the sounds of what infants produce into account and respond with vocalisations and timings that suit infants' developing interests and skills (Braarud & Stormark, 2008; Elmlinger et al., 2019, 2023; Goldstein & Schwade, 2008; Gros-Louis, 2006, 2014; Miller et al., 2018; Smith & Trainor, 2008). Any sense of directionality between child and caregiver in the interactive process remains unclear because the timing and sounds of infants' vocalisations also display contingency with caregiver vocalisations (Hilbrink, 2015; Laing

& Bergelson, 2020). It should also be noted that the interactive component of infant vocal development is not the only factor at play; there is also a strongly endogenous and independent nature to early child vocalisations (Vihman, 2014, 2017), as shown by evidence of infants exhibiting a consistent rate of vocalisation regardless of whether they are alone or have adults nearby (Oller, Caskey, et al., 2019; Oller, Griebel, et al., 2019). These findings are not at odds with the reciprocal nature of early vocal exchanges in child-caregiver interactions, but rather highlight that infants' vocal development is also led by their internal motivation to explore speech sounds and to align what they can produce with what they perceive through interactions with their caregivers (Cox et al., 2020). These findings indicate that social interaction provides the locus for children's linguistic input, and that language is crucial for children's daily interactions (Lopez et al., 2020; Warlaumont et al., 2014).

The back-and-forth dynamics of early vocal exchanges can straightforwardly be extended to other dimensions and contexts of child-caregiver interaction. These co-produced structures can also appear, for example, when a child stacks cups to make a tower, and the caregiver stacks another set of cups to make a tower, and the actions are repeated. In these back-and-forth interactions, each interactant creates social affordances for the other to respond without necessarily relying on prompts from the interlocutor (Schertz et al., 2018). At the heart of turn-taking lies mutual interaction with the aim of sharing social interest (Harrist & Waugh, 2002; Schertz et al., 2018, 2023).

The interactive dynamics of turn-taking highlight that consideration of the content of IDS without the interactional structures alongside it severely limits our understanding. Directing language to infants in the form of IDS only represents one component of a complex interactive speech style that also involves providing social feedback and encouraging children to initiate and play an active role in interactive context (Ferjan Ramírez et al., 2020). Children's motivation to interact over and above the capacity to generate linguistic meaning reflects how central human interactional capacities are to language development (Bruner, 1974; Donnelly & Kidd, 2021; Goldstein & Schwade, 2008; Levinson, 2006; Smith & Trainor, 2008; Weisleder & Fernald, 2013). The social nature of early language development provokes questions about the nature of the underlying mechanisms of turn-taking, which I explore further in **Section 1.2.2**.

1.2.2 Cognitive Mechanisms & Models of Turn-Taking

The parallel development of turn-taking and language has provoked discussions about what sort of mechanisms are involved in this interactive skill. Because turn-taking develops early in ontogeny (Hilbrink et al., 2015), exhibits similar structures across languages (Stivers, 2009) and modalities (de Vos et al., 2015; Horton & Singleton, 2022), and has deep evolutionary roots (Levinson, 2006), these intricate timing structures may reflect a fundamental interactional infrastructure dating back to phylogeny prior to language (Levinson, 2016). Turn-taking has been argued to be part of and only party derivable from a package of fundamental capacities that enable human social behaviour, such as intention attribution, recipient design, social motivation, cooperative behaviour, multimodal integration, temporal sensitivities, and so forth (Levinson, 2006). This package of interactional abilities has been dubbed the Interaction Engine (Levinson, 2006; Levinson, 2016). These interactional abilities provide the foundations for language and place fundamental constraints on its characteristics. A substantial part of developing language is to learn to process and plan speech within the strict timing pressures of conversation (Christiansen & Chater, 2016). According to this theory, the slowdown in the timing of infants' turns at around one year of age – at the beginnings of sociocognitive (Tomasello, 1999) and verbal development (Bergelson & Swingley, 2018; Cychosz et al., 2021; Vihman, 2017) – reflects the pressure of coordinating several new cognitive processes within the strict timing structures of conversation; that is, infants' coordination of their existing interactional skills with their developing linguistic ones (Levinson, 2016). The Interaction Engine theory remains fairly underspecified with respect to the developmental trajectories that we might expect to see during early infancy and the respective contributions of individual differences in linguistic and social skills (Fusaroli et al., 2019; Warlaumont et al., 2014). In the meta-analysis on turn-taking in ***Study III***, we investigated the influence of key moderators to attempt to gain more information about the potential cognitive mechanisms underlying children's turn-taking abilities, although the theory requires greater degree of specification, as argued further in **Section 3.3**.

Whereas the Interaction Engine focuses on the underlying cognitive mechanisms required for turn-taking, computational modelling approaches provide a different perspective and attempt to capture the temporal structures that emerge from an interactant mutually adapting their timing to that of another interactant (Greenfield, 1994; Ravignani & de Reus, 2019; Takahashi, 2013). These models often draw inspiration from the concept of coupled oscillator dynamics, where individuals possess internal oscillators that become synchronised through the coordination of response times. For instance, the Coupled Oscillators Model (Wilson, 2005) suggests that conversational partners synchronise their vocalisations by establishing a shared tempo, ensuring that response times align with multiples of this shared tempo. This simplifies the task of when to begin a turn because specific timing intervals are preferred, reducing the likelihood of simultaneous speech. While the applicability of this model to human conversations has received little support (O'Dell, 2012), it finds some support in studies of animal turn-taking behaviour

(Takahashi, 2013). In **Section 3.3.4**, I argue that turn-taking research in humans can benefit from exploring the space of potential turn-taking structures by investigating non-human animal turn-taking in a shared computational framework (Cox, Templeton & Fusaroli, 2023; Verga et al., 2023).

Interim Conclusion

Infants' development of turn-taking skills aligns with developmental theories that construe language development as embedded within temporally contingent social interaction (Bolis & Schillbach, 2020; Vygotsky, 1978; Tomasello, 1999). A better overview of the developmental trajectories of turn-taking during early development would offer a more complete understanding of the underlying skills and cognitive mechanisms required. To map out research on developmental trajectories of child turn-taking and to provide recommendations for future investigations, we conducted a meta-analysis of child response latencies in ***Study III*** (**Section 2.3**).

1.3 Going Beyond Individual Studies

An idealistic portrayal of science would describe it as a collaborative and cumulative endeavour of iterative theory development that strives to establish regularities across different populations, conditions and contexts while recognising the limits of its own generalisations. Study-specific results in such a cumulative science play a processual and instrumental role in building better models and arriving at better explanations (Buzbas et al., 2023; Devezer et al., 2019; Devezer & Buzbas, 2023). In practice, however, scientists often construe the process of science as being concerned with establishing facts (Marks, 2009). This facts-focused approach to science inflates the contribution of individual studies to epistemic gain and underlies a number of problematic patterns in scientific culture and practice (Devezer et al., 2019; Devezer & Buzbas, 2023; Smaldino & McElreath, 2016; Yarkoni, 2022; Yarkoni & Westfall, 2017). These are important considerations to keep in mind when integrating empirical data into existing bodies of scientific knowledge through systematic meta-analyses, as is done in this dissertation. This section argues for the crucial importance of adopting cumulative science practices in the scientific process to ensure iterative and methodical development of theories over time. The first section outlines the current ecosystem of science and describes why theory development requires moving beyond reliance on individual studies (**Section 1.3.1**). The second section highlights how we can improve and evaluate the generalisability of phenomena by pooling results from individual studies via meta-analyses and Bayesian posterior passing (**Section 1.3.2**).

1.3.1 The Current Ecosystem of Science & Limitations of Individual Studies

Cumulative science fosters the continuous development of scientific understanding by systematically incorporating new research findings into an existing body of knowledge and improving our capacity to explain regularities (Koile & Cristia, 2021; Fusaroli et al., 2022; Ledgerwood, 2014). Theories play a key role in this process; they capture observed patterns and generate new predictions, providing explanatory power by isolating the causes and uncovering the mechanisms generating empirical regularities (Craver, 2009, 2014; Rohrer, 2018). Individual studies in this iterative process occupy patterns localised to one specific place in the methodological landscape, and therefore remain uninformed as to the generalisability of the result to other regions of the space with different dimensions, such as different experimental designs or sample characteristics (Machery, 2020; Yarkoni, 2022). In isolation, then, a study can only tell us so much, and one set of experimental results cannot be viewed as truth, nor can most experiments produce incontrovertible answers or seminal findings (Brand et al., 2019; Devezer et al., 2019). Rather, science rests on a process of epistemic iteration where successive stages of knowledge build upon preceding ones. This iterative construal of science has important consequences for inferential success and theory development. In the context of IDS research, for example, investigators initially observed its widespread use across various cultures and languages and hypothesised universality in its prevalence and acoustic features (Ferguson, 1964; Fernald, 1989). However, as reviewed above, increasing amounts of studies have found exceedingly low quantities of IDS in some cultures (Casillas et al., 2020; Cristia, 2023) and that language development progresses regardless (Casillas et al., 2020; Crago & Allen, 1997). Our theory of IDS thus needs to change to incorporate both the widespread acoustic features of IDS as a potential reflection of biophysical biases, but also recognise the influence of culture on these variables and the limits of generalisability as a result of bias in the field (Kidd et al., 2023; Kidd & Garcia, 2022a).

The specificity and unreliability of individual studies need to be recognised when we build theories and explanations about developmental phenomena. A single study on IDS directed to infants of a particular age in a given language and culture using a specific method and analysis to extract insights about behaviour in a specific interactive task cannot on its own inform a research question that extends to other contexts; rather, it is only through synthesis and iterative theory development that insights can be extracted and contextualised. The problem of study specificity is exacerbated by the ‘hidden universe of uncertainty’ of studies due to methodological and analytic choices (Yarkoni, 2020). Substantial unexplained variability across studies stems from researchers deciding how to operationalise a variable, design stimuli, analyse data, *et cetera ad infinitum*. Recent many-analyst studies where each research team tests the same hypothesis by analysing the same dataset show almost as many different interpretations as research teams (Botvinik-Nezer et al., 2020; Coretta et al., 2023; Silberzahn et al., 2018), although a large part of the variation in results stems from different interpretations of unclear and underdetermined research questions (Auspurg & Brüderl, 2021). The specificity of individual studies thus arises as a product of the multiverse of specifications resulting from choices of experimental methodology, participant samples, and data analysis pipelines. Neglecting these inherent study-specific sources of variability contributes to the lack of generalisability to other dimensions in the methodological landscape (Rocca & Yarkoni, 2021; Yarkoni, 2022).

Another reason for not being able to rely heavily on the results of a single study involves the inflation of findings and spurious values as well as low replication rates (Yarkoni, 2022; Ioannidis, 2005; Open Science Collaboration, 2015) as a result of questionable research practices (John et al., 2012; Simmons et al., 2011), selection pressures (Smaldino & McElreath, 2016; Francis, 2012), and misaligned incentives for scientific researchers (Brischoux & Angelier, 2015; Smaldino & McElreath, 2016). Many of these problems result from the common but dangerous tendency to construe statistics as an automatic way to turn data into incontrovertible conclusions (Devezer & Buzbas, 2023). This is a dangerous tendency because a single scientific hypothesis can receive support from many different statistical hypotheses and can therefore result in unreasonable flexibility in both directions: that is, *post hoc* formation of scientific hypotheses to fit the statistical analyses (Kerr, 1998; Leung, 2011) and biased selection of statistical analyses to align with the scientific hypothesis (Gelman & Loken, 2013).

Suggested paths of improvement involve restricting degrees of freedom in the data collection and analysis process through preregistration (Nosek et al., 2018) or registered reports (Chambers & Tzavella, 2022), and encouraging researchers to share open data (Morey et al., 2016; Nosek & Bar-Anan, 2012) and code (Clyburne-Sherin et al., 2019; Hardwicke et al., 2018) to facilitate verification of published results. However, these constraints have been argued to be insufficient measures to cure the epistemic problems of our discipline in that these improvements do not contribute to theory development (Szollosi et al., 2020; van Rooij & Baggio, 2020). There is some support for this from recent studies into the efficacy of preregistration showing that this procedure does not solve the problem of underdetermined hypotheses and unclear claims in both preregistrations and published papers (Claesen et al., 2021; van den Akker et al., 2023).

The root of the replicability crisis may be better construed as a crisis of inference (Guest & Martin, 2021; Rotello et al., 2015; Starns et al., 2019), which requires more conceptual and theoretical work to ameliorate it (Buzbas et al., 2023; Devezer et al., 2019; Devezer & Buzbas, 2023; Smaldino, 2019; Smaldino, 2017; Smaldino et al., 2015; Szollosi et al., 2020). Explicit theory development is a crucial component in a field like developmental psychology, where infant behaviours are noisy and highly variable no matter the degree of experimental control (Cristia et al., 2016; Schreiner et al., 2022). Replicability may thus be too tall of an order for infants to conform to, and more effort should accordingly be dedicated to theory development (Smaldino, 2019; Yarkoni, 2022). Theory development involves specifying the link between a verbal formulation and its quantitative operationalisation; that is, for a result to inform a verbal formulation, the measure must be a proper operationalisation of the verbal formulation (Devezer et al., 2019). If the mathematical expression is a poor operationalisation of the state of affairs of the verbal hypothesis, then the results from the statistical cannot inform the verbal formulation

(Devezer et al., 2019). If the connection between a conceptual construct and quantitative operationalisation is underspecified, then we run the risk of ‘not even being wrong’ (Scheel, 2022). An underspecified connection makes statistical results uninterpretable because it obscures which observations count as evidence for or against the claim. This link can often be evaluated before seeing data; for example, it is straightforward to evaluate the theoretical basis of a study showing that a dead salmon exhibits neural activity in response to photographs (Bennett et al., 2009). In a more realistic scenario, however, false discoveries based on poor or absent theory can be harder to identify (Gigerenzer, 2004; Ioannidis, 2005; Smaldino & McElreath, 2016).

1.3.2 How to Create Cumulative and Generalisable Science?

Science stands on the shoulders of – not giants – but normal human beings who are as susceptible to confirmation and selection biases as everyone else. In light of the challenging circumstances surrounding the ecosystem of science, how can we arrive at sound scientific explanations for empirical regularities? A cumulative and collective construal of science would argue that investigating regions of the methodological space in an exploratory and systematic way is crucial for theory development (Koile & Cristia, 2021; Devezer et al., 2019). The self-reflective nature of meta-science is useful for this exploratory purpose (Fusaroli et al., 2022). The accumulation of results across methodological intersections of different studies can offer critical self-reflection on the generalisability and limitations of our knowledge and offer data-driven suggestions of where to allocate resources for future research studies (Devezer & Buzbas, 2023; Zettersten, Cox, et al., 2023).

The limitations of relying on single studies have motivated a push towards methods of synthesising quantitative evidence in the form of meta- and mega-analyses (Koile & Cristia, 2021; Liu & Almeida, 2023; Zettersten et al., 2023). To counter selection biases and obtain an objective picture of a given field, there exist detailed procedures restricting researcher degrees of freedom in the process (Moher et al., 2009; Page et al., 2021). Synthesising evidence from studies in a cumulative approach offers insights to the generalisability and heterogeneity of the construct across a wide variety of experimental designs, participants and stimuli (Cristia et al., 2022; Cruz Blandón et al., 2023; Koile & Cristia, 2021; Tsuji et al., 2014). There are limitations to this process of estimation across studies, such as issues in transparency and reproducibility (Lakens et al., 2016, 2017; Nuijten et al., 2020), considerable errors of extraction (Zettersten, Cox, et al., 2023), reliance on published literature and concomitant bias (Moher et al., 2009; Page et al., 2021), as well as substantial between-study heterogeneity threatening practical interpretation of results (Mathur & VanderWeele, 2019). There are statistical approaches that attempt to correct for publication bias and to account for heterogeneity; however, major concerns about the validity of meta-analytic estimates remain (cf., Kvarven et al., 2020). These concerns primarily revolve around the reliability of meta-analytic estimates rather than the synthesis of insights and exploration of the sampling space of individual studies (Mathur & VanderWeele, 2020; Zettersten, Cox, et al., 2023).

A related cumulative method that synthesises findings across multiple individual experiments involves multi-lab replication studies. In such designs, multiple labs conduct replications of original studies by implementing a common experimental protocol to test a research question across sites (ManyBabies Consortium, 2020). By varying experimenter identity and increasing sample diversity, these large-scale studies contribute to a greater likelihood of generalisability and precision (Byers-Heinlein et al., 2021; ManyBabies Consortium, 2020). However, the high degree of uniformity in methodological and analytic implementation can lead to less generalisability across other conditions than a meta-analysis.

In a recent study (Zettersten, Cox, et al., 2023), we attempted to reconcile the results from a multi-lab replication study (ManyBabies Consortium, 2020) and a community-augmented meta-analysis on infants' preference to attend to IDS over ADS. The findings from each source individually provided an incomplete picture of moderators of the IDS preference effect; for example, the two studies differed with respect to their findings on the effect of infant age, with a linear increase in the multi-lab replication study and a finding of stability across infant ages in the meta-analysis. As we suggest in the paper, this conflict may originate in factors beyond the underlying construct. Investigators in the meta-analysis had the freedom to tailor their stimuli and methods to the particular infant age investigated, which may in turn mask age-related changes in the strength of the IDS preference effect. In contrast, the speech stimuli in the multi-lab replication study were held uniform across participating laboratories and may have been better suited for children in older age ranges (ManyBabies Consortium, 2020). The synthesis of these results underscores the value of being able to generalise across various populations, experimental methodologies and infant ages, but also highlights the importance of acknowledging the inferential limitations of evidence synthesis methods (Lakens et al., 2017). Seeing scientific advancement as an iterative procedure involving data accumulation and theory development empowers us to map out the diversity of samples in earlier research, scrutinise the possibilities for generalisability, and point to future directions of research (Fusaroli et al., 2022; Zettersten, Cox, et al., 2023).

Beyond meta-analytic synthesis and multi-lab replication strategies of generalising across different methodologies and samples, how can the cumulative nature of the scientific process be reflected in our choice of statistical methods? At the heart of cumulative science approaches lies the idea of gradual accumulation of knowledge by updating the generalisability of our beliefs with new empirical data. Updating a belief based on new evidence is central to Bayesian approaches to statistics (Brand et al., 2019; McElreath, 2018). A prior probability represents initial beliefs, which are then updated with new data to obtain a posterior distribution that reflects updated beliefs. We can apply this process of Bayesian inference to the scientific process via a statistical method called posterior passing (Beppu & Griffiths, 2009). This method involves using a posterior estimate from an earlier study as the prior for the next study, creating a string of connected studies where each study incorporates the data collected in earlier studies and contributes to the precision of the next posterior estimate (Beppu & Griffiths, 2009). The passing of posteriors across studies reduces the influence of any single experimental dataset and generates critical reflection on how to contextualise and incorporate the results of previous studies into the analysis (Brand et al., 2019). For example, how to best incorporate uncertainty in the prior and ensure appropriateness involves consideration of how individual studies address the same theoretical effect, and whether differences in analytic technique or experimental design should result in the inflation of prior uncertainty (Brand et al., 2019; Wesner & Pomeranz, 2021). Experiments can be combined in posterior passing if they are addressing the same theoretical effect. For example, Dideriksen et al. (2023) investigate task-oriented conversational structures and use informed priors based on the previous literature on casual conversations. Similarly, in ***Study II*** of this dissertation, we employ the average meta-analytic estimates of IDS acoustic features from ***Study I*** to encode our data-driven prior expectations and to critically evaluate the extent to which Danish IDS follows or contradicts cross-linguistic tendencies. This statistical method thus offers a back-and-forth conversation between prior studies and new results, but it does not obviate the need for detailed modelling of data from individual behavioural studies. The cumulative nature of the ideal scientific process, then, can be reflected in our choice of statistical methods and create iterative improvement of our estimates and theories over time (Brand et al., 2019).

Interim Conclusion

In a cumulative perspective of science, individual studies represent patterns in specific regions within the methodological landscape, and based on these patterns, scientists gradually build up theories and explanations for variability and generalisability of results across studies. By aggregating empirical results in meta-analyses and using these aggregations to encode our prior expectations in statistical models, these methods offer an explicit method to integrate new research findings into existing frameworks of knowledge. This cumulative perspective on scientific exploration should be evident in the aims and motivations of the studies in this dissertation.

1.4 Aims and Motivations of the Studies

The reviewed literature indicates that behavioural adaptations emerge from social feedback loops between interactants and are shaped by changes across ontogeny, cross-cultural norms and form-function relationships (Elmlinger et al., 2019, 2023; Goldstein & Schwade, 2008; Laing & Bergelson, 2020; Warlaumont et al., 2014; Bolis & Schilbach, 2018, 2020; Fusaroli & Tylén, 2012). The remainder of **Section 1.4** provides brief overviews of the specific aims of each of the studies included in this dissertation.

1.4.1 Aims & Motivations of *Study I*

Study I is a systematic review and Bayesian meta-analysis of the acoustic features of IDS and quantifies the available evidence for each of the following five acoustic features: f_o, f_o variability, articulation rate, vowel duration, and vowel space area. The specific aims of the meta-analysis were thus to examine the following three main questions:

1. How do the five acoustic properties of IDS change over the course of early child development?
2. To what extent do these features exhibit cross-linguistic variability?
3. To what extent do these features provide commensurable measurements according to the experimental task and recording environment?

1.4.2 Aims & Motivations of *Study II*

Study II goes hand in hand with the meta-analysis on IDS and builds on the insights from previous studies of Danish IDS (Bohn, 2013; Dideriksen & Fusaroli, 2018) and IDS across distinct languages (Cox et al., 2023). We turned our focus to the following three questions concerning the acoustic expression of Danish IDS and ADS:

1. To what extent do Danish caregivers produce differences across IDS and ADS in terms of the five acoustic features explored in the meta-analysis: f_o, f_o variability, articulation rate, vowel duration, and vowel space area?
2. Do Danish caregivers exhibit acoustic changes in their IDS across different infant ages in early ontogeny?

3. What is the balance between vowel separability and within-vowel variability in Danish caregivers’ IDS and ADS?

1.4.3 Aims & Motivations of *Study III*

Study III provides a systematic review and Bayesian meta-analysis on the developmental trajectory of children’s turn-taking abilities and attempts to identify key moderators affecting child response latencies. We asked the following questions:

1. What is the developmental trajectory of infant response latencies in early child-caregiver interactions?
2. To what extent do these developmental changes in child response latency change with child age, adult response latency, interlocutor familiarity, interactional setting, activity type, language and developmental atypicalities?
3. Can we gain more information on the potential mechanisms underlying turn-taking?

Study I

A Systematic Review and Bayesian Meta-Analysis of the Acoustic Features of Infant-Directed Speech

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

[1] Department of Linguistics, Cognitive Science & Semiotics, Aarhus University
[2] Interacting Minds Centre, Aarhus University
[3] Department of Language & Linguistic Science, University of York
[4] Max Planck Institute for Psycholinguistics, Nijmegen
[5] Department of Communication, University of California, Los Angeles

Cox, C., 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.

Acknowledgements

The authors declare no competing interests, nor any conflicts of interest with regard to the funding source for this study. RF has been a paid consultant for F. Hoffman La Roche on related but non-overlapping topics. This project has been supported by seed funding from the Interacting Minds Centre at Aarhus University. All of the computation done for this project was performed on the UCloud interactive HPC system, which is managed by the eScience Center at the University of Southern Denmark. The analysis and visualization code and a reproducible R Markdown manuscript are available and permanently archived in the following open repository: https://osf.io/hc7me/.

Correspondence: chris.mm.cox@gmail.com (Chris)

University of York
York Graduate Research School
Research Degree Thesis Statement of Authorship

Candidate name	Christopher Martin Mikkelsen Cox
Department	Department of Linguistics & Cognitive Science
Thesis title	How Social Contingency Shapes Early Child-Caregiver Interactions

Title of the work (paper/chapter)	A systematic review and Bayesian meta-analysis of the acoustic features of infant-directed speech	
Publication status	**Published**	X
	Accepted for publication	
	Submitted for publication	
	Unpublished and unsubmitted	
Citation details (if applicable)	Cox, C., 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.	

Description of the candidate's contribution to the work*	Conceptualization, Methodology, Software, Validation, Formal analysis, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision, Project administration
Approximate percentage contribution of the candidate to the work	70 %
Signature of the candidate	
Date (DD/MM/YY)	24-11-23

Co-author contributions

By signing this Statement of Authorship, each co-author agrees that:
(i) the candidate has accurately represented their contribution to the work;
(ii) if required, permission is granted for the candidate to include the work in their thesis (note that this is separate from copyright considerations).

Name of co-author	Christina Bergmann
Contact details of co-author	Chbergma@uos.de
Description of the co-author's contribution to the work*	Conceptualization, Writing (original draft and revisions)
Approximate percentage contribution of the co-author to the work	5%
Signature of the co-author	
Date (DD/MM/YY)	24-11-2023

Name of co-author	Tamar Keren-Portnoy
Contact details of co-author	tamar.keren-portnoy@york.ac.uk
Description of the co-author's contribution to the work*	Participated in the writing and interpretation of the data.
Approximate percentage contribution of the co-author to the work	5%
Signature of the co-author	Tamar K
Date (DD/MM/YY)	3-12-2023

Name of co-author	Greg Bryant
Contact details of co-author	UCLA Dept of Communication, 2225 Rolfe Hall, LA, CA 90042 US Email: gabryant@ucla.edu
Description of the co-author's contribution to the work*	Formulation of ideas; writing.
Approximate percentage contribution of	10%

the co-author to the work	
Signature of the co-author	
Date (DD/MM/YY)	24-11-23

Name of co-author	Emma Fowler
Contact details of co-author	EmmaFowler1@hotmail.com
Description of the co-author's contribution to the work*	Data Collection
Approximate percentage contribution of the co-author to the work	5%
Signature of the co-author	Emma Fowler
Date (DD/MM/YY)	27-11-23

Name of co-author	Riccardo Fusaroli
Contact details of co-author	fusaroli@cc.au.dk
Description of the co-author's contribution to the work*	Conceptualization, Methodology, Software, Validation, Formal analysis, Resources, Data Curation, Writing - Original Draft, Visualization, Project administration
Approximate percentage contribution of the co-author to the work	5%
Signature of the co-author	
Date (DD/MM/YY)	24-11-23

A Systematic Review and Bayesian Meta-Analysis of the Acoustic Features of Infant-Directed Speech

Abstract

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

1 Introduction

1.1. The form and function of infant-directed speech

Speaking to infants presents caregivers with a significant challenge. Because infants are not linguistically competent, older individuals modify their speech to them in a variety of ways to communicate. The ways in which caregivers produce infant-directed speech (IDS) have been widely documented, and some clear patterns have emerged across multiple languages. For example, speakers often increase their vocal pitch and pitch variability, slow down their speech, and articulate more clearly (e.g., Fernald, 1989; Fernald et al., 1989; Stern et al., 1983). The discovery of similar acoustic properties of IDS across so many languages and cultures strongly suggests that this speech style plays an important role in linguistic and social development (Golinkoff et al., 2015).

In the study of signal design in humans and nonhuman animals, form-function analysis is used to understand how structural characteristics of signals are shaped by the communicative functions they serve. This approach applies well to the study of IDS (Bryant & Barrett, 2007; Fernald, 1992; Owren & Rendall, 2001). For instance, the loud, low-pitched, abrupt onset of a prohibitive yell could be designed to interrupt the behavior of a baby by exploiting the startle reflex, which quickly re-orients a target infant's attention to the sound source (Bryant & Barrett, 2007). Similarly, approval vocalizations may induce positive emotions through raised pitch and pitch variability, faster speech, and modulated loudness reflecting speakers' positive valence and heightened arousal (Bryant & Barrett, 2007; Kitamura & Lam, 2009; Lam & Kitamura, 2006). But communicative functions overlap and interact as the cognitive and linguistic skills of the infant develop, and their interactional affordances change (Kitamura & Notley, 2009; McRoberts et al., 2009; Panneton et al., 2006).

One prominent hypothesis holds that the acoustic features of IDS may help infants learn aspects of language (Golinkoff et al., 2015). The benefits of IDS to language development are generally attributed to its tendency to increase the clarity of the speech input (e.g., Kalashnikova & Burnham, 2018; Kuhl et al., 1997; Liu et al., 2003). This hypothesis receives substantial support from longitudinal studies showing positive correlations between parents' tendency to produce acoustically exaggerated vowels and speech discrimination skills (Liu et al., 2003) as well as expressive vocabulary size (Hartman et al., 2017; Kalashnikova & Burnham, 2018). Other studies show that acoustically exaggerated vowels induce more mature neural processing of vowel categories in infants (Peter et al., 2016) and faster word recognition (Song et al., 2010). The cross-linguistic tendency for caregivers to exaggerate the differences between vowel categories might facilitate infants' language development by increasing category separability in the speech stream. An increase in vowel category separability in speech has been shown to co-occur with a greater degree of within-category variability (Cristia & Seidl, 2014; Martin et al., 2015; McMurray et al., 2013; Miyazawa et al., 2017; Rosslund et al., 2022), which may work in parallel with separability to increase the robustness and generalizability of the categories (Eaves et al., 2016; Perry et al., 2010; Rost & McMurray, 2009, 2010).

The functions of IDS have been posited to exhibit change over the course of early infant development, with the speech style initially serving primarily to direct infants' attention and express affect, and later on serving more specific linguistic purposes (Fernald, 1992). According to a form-functions analysis, these age-related changes in the functions of IDS should manifest themselves in the acoustic properties of caregivers' speech. Despite the implications of unidirectionality in its name, however, IDS also includes feedback from infants – IDS involves reciprocity and interaction where the interdependence of infants' active participation and caregiver responsiveness plays a crucial role (Goldstein & Schwade, 2008; Ko et al., 2016; Murray & Trevarthen, 1986; Nguyen et al., 2022; Warlaumont et al., 2014). The benefits of IDS should be construed as originating in the mutual feedback loops between infant and caregiver, where infants provide an important source of feedback about which signals they prefer to attend to and interact with (Goldstein & Schwade, 2008; Ko et al., 2016; Murray & Trevarthen, 1986; Nguyen et al., 2022; Warlaumont et al., 2014).

1.2. Development of infants' IDS preference

Many studies have demonstrated that infants prefer to listen to IDS over ADS (Cooper & Aslin, 1990; Fernald, 1989; Fernald et al., 1989; Fernald & Simon, 1984; Kuhl et al., 1997; ManyBabies Consortium, 2020; Pegg et al., 1992; Werker & McLeod, 1989). This preference persists when presented speech is in a foreign language (ManyBabies Consortium, 2020; Werker & McLeod, 1989), or when low-pass-filtered and containing only global prosodic information (Fernald & Kuhl, 1987). Even infant-directed songs in a foreign language induce relaxation in babies (Bainbridge et al., 2021). A recent large-scale, multi-lab replication study found that infants exhibit linear increases in their IDS preference until at least 15 months of age, the oldest age tested (ManyBabies Consortium, 2020). This trajectory was similar to the findings of a meta-analysis reporting a general increase in looking times toward IDS in preverbal infants from 0 to 9 months (Dunst et al., 2012). In contrast, two studies have reported that infants' IDS preference exhibits a U-shaped pattern. Hayashi et al. (2001) found that while both groups of 4-6- as well as 10-14-

month-old infants paid more attention to IDS than ADS, but 7-9-month-old infants did not exhibit a preference. Similarly, Newman & Hussain (2006) found a preference for IDS in 5-month-old infants, but not in 9- or 13-month-olds.

Infants' shifting preferences for IDS over ADS over the first year of life could reflect dynamic changes in the acoustic features they attend to. For example, Panneton et al. (2006) reported that 4-month-old infants listened longer to speech with a higher positive affect (i.e., relatively higher emotion content) and slowed duration, but 8-month-old infants preferred speech with normal duration and lower relative affect. Other studies examining differences in preferences have demonstrated various effects suggesting that infants, even during their first year, might be attending differentially to many aspects of IDS (Kitamura & Lam, 2009; Kitamura & Notley, 2009; Lam & Kitamura, 2006). For example, younger infants have been shown to preferentially attend to the intonational variability and positive affect of IDS (Kitamura & Burnham, 1998; Singh et al., 2002). At this early developmental stage, the tendency for IDS to contain increased pitch variability, modulated loudness contours, and rhythmic alterations (Fernald & Mazzie, 1991; Fernald & Simon, 1984) likely serves the function of effectively communicating intentions, including getting an infant's attention, expressing emotions and encouraging behavior (Fernald, 1992). As infants get older and become more advanced in language development, their attention might shift toward aspects of IDS that provide linguistic information (Kitamura & Notley, 2009; McRoberts et al., 2009; Segal & Newman, 2015). If caregivers adapt the acoustic properties of their IDS to suit infants' developmental needs, we may see systematic shifts in acoustic properties over the course of early infancy, such that exaggerated prosodic features associated with communicating intent to young infants should decline, and linguistically-relevant properties should be emphasized more for older children, including expansion of the vowel space area (Fernald, 1992; Kuhl et al., 1997).

1.3. IDS across cultures

The study of IDS across cultures has a long interdisciplinary history. Early linguistic research revealed many regularities in IDS across disparate languages and cultures, as well as language-specific phenomena. In this work, many of the reported features were not acoustic but concerned phenomena such as modified morphemes and grammatical constructions as well as lexical innovations (e.g., Ferguson, 1964). Naturally, these kinds of features should vary cross-culturally, and variations were noted within villages, including features that were unique to single families, or that might spread to a few households at most. Ferguson (1964) also discussed cultural variations in attitudes toward babytalk, including its use in public and whether it was more appropriate for men or women to produce it. Other studies have shown that the frequency of speaking to infants in any manner can vary dramatically, with some cultural groups not speaking to infants very much at all (Casillas et al., 2020; Cristia et al., 2019; Shneidman & Goldin-Meadow, 2012). A high degree of variability in the rate of IDS use, however, does not preclude universality (Bryant, 2022); rather, IDS may represent a continuum across cultures that exhibits cross-linguistic variability in its rate and acoustic properties. Early rejections of the universality of IDS often conflated the issues of incidence with form; that is, how often IDS occurs during interaction is separate from its acoustic features when it is actually produced. Later analyses focusing on acoustic characteristics of IDS across languages have revealed striking similarities (Fernald et al., 1989; Grieser & Kuhl, 1988; Papoušek et al., 1991). Recent large-scale studies have shown that these features occur widely, and the recognition of IDS and infant-directed song is robust (Hilton et al., 2022; ManyBabies Consortium, 2020). Questions regarding within- and between-culture variation are crucial to address when issues of universality are raised (Bryant, 2022).

Researchers have now started using day-long recordings of infants (e.g., Räsänen et al., 2021; Xu et al., 2009) and open archives of acoustic data (e.g., MacWhinney, 2014), allowing for the analysis of more ecological data to investigate infants' linguistic and emotional development through quantitative and computational means (e.g., Warlaumont et al., 2014). These archives provide data from diverse cultures (e.g., Casillas et al., 2020) and offer new insights into the role of linguistic input on early language development. For example, US English speakers appear to produce a particularly exaggerated form of IDS relative to other speakers (Fernald et al., 1989; Floccia et al., 2016; Shute & Wheldall, 1999). Because such a high proportion of studies on IDS examine US English (cf. Figure 1 and Table S10.1 in Supplementary Information), the field may have a biased view of how IDS manifests itself, and how it may affect language development (Floccia et al., 2016; ManyBabies Consortium, 2020). Figure 1 shows the proportion of languages for which IDS has been analyzed compared to the total number of languages listed on the World Atlas of Language Structures (Haspelmath, 2009). Although this world map suggests a considerable bias in the types of languages and cultures investigated, increasing linguistic diversity – while valuable in and of itself – is unlikely to improve our understanding of the cognitive underpinnings of IDS alone. More fine-grained, hypothesis-driven comparisons are also required (Christiansen et al., 2022; Deffner et al., 2021; Trecca et al., 2021), as discussed further in sections 4.2-4.3. In order for such comparative approaches to be useful, we need a more careful and theory-driven analysis of the extant IDS literature and how IDS varies across infant ages, languages, experimental tasks and recording environments. It should also be noted that the participants in the studies included in this meta-analysis largely consist of female caregivers residing in Western, educated, industrialized, rich, developed countries (Nielsen et al., 2017). Due to the sparsity of the data on other speaker types and populations, the meta-analysis could not analyze these factors as potential sources of variability in the acoustic measures (e.g., kin vs. non-kin caregivers), as discussed further in sections 4.2-4.3.

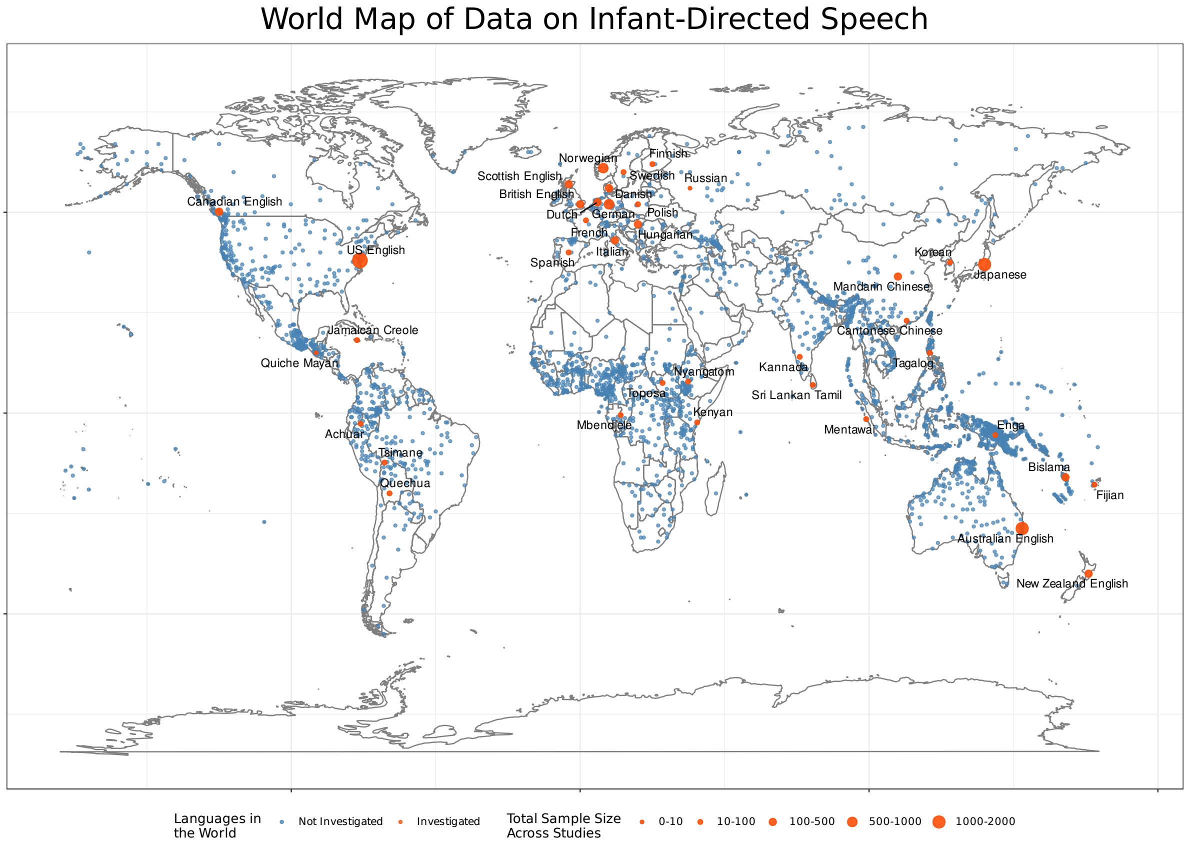


Figure 1: World map of IDS data. This plot provides a coarse overview of languages for which IDS has been analyzed. It compares the languages included in this meta-analysis to the languages listed on the World Atlas of Language Structures (WALS) [60]*. Data extracted from: https://github.com/cldf-datasets/wals. Each point represents a language, the color of which indicates whether the language is included in this meta-analysis, and the size of which indicates the cumulative sample size. It should be noted that the represented languages may not be accurate in terms of exact positions on the map and that WALS includes languages with very small speaking communities (e.g., the vast number of languages in USA).*

1.4. Acoustic changes in IDS across development

Many studies have demonstrated that caregivers exhibit age-related changes in the acoustic properties of their IDS. Here we will provide an overview of how each of the acoustic features in IDS that we investigated in our meta-analysis have been shown to change as a function of infants' age. See Figure 2 for a summary visualization.

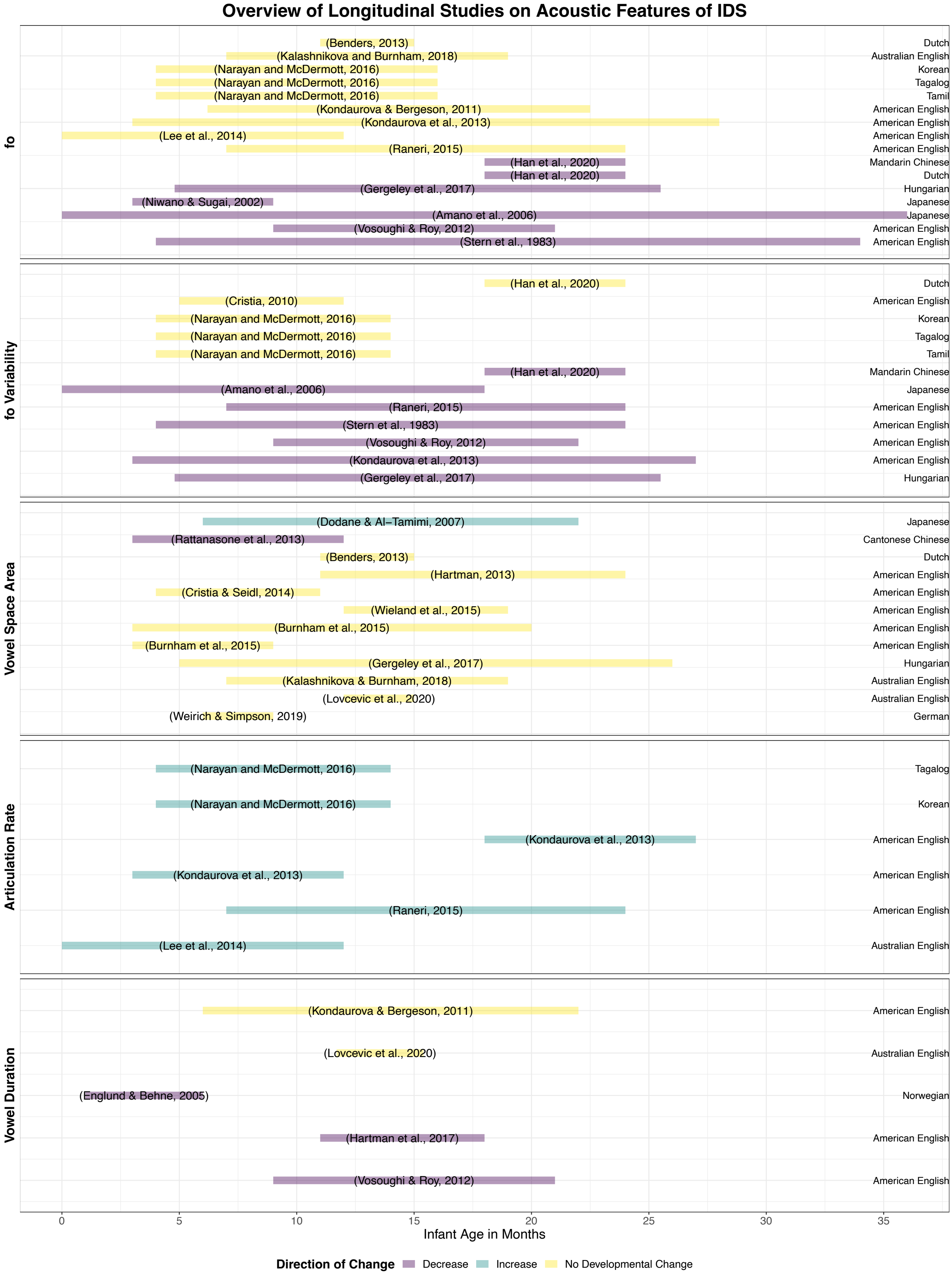

Overview of Longitudinal Studies on Acoustic Features of IDS
fo
(Benders, 2013) Dutch
(Kalashnikova and Burnham, 2018) Australian English
(Narayan and McDermott, 2016) Korean
(Narayan and McDermott, 2016) Tagalog
(Narayan and McDermott, 2016) Tamil
(Kondaurova & Bergeson, 2011) American English
(Kondaurova et al., 2013) American English
(Lee et al., 2014) American English
(Raneri, 2015) American English
(Han et al., 2020) Mandarin Chinese
(Han et al., 2020) Dutch
(Gergeley et al., 2017) Hungarian
(Niwano & Sugai, 2002) Japanese
(Amano et al., 2006) Japanese
(Vosoughi & Roy, 2012) American English
(Stern et al., 1983) American English
fo Variability
(Han et al., 2020) Dutch
(Cristia, 2010) American English
(Narayan and McDermott, 2016) Korean
(Narayan and McDermott, 2016) Tagalog
(Narayan and McDermott, 2016) Tamil
(Han et al., 2020) Mandarin Chinese
(Amano et al., 2006) Japanese
(Raneri, 2015) American English
(Stern et al., 1983) American English
(Vosoughi & Roy, 2012) American English
(Kondaurova et al., 2013) American English
(Gergeley et al., 2017) Hungarian
Vowel Space Area
(Dodane & Al−Tamimi, 2007) Japanese
(Rattanasone et al., 2013) Cantonese Chinese
(Benders, 2013) Dutch
(Hartman, 2013) American English
(Cristia & Seidl, 2014) American English
(Wieland et al., 2015) American English
(Burnham et al., 2015) American English
(Burnham et al., 2015) American English
(Gergeley et al., 2017) Hungarian
(Kalashnikova & Burnham, 2018) Australian English
(Lovcevic et al., 2020) Australian English
(Weirich & Simpson, 2019) German
Articulation Rate
(Narayan and McDermott, 2016) Tagalog
(Narayan and McDermott, 2016) Korean
(Kondaurova et al., 2013) American English
(Kondaurova et al., 2013) American English
(Raneri, 2015) American English
(Lee et al., 2014) Australian English
Vowel Duration
(Kondaurova & Bergeson, 2011) American English
(Lovcevic et al., 2020) Australian English
(Englund & Behne, 2005) Norwegian
(Hartman et al., 2017) American English
(Vosoughi & Roy, 2012) American English
0
5
10
15
20
25
30
35
Infant Age in Months
Direction of Change
Decrease
Increase
No Developmental Change

Figure 2: An overview of the findings from longitudinal studies for each of the acoustic features. The x-axis indicates infant age in months, while the colours indicate whether the acoustic feature exhibits an increase (orange) or decrease (teal) or no change (purple) over the course of the ages investigated. Studies are sorted by the main conclusion regarding any feature change.

The most common finding in studies examining the acoustic features of IDS is that IDS utterances, on average, have higher f_o and f_o variability than ADS, resulting in the salient perceptual effects of perceived higher pitch and pitch variation (e.g., Fernald, 1989; Stern et al., 1983). Interestingly, many longitudinal studies on f_o show that caregivers decrease their overall vocal pitch to infants over the course of development (Amano et al., 2006; Gergely et al., 2017; Han et al., 2020; Niwano & Sugai, 2002; Stern et al., 1983; Vosoughi & Roy, 2012), but findings are mixed with other studies reporting no change over time (Benders, 2013; Kalashnikova & Burnham, 2018; Kondaurova et al., 2013; Kondaurova & Bergeson, 2011; Lee et al., 2014; Narayan & McDermott, 2016; Raneri, 2015). Variability in f_o shows a similar pattern. Pitch variation reflects intonational contours that provide information about speakers' expression of affect and intentions (e.g., Fernald & Simon, 1984; Knoll & Costall, 2015). Longitudinal studies of f_o variability in IDS indicate a peak before infants turn 12 months of age, with a subsequent decrease over the course of development (Amano et al., 2006; Cristia, 2010; Gergely et al., 2017; Han et al., 2020; Kondaurova et al., 2013; Narayan & McDermott, 2016; Raneri, 2015; Stern et al., 1983; Vosoughi & Roy, 2012).

The tendency for caregivers to expand their vowel space area in IDS represents one of the more subtle adaptations of speech directed to infants. The most common measure calculates the area in acoustic space encompassed by the mean formant values of the three corner vowels: /i/, /a/, and /u/. Because these three vowels represent articulatory extremes and occur in the majority of the world's languages (e.g., Liljencrants & Lindblom, 1972), studies focus on how caregivers adapt the acoustic-phonetic characteristics of these vowels in their IDS. Thus, vowel space area is used as a measure of how much caregivers clarify their speech to infants (e.g., Hartman et al., 2017; Liu et al., 2003; although cf., Cristia & Seidl, 2014; Martin et al., 2015; Miyazawa et al., 2017). A majority of studies do not find evidence of any shift in vowel space area at a variety of age ranges (Benders, 2013; Burnham et al., 2015; Cristia & Seidl, 2014; Gergely et al., 2017; Hartman, 2013; Kalashnikova & Burnham, 2018; Lovcevic et al., 2020; Weirich & Simpson, 2019; Wieland et al., 2015). But some studies have shown changes over time, although there are differences in the direction of the shift (Dodane & Al-Tamimi, 2007; Rattanasone et al., 2013).

Articulation rate measures the speed at which people speak, which can have important consequences for how easily language is processed. This is true not only for young infants, but adults as well, including second-language learners and listeners with other impairments (e.g., Huettig & Guerra, 2019). Speaking too fast can prevent proper processing, which could have effects on phonological perception, emotional communication, and other comprehension issues. Several longitudinal studies of articulation rate have shown that caregivers increase their rate of articulation (i.e., speed up speech) over the course of infant development (Kondaurova et al., 2013; Lee et al., 2014; Narayan & McDermott, 2016; Raneri, 2015). Vowel duration, lastly, plays a crucial role in phonological processing (e.g., Fernald et al., 1989) as well as in modulating infant attention and facilitating language development (Gleitman et al., 1984). The exaggeration of the duration of vowels in IDS may make relevant phonological differences more salient to children, thereby facilitating their detection of clause and phrase boundaries (Seidl & Cristià, 2008; Soderstrom et al., 2003). Longitudinal studies in several languages indicate that caregivers often decrease relative vowel duration differences in IDS and ADS as infants get older (Englund & Behne, 2005; Hartman et al., 2017; Vosoughi & Roy, 2012).

1.5. Objectives of the meta-analyses

In the current meta-analysis, we aimed to investigate the acoustic properties of IDS across infant ages and languages, and understand these results in relation to the purported functions of IDS. We conducted this investigation by examining the influence of four moderator variables on possible acoustic differences between ADS and IDS: i) age, ii) language, iii) experimental task and iv) recording environment. The justification for each is described in brief. First, by pooling together data from the studies and quantifying the acoustic changes in IDS as a function of infant age, we can examine which of the acoustic properties of IDS change to become more similar to ADS over early infant development. Specific changes in the acoustic properties of IDS over developmental time would suggest that caregivers exhibit sensitivity to infants' shifting socio-emotional and linguistic needs and adapt their speech accordingly. Concretely, if IDS in early development serves primarily to convey affect and only later serves a linguistic function, then we might expect to see developmental shifts in the acoustic properties that are primarily associated with linguistic facilitation (e.g., vowel space area and vowel duration). Whether these linguistic features are present from birth or become gradually more exaggerated in IDS as infants exhibit linguistic development remains an open empirical question. Over longer timescales (not covered by the studies in this meta-analysis), we would expect all of the acoustic properties of IDS to gradually become indistinguishable from those of ADS. Second, to quantify the amount of cross-linguistic variation that could be observed, we analyzed language as a moderator variable. For each acoustic variable, we provided language-specific estimates for each of the languages under investigation, as shown in Tables S9.1-9.5 in the Supplementary Information. The data were too sparse to allow for an investigation of an interaction between infant age and the language spoken (cf. Figure S7 in the Supplementary Information). Last, we analyzed experimental task (i.e., spontaneous vs. read speech) and recording environment (i.e., naturalistic vs. laboratory) as moderators to examine whether the studies provided commensurable measurements across different conditions.

Whether the acoustic properties of caregivers' IDS change according to experimental task and recording environment remains an open question and an important consideration for future studies of IDS (Dunst et al., 2012). A cross-tab plot showing how the acoustic measures were distributed across the conditions of task and environment is shown in Figure S8 in the Supplementary Information. In addition to these moderator analyses, we also conducted sensitivity analyses in order to quantify the robustness of our findings and to assess the evidentiary strength for each acoustic feature in light of publication bias. We computed the worst-case effect size estimate based only on non-affirmative studies and investigated how sensitive the meta-analytic results were to a potential bias for significant results in the field.

2 Methodology and acoustic measures

In order to obtain a comprehensive sample of the available literature on acoustic properties of IDS, we conducted a systematic literature search on PubMed and Web of Science, in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Guidelines (Page et al., 2021) (cf. Figure S1.1 and Tables S1.2-3 in Supplementary Information). The search terms used were ("motherese" OR "baby talk" OR "child-directed speech" OR "infant-directed speech" OR "caretaker speech" OR "parentese"), with no search limits in the query in order to target studies broadly. The first systematic search was conducted independently by two of the authors (RF & EF) in June 2017 and updated by a third author (CC) in December 2021; CC screened for missed studies from before and after the date of the first systematic search. Disagreements in the screening of papers were resolved with discussions in the first phase between EF and RF and in the second phase between CC and RF; if the paper was thought to contain relevant data for the meta-analysis (see below), the paper was included for the successive phase of the review. Disagreements were therefore rare and mainly motivated by studies where relevant information was reported only in the Supplementary Information. As of December 2021, the search strategy yielded a total of 602 papers, which were manually screened for inclusion according to the following criteria: i) infants had to be typically-developing, ii) studies had to include quantification of an acoustic feature, iii) studies had to include a comparison condition with ADS, iv) the speech had to be spoken to an infant by one or both of their primary caregivers.

Based on the initial set of 602 papers, we used Connected Papers and Research Rabbit to find an additional 48 relevant studies. After excluding 54 duplicate studies, we screened the titles of 596 studies and excluded a further 302 studies that were unrelated to the current investigation. We read the abstracts of the remaining 294 studies and evaluated each with reference to the above exclusion criteria. Of the 294 papers, 175 studies had no relation to IDS, 17 studies had no comparison condition with ADS, and 15 studies examined atypical populations and had no relevant control sample of typically-developing infants to extract data on. We discuss the importance of future studies investigating more diverse speaker characteristics further in Section 4.2. To the best of our knowledge, the present review of a total of 88 studies represents a comprehensive sample of the literature on IDS.

In order to assess the state of the literature and to explore the extent to which the studies build a common discourse with reciprocal references, we used the R package *bibliometrix* (Aria & Cuccurullo, 2017) to build coupling and direct-citation networks of the studies, as shown in Figure S2.1 in Supplementary Information. The studies cluster together into three main groups and exhibit considerable overlap in the studies they cite. Further, they cite each other somewhat independently of the acoustic measure reported. The collection of studies investigated here, then, represents a coherent intersection of papers that builds a common discourse on a variety of relevant aspects of IDS.

2.1. Data Extraction

The following meta-analyses allowed us to explore how each acoustic measure differed across infant ages, languages, experimental tasks, and recording environments. We classed the 84 relevant papers into 5 clusters based on the acoustic measure reported: i) f_o, ii) f_o variability, iii) vowel space area, iv) articulation rate, and v) vowel duration. If an individual study reported multiple acoustic measures, the study was included in all of the relevant clusters. It should be noted that other acoustic measures of IDS were reported in some of the studies under investigation (e.g., syllable duration (3 studies), pause duration (5 studies), and intensity (5 studies)); however, the studies provided insufficient data for meta-analysis.

In order to standardize the measures and to allow for comparison among the studies, we calculated Hedges' *g*, an effect size variant that is preferred for small sample sizes (Morris, 2000). For our purposes, this effect size represents the standardized mean difference between ADS and IDS; that is, the bigger the effect size, the larger the difference between the speech styles. A positive effect size indicates that the value for IDS is greater than that for ADS, and vice versa.

This implies that an acoustic property of IDS that becomes more similar to ADS over the course of development would manifest as a shift towards an effect size of zero.

When the raw means and standard deviations were reported in the papers, we calculated the effect sizes with standard formulae for Hedges' g (i.e., $hedges'g = \frac{mean_1 - mean_2}{SD_{pooled}}$), where $SD_{pooled} = \frac{(n_1-1)SD_1^2+(n_2-1)SD_2^2}{n_1+n_2-2}$, as formulated in (Hedges, 1981) where the standard deviation of each group is weighted by its sample size, using the R package *esc* (Lüdecke, 2019). For the remaining studies that did not report the raw data, the effect sizes were calculated either by using the reported *d*-values, one-sample *t*-values, or by digitally extracting the raw data from published plots using the WebPlotDigitizer application (Rohatgi, 2014). In certain cases, the standard deviation of the effect size could not be calculated from the reported data or plots. In order to include these effect sizes in the meta-analysis, these missing standard deviation values (n = 110) were imputed by using multivariate imputation by chained equations based on a Bayesian linear regression model in the R package *mice* (Van Buuren & Groothuis-Oudshoorn, 2011), as described further in S3 of the Supplementary Information. We checked that this process of multiple imputation did not bias the estimation of the overall effect size for each acoustic measure by comparing the estimates of the intercepts-only models for the imputed and non-imputed datasets. The results of these analyses are shown in Table S3.1 in Supplementary Information. All hierarchical Bayesian models in this paper pool the results of analyses performed on the imputed datasets. In Table S11.1 in Supplementary Information (cf., Figure 1 as well), we provide more information about the size of the sample investigated for each language. The raw data and code will be made available on MetaLab upon publication of this manuscript (https://osf.io/hc7me/?view_only=93ca379395414057b8704cb23aab4ded).

2.2. Hierarchical Bayesian model

In the following meta-analyses of five acoustic features, we combined the weighted results of comparable studies and provided a pooled estimate of the overall effect sizes. We estimated and adjusted for heterogeneity in population samples and methodologies by allowing the estimate to vary by study. The hierarchical structure of the random-effects model posits that the true effect size may be study-specific and thereby accounts for repeated measures (Hedges & Olkin, 1985; Raudenbush & Bryk, 1985; Fernández-Castilla et al., 2020).The credible interval of the pooled estimate thus aggregates information from both within-study sampling error and between-study variance (Hedges & Vevea, 1998). The hierarchical Bayesian robust regression models were fitted to the meta-analytic data using a Student's t-likelihood. With this type of robust regression model, longer-tailed distributions are implemented in order to reduce the influence of outliers. This method incorporates outliers without allowing them to dominate non-outlier data (Jylänki et al., 2011). See S5 in the Supplementary Information for a detailed account of the models and choice of priors (S5.1), prior and posterior predictive checks (S5.2), prior-posterior update plots (S5.3) as well as prior sensitivity analyses for the model estimates (S5.4.1) and evidence ratios (S5.4.2) of intercept and age.

2.3. Moderator Analyses

We began by building intercepts-only models in order to condition the data for each of the acoustic measures on the variance associated with individual studies. With this model, we posited that effect sizes were nested within languages and within studies. In order to quantify the within-language variability due to different studies reporting data on the same language and repeated measures within these studies, we included nested effects of study and measures within the random-effects term (i.e., (1 | Language/StudySite/measurement)). We used these three-level intercepts-only models to assess the within-language, between-study heterogeneity and report how the effect size estimates of each study deviate from the pooled effect size estimate (cf. Figures S6.1-6.5 in Supplementary Information).

We then constructed a second model to analyze the influence of potential moderators on the variation of effect sizes across studies. This second model allowed us to explore the effect of the following predictors on each of the acoustic measures: i) infant age, ii) language, iii) experimental task and iv) recording environment, the justifications for which were described in section 1.5. We refer to this second model as the full model for the remainder of this paper.

We performed pairwise leave-one-out-information criterion-based model comparison (Vehtari et al., 2017) between the full model and models without each of the predictor variables. We report leave-one-out (loo) stacking weights (Yao et al., 2018) in favor of the model. Stacking weights indicate the probability that the model including the variables is better than the model without the predictor variables.

All computations were performed in R 4.2.0 (R Core Team, 2020) using brms 2.17 (Bürkner, 2017) and Stan 2.21 (Carpenter et al., 2017) in R Studio 1.4 (RStudio Team, 2020).

For each acoustic measure, we provide the estimates from the full model and report 95% credible intervals, evidence ratios, credibility scores and loo stacking weights for each of the models. Each will be described in turn: i) credible intervals (henceforth CrI) refer to the range of values within which there is a 95% probability that the true value of the parameter is included given the assumptions of the model, ii) the evidence ratio provides the ratio of likelihood in favor of a hypothesis; that is, an evidence ratio of 5 indicates that the hypothesis is 5 times more likely than the alternative, while an evidence ratio of ‘Inf’ (infinite) suggests that all of the posterior samples are compatible with the hypothesis and not with the alternatives (Bürkner, 2017), iii) the credibility score refers to the percentage of posterior samples in the direction of the hypothesis under investigation (Bürkner, 2017), and lastly, iv) stacking weight refers to the probability that the model including a predictor provides a better model of the data than the model without the predictor (Vehtari et al., 2017). The estimates from the best model for each acoustic variable are reported in Tables S8.1-8.5 in the Supplementary Information.

We chose to assess publication bias by conducting quantitative sensitivity analyses and estimating the severity of the publication bias required to attenuate the credible interval of the pooled effect size to include null values (Mathur & VanderWeele, 2020). Traditional assessments of publication bias rely on spearman rank correlations between effect size and standard error and exhibit certain limitations. These traditional methods, for example, provide binary decisions either rejecting the null hypothesis of no publication bias or not and fail to control for Type I error rates when used with standardized mean difference effect sizes and conventional variance estimates (Jin et al., 2015; McShane et al., 2016). This is especially the case when within-study sample sizes are relatively small or between-study heterogeneity is high (Pustejovsky & Rodgers, 2019). We therefore chose to assess how robust the meta-analytic estimates would be to varying assumptions of publication bias (Mathur & VanderWeele, 2020). These methods assume that meta-analytic studies represent samples from an underlying (possible) population of published and unpublished studies, where the probability of selection for significant studies is higher. The potential presence of publication bias is thereby assessed i) by varying assumptions as to how much more likely significant studies are to be published than non-significant studies, and ii) by calculating the amount of publication bias required to attenuate the estimates so that the evidence in favor of an effect becomes negligible. This method has limitations, such as relaxing certain distributional assumptions on the population effects and assuming that the non-significant findings available are representative of the whole population of unpublished studies (Mathur & VanderWeele, 2020), However, the method still offers substantial benefits over classical funnel plot methods and selection models (cf., Jin et al., 2015; McShane et al., 2016; Pustejovsky & Rodgers, 2019 for reviews). It should be noted that this method of analyzing publication bias sensitivity cannot comment on the severity of publication bias in practice, nor the opposite; rather, this analysis provides results that allow us to assess the extent to which an effect would be present even if publication bias were a severe issue in the literature. For each acoustic measure, we report the worst-case effect size estimate based solely on the non-significant studies and make sensitivity plots and significance funnel plots (cf. S10.1-10.2 in Supplementary Information).

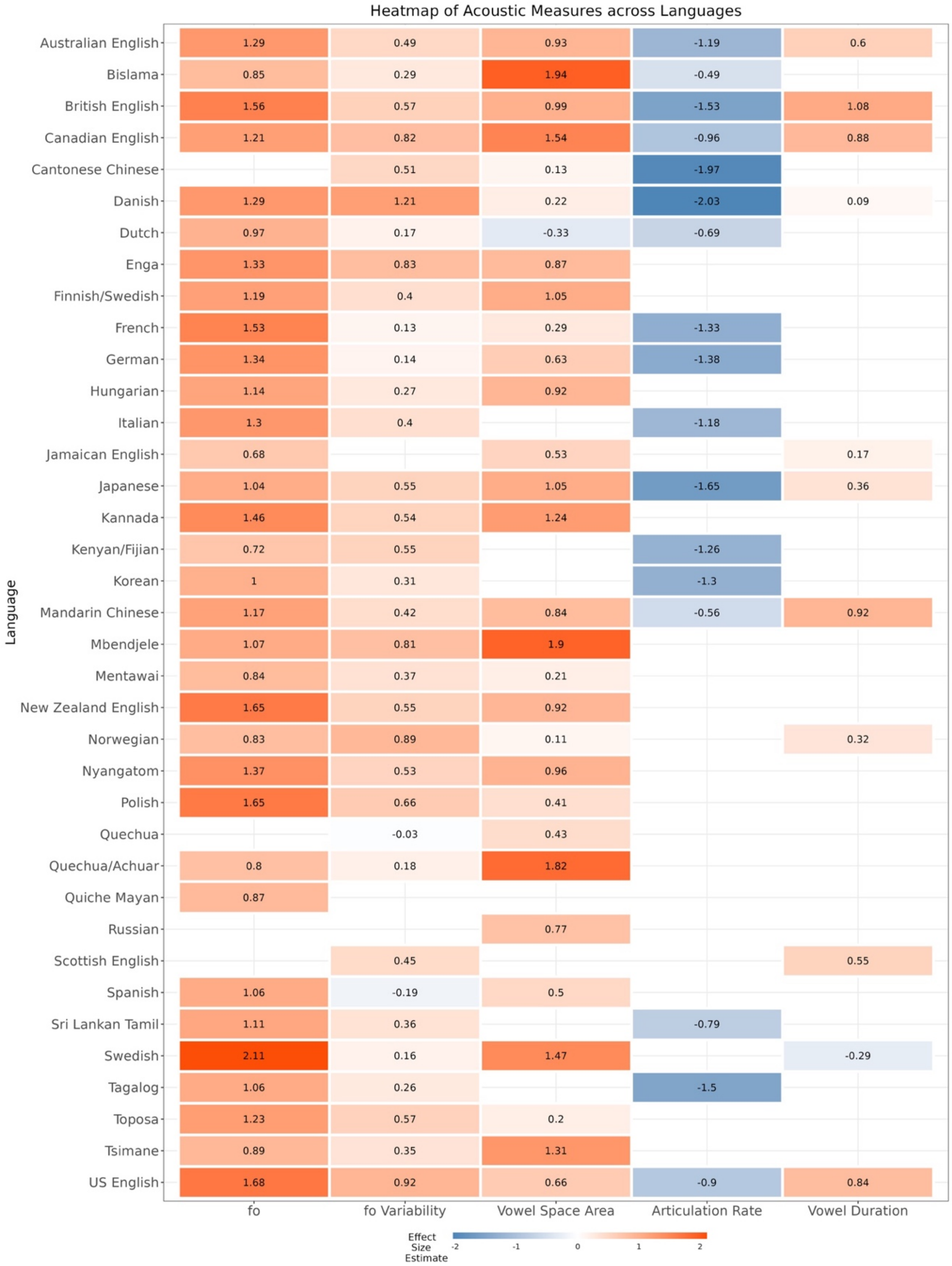
3 Results of Meta-Analyses

3.1 Summary of Results

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

Acoustic Feature	No. of Studies	No. of ES	Average Effect Size	ER	Study SD	Robust Predictors
f_0	60	262	1.19 [0.81; 1.58]	Inf	0.91 [0.72; 1.14]	Language, Age, Task, Environment
f_0 Variability	44	202	0.46 [0.21; 0.71]	817.18	0.76 [0.60; 0.95]	Language, Task
Vowel Space Area	33	84	0.81 [0.44; 1.16]	1799	0.61 [0.41; 0.86]	Language
Articulation Rate	17	60	-1.11 [-1.80; -0.39]	390.3	0.74 [0.42; 1.19]	Language, Age, Task
Vowel Duration	26	81	0.51 [0.16; 0.86]	67.7	0.50 [0.12; 0.92]	Language, Age

Table 1. A summary of the results for the best models for each acoustic variable. ES refers to Effect Sizes. ER refers to Evidence Ratio. Inf means that all posterior samples are in the direction of the hypothesis. The average effect size refers to the average effect size across infant ages and languages in the best model for the acoustic measure.



Language	fo	fo Variability	Vowel Space Area	Articulation Rate	Vowel Duration
Australian English	1.29	0.49	0.93	-1.19	0.6
Bislama	0.85	0.29	1.94	-0.49	
British English	1.56	0.57	0.99	-1.53	1.08
Canadian English	1.21	0.82	1.54	-0.96	0.88
Cantonese Chinese		0.51	0.13	-1.97	
Danish	1.29	1.21	0.22	-2.03	0.09
Dutch	0.97	0.17	-0.33	-0.69	
Enga	1.33	0.83	0.87		
Finnish/Swedish	1.19	0.4	1.05		
French	1.53	0.13	0.29	-1.33	
German	1.34	0.14	0.63	-1.38	
Hungarian	1.14	0.27	0.92		
Italian	1.3	0.4		-1.18	
Jamaican English	0.68		0.53		0.17
Japanese	1.04	0.55	1.05	-1.65	0.36
Kannada	1.46	0.54	1.24		
Kenyan/Fijian	0.72	0.55		-1.26	
Korean	1	0.31		-1.3	
Mandarin Chinese	1.17	0.42	0.84	-0.56	0.92
Mbendjele	1.07	0.81	1.9		
Mentawai	0.84	0.37	0.21		
New Zealand English	1.65	0.55	0.92		
Norwegian	0.83	0.89	0.11		0.32
Nyangatom	1.37	0.53	0.96		
Polish	1.65	0.66	0.41		
Quechua		-0.03	0.43		
Quechua/Achuar	0.8	0.18	1.82		
Quiche Mayan	0.87				
Russian			0.77		
Scottish English		0.45			0.55
Spanish	1.06	-0.19	0.5		
Sri Lankan Tamil	1.11	0.36		-0.79	
Swedish	2.11	0.16	1.47		-0.29
Tagalog	1.06	0.26		-1.5	
Toposa	1.23	0.57	0.2		
Tsimane	0.89	0.35	1.31		
US English	1.68	0.92	0.66	-0.9	0.84

Extended Data Figure 1: A heatmap providing an overview of the effect size estimates for each of the acoustic variables and languages. Dark orange shading indicates a strong effect size value on the positive scale. Dark blue shading indicates a strong effect on the negative scale.

3.2. Fundamental frequency (f_o)

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

3.2.1. f_o across studies

The Bayesian hierarchical intercepts-only model of f_o revealed an overall estimated effect size of $g = 1.17$ with 95% CrI of [0.86; 1.45], a between-languages heterogeneity of $g = 0.34$ [0.05; 0.67], a heterogeneity between studies within languages of $g = 0.90$ [0.71; 1.11], as well as a between-measures heterogeneity of $g = 0.07$ [0.00; 0.21]. A standardized mean difference of this size implies that approximately 87.9% of IDS speech samples are expected to exhibit a higher f_o than ADS speech samples. An overview of how the studies varied with respect to the f_o estimate is shown in the forest plot in Figure S6.1 in the Supplementary Information. All of the studies exhibited effect size estimates on the positive scale, with only 16 of the 60 studies including the null in the lower-bound of their credible intervals.

3.2.2. f_0 as a function of language

The estimates from the full model are shown in Figure 3 below. All of the point estimates for the languages under investigation appear in the positive range of effect sizes. The cross-linguistic differences between IDS and ADS in f_0 across languages thus mainly vary according to the extent to which f_0 is higher in IDS relative to ADS (cf. Table S9.1 in the Supplementary Information for language-specific estimates and credible intervals).

3.2.3. f_0 as a function of age

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

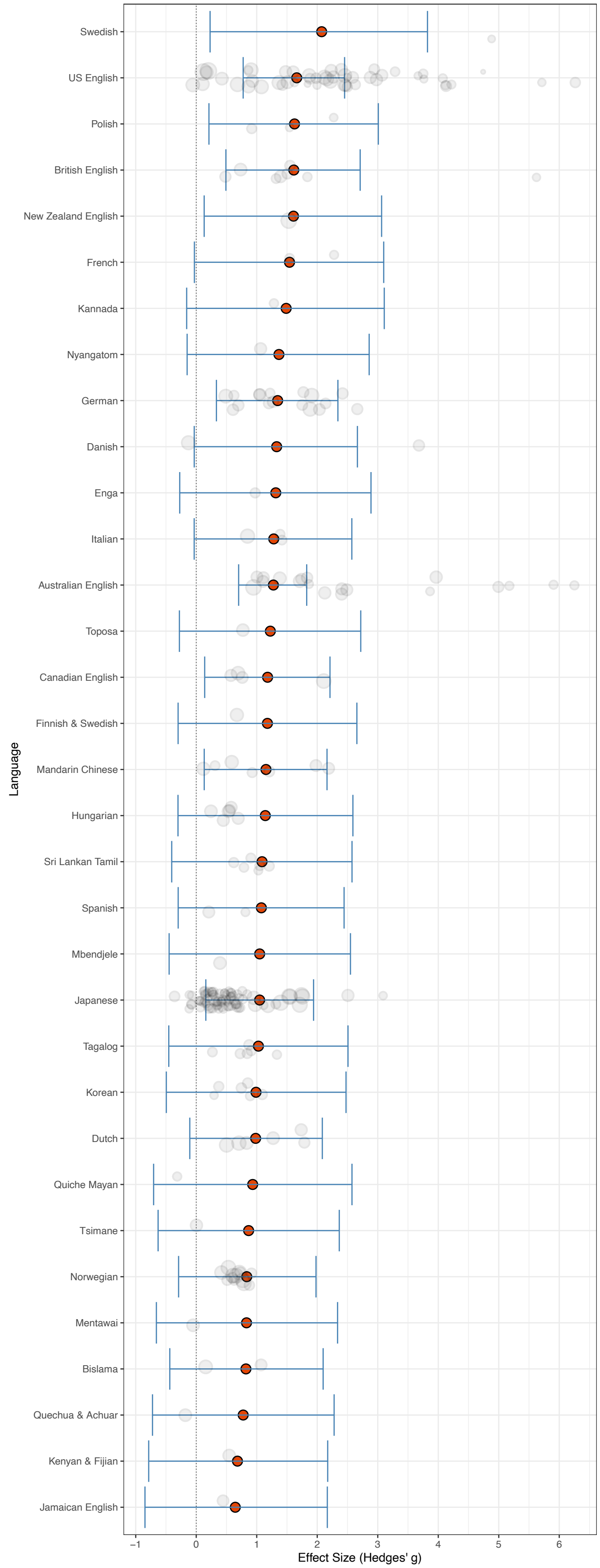
3.2.4: f_0 as a function of task and environment

As shown in the middle-right plot in Figure 3, caregivers produced a greater f_0 difference between the two speech styles in experimental tasks designed to elicit spontaneous speech (estimate: 0.43 with 95% CrI [0.13; 0.74], evidence ratio: 94.54, credibility: 0.99). As shown in the lower-right-hand plot in Figure 3, recording parents in a naturalistic setting as opposed to in the laboratory shows a smaller difference between IDS and ADS in terms of f_o (estimate: -0.48 with 95% CrI [-0.87; -0.07], evidence ratio: 36.54, credibility: 0.97).

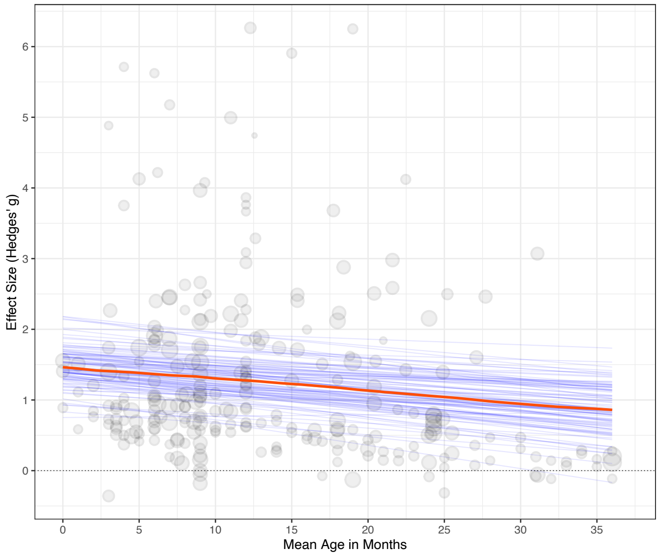
3.2.5. Publication bias for f_o

The sensitivity analysis of publication bias for f_o indicated that no amount of publication bias would be able to attenuate the effect size estimate for the credible interval to include null effects, as depicted in Figure S10.1 in the Supplementary Information. The worst-case effect size estimate based solely on non-significant studies is 0.60 with 95% CrI [0.37; 0.83], as shown in Figure S10.2 in the Supplementary Information. This analysis suggests that the effect size estimates might be quite robust to even severe levels of publication bias - assuming that effect size estimates of non-significant studies are representative of those of unpublished studies.

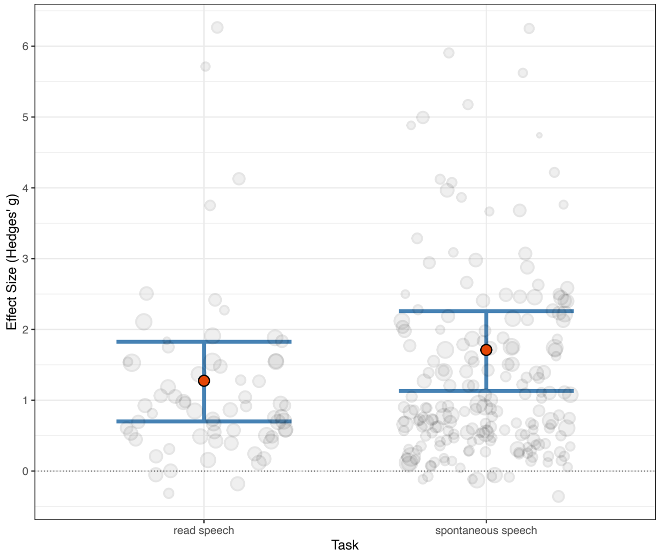
f_o as a Function of Language



f_o as a Function of Age



f_o as a Function of Task



f_o as a Function of Environment

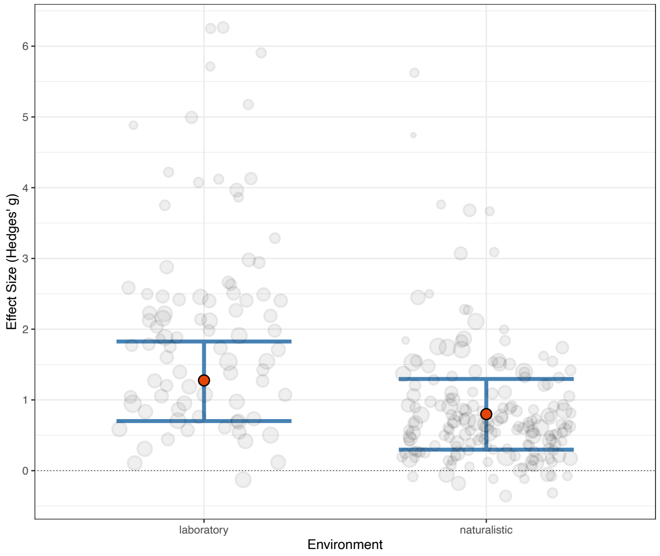


Figure 3. Panel showing model estimates for a total of 3401 participants across 60 studies investigating 33 distinct languages. The panel consists of i) A plot of effect size estimates for f_o *according to language (left). The centres of the error bars (orange points) indicate the posterior effect size estimate for each language pooled across studies. The error bars provide the 95% credible interval and the grey points are the raw effect size data). The size of the points is proportional to the inverse of the standard error of the effect size (i.e., the larger the point, the smaller the standard error). ii) A spaghetti plot showing 100 posterior model predictions for the effect size estimates for* f_o *as a function of age (top right). iii) A plot showing the distribution of effect size estimates across experimental tasks (middle-right). The orange points indicate the posterior effect size estimate for each experimental condition. The error bars provide the 95% credible interval and the grey points are the raw effect size data). The size of the points is proportional to the inverse of the standard error of the effect size (i.e., the larger the point, the smaller the standard error). iv) A plot showing the distribution of effect size estimates across recording environments (bottom-right). The orange points indicate the posterior effect size estimate for each recording condition. The error bars provide the 95% credible interval and the grey points are the raw effect size data). The size of the points is proportional to the inverse of the standard error of the effect size (i.e., the larger the point, the smaller the standard error).*

3.3. f_o variability

Some of the studies reported f_o range (*n*=25) and others reported the standard deviation of f_o (*n*=20). As these measures both capture change in f_o over the course of the utterance, we grouped them together into a single category. If a study reported both measures, we used standard deviation because range consists of the difference between the highest and the lowest value, therefore being highly sensitive to even one outlier or measurement error. Therefore, standard deviation remains less sensitive to extreme values and represents the more reliable measure of the two. The effect size distributions of f_o range and f_o standard deviation were shown to be strongly correlated and exhibit no notable differences, as shown in Figures S4.1-S4.2 in the Supplementary Information. We extracted 223 effect sizes from 44 of the 88 studies. In this context, a positive Hedges' *g* value signifies a higher degree of f_o variability in IDS, and vice versa. The model with task, age, and language as predictors provided a better account of the data (stacking weight: 0.681) than the model including both task and environment (stacking weight: 0.218), the model excluding task (stacking weight: 0.017), and the model excluding both task and environment (stacking weight: 0.084).

3.3.1. f_o variability across studies

The Bayesian hierarchical intercepts-only model of f_o variability showed an overall estimated difference of $g = 0.69$ with 95% CrI [0.44; 0.92] and a between-languages heterogeneity of $g = 0.25$ [0.02; 0.52], a heterogeneity between studies within languages of $g = 0.71$ [0.56; 0.88], as well as a between-measures heterogeneity of $g = 0.11$ [0.01; 0.23]. With a standardized mean difference of this size, this implies that approximately 83% of IDS speech samples would show a higher degree of f_o variability compared to that of ADS speech samples. An overview of how the studies varied with respect to the f_o variability estimate is shown in the forest plot in Figure S6.2 in the Supplementary Information. The estimated effect sizes were primarily distributed on the positive scale, indicating that the studies provided evidence for greater f_o variability in IDS than in ADS. Only one out of the 43 studies on f_o variability had a negative effect size point estimate, which the authors posit could be a result of caregivers' tendency to produce utterances with a higher minimum f_o in IDS, thereby reducing the possible f_o range (Outters et al., 2020).

3.3.2. f_o variability as a function of language

As shown in Figure 4 below, most of the point estimates for the languages appeared to be in the positive range of effect sizes (cf. Table S9.2 in the Supplementary Information for language-specific estimates and credible intervals). The cross-linguistic differences between IDS and ADS in f_0 variability mainly related to the degree of exaggeration.

3.3.3. f_0 variability as a function of age

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

3.3.4. f_0 variability as a function of task and environment

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

3.3.5. Publication bias for f_0 variability

A sensitivity analysis with a random-effects specification indicates that no amount of publication bias would be able to attenuate the effect size estimate for the credible interval to include null effects, as depicted in Figure S10.1 in the Supplementary Information. The uncorrected worst-case estimate for the effect size based solely on non-significant studies is 0.33 with 95% CrI [0.18; 0.47], as shown in in Figure S10.2 in the Supplementary Information.

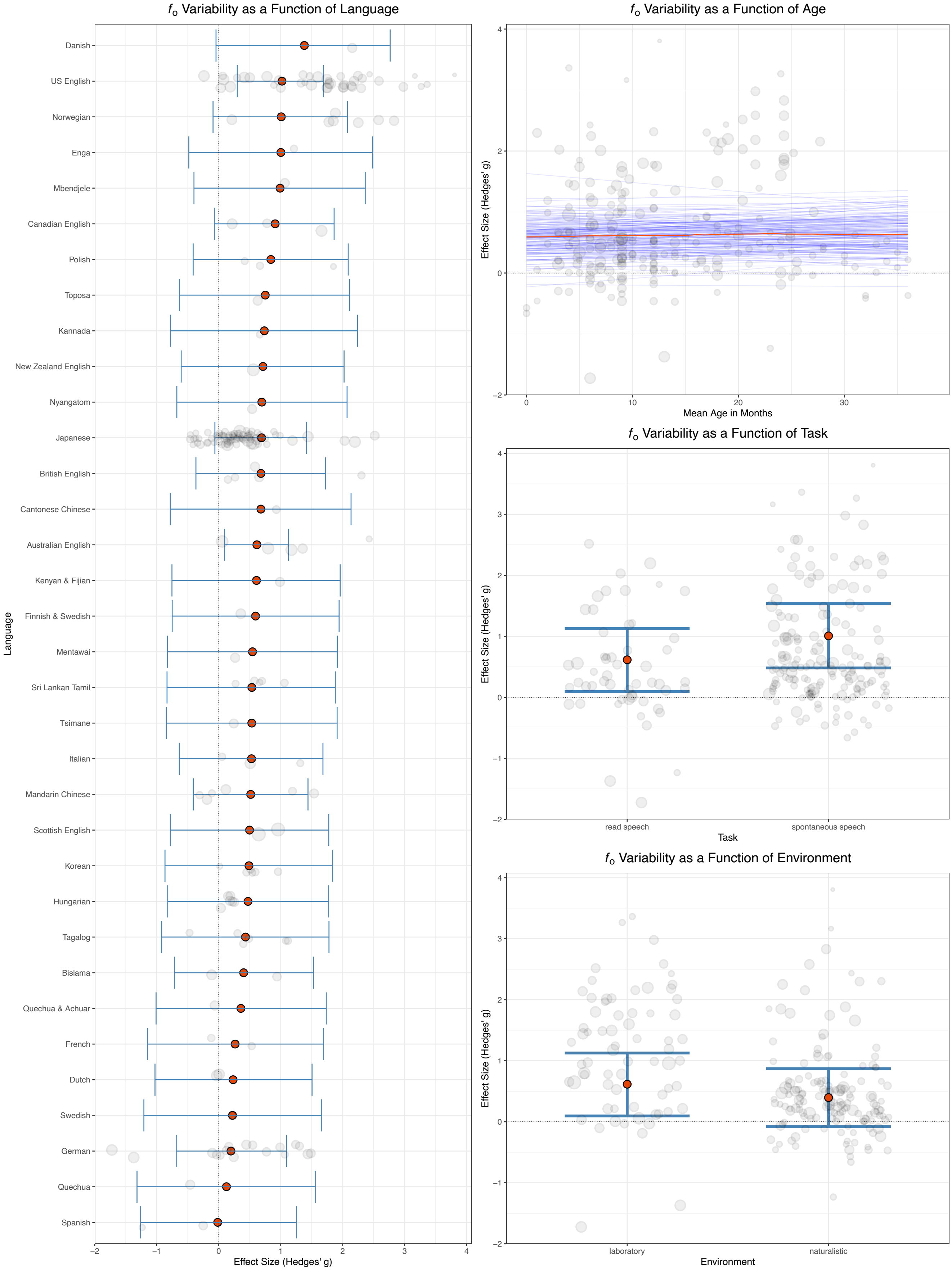

f_o Variability as a Function of Language
Language
Danish
US English
Norwegian
Enga
Mbendjele
Canadian English
Polish
Toposa
Kannada
New Zealand English
Nyangatom
Japanese
British English
Cantonese Chinese
Australian English
Kenyan & Fijian
Finnish & Swedish
Mentawai
Sri Lankan Tamil
Tsimane
Italian
Mandarin Chinese
Scottish English
Korean
Hungarian
Tagalog
Bislama
Quechua & Achuar
French
Dutch
Swedish
German
Quechua
Spanish
−2
−1
0
1
2
3
4
Effect Size (Hedges' g)
f_o Variability as a Function of Age
Effect Size (Hedges' g)
4
2
0
−2
0
10
20
30
Mean Age in Months
f_o Variability as a Function of Task
Effect Size (Hedges' g)
4
3
2
1
0
−1
−2
read speech
spontaneous speech
Task
f_o Variability as a Function of Environment
Effect Size (Hedges' g)
4
3
2
1
0
−1
−2
laboratory
naturalistic
Environment

Figure 4. Panel showing model estimates for a total of 3006 participants across 44 studies investigating 34 distinct languages. The panel consists of i) A plot of effect size estimates for f_o variability according to language (left). The centres of the error bars (orange points) indicate the posterior effect size estimate for each language pooled across studies. The error bars provide the 95% credible interval and the grey points are the raw effect size data). The size of the points is proportional to the inverse of the standard error of the effect size (i.e., the larger the point, the smaller the standard error). ii) A spaghetti plot showing 100 posterior model predictions for the effect size estimates for f_o variability as a function of age (top right), iii) A plot showing the distribution of effect size estimates across experimental tasks (middle-right). The orange points indicate the posterior effect size estimate for each experimental condition. The error bars provide the 95% credible interval and the grey points are the raw effect size data). The size of the points is proportional to the inverse of the standard error of the effect size (i.e., the larger the point, the smaller the standard error). iv) A plot showing the distribution of effect size estimates across recording environments (bottom-right). The orange points indicate the posterior effect size estimate for each recording condition. The error bars provide the 95% credible interval and the grey points are the raw effect size data). The size of the points is proportional to the inverse of the standard error of the effect size (i.e., the larger the point, the smaller the standard error).

3.4. Vowel space area

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

3.4.1. Vowel space area across studies

The Bayesian hierarchical intercepts-only model of vowel space area showed an overall estimated difference in vowel space area of $g = 0.66$ with 95% CrI of [0.34; 0.98], a between-languages heterogeneity of $g = 0.55$ [0.12; 0.97], a heterogeneity between studies within languages of $g = 0.66$ [0.43; 0.92], as well as a between-measures heterogeneity of $g = 0.11$ [0.00; 0.28]. A standardized mean difference of this size implies that approximately 74% of IDS speech samples overall will show an expanded vowel space area compared to those of ADS speech samples. An overview of how the studies varied with respect to the vowel space area estimate is shown in the forest plot in Figure S6.3 in the Supplementary Information. The studies were generally distributed across positive effect sizes; however, 19 of the 32 studies included the null in the lower bound of their credible intervals, and 2 of the 32 studies provided evidence for the opposite effect, namely that ADS exhibited an expanded vowel space area compared to IDS (Benders, 2013; Steen & Englund, 2021). The pooling of data from these studies on vowel space area, then, indicated a moderate effect size, with some of the studies providing conflicting results, possibly due to cross-linguistic differences, as discussed further below and in section 4.1.

3.4.2. Vowel space area as a function of language

As shown in Figure 5, most of the point estimates for the languages appeared to be in the positive range of effect sizes (cf. Table S9.3 in the Supplementary Information for language-specific estimates and credible intervals); however, there appears to be substantial cross-linguistic variation in the extent to which caregivers expand their vowel space area when speaking to infants.

3.4.3. Vowel space area as a function of age

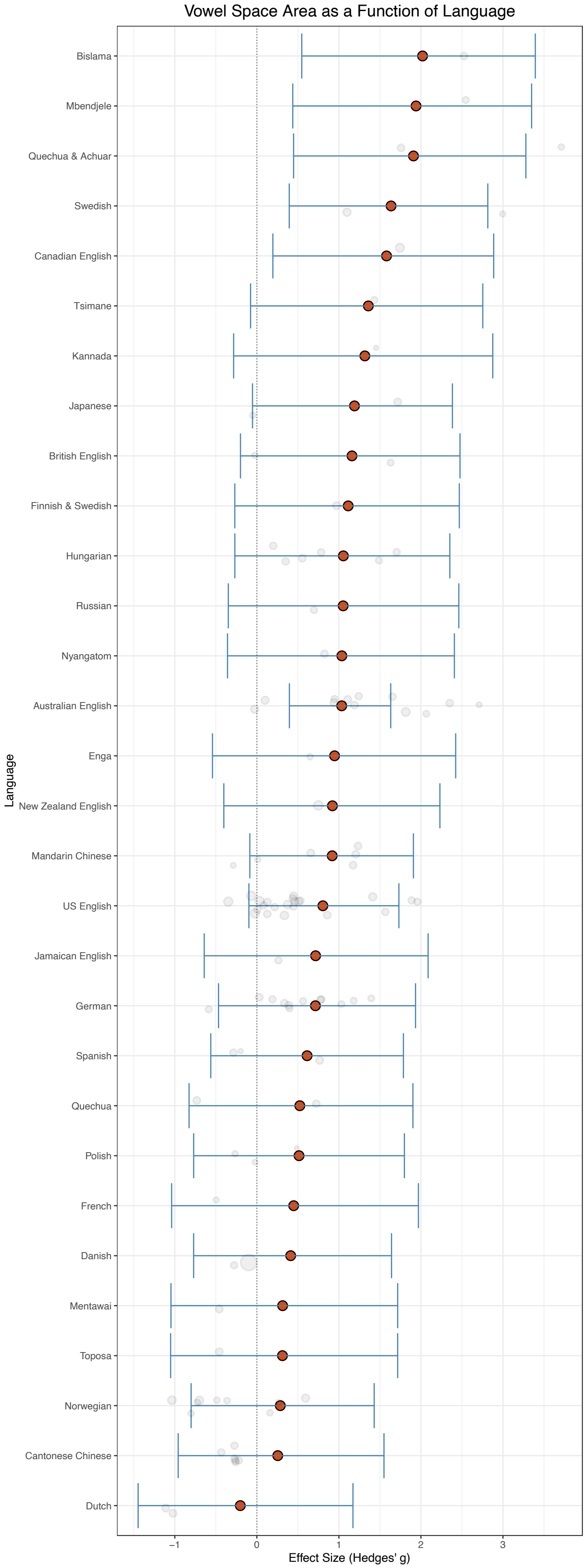
As shown in the top-right of Figure 5, the model indicated no evidence for an effect of infant age. The estimate is -0.00 with 95% CrI [-0.02; 0.01], evidence ratio: 2.04, credibility: 0.66.

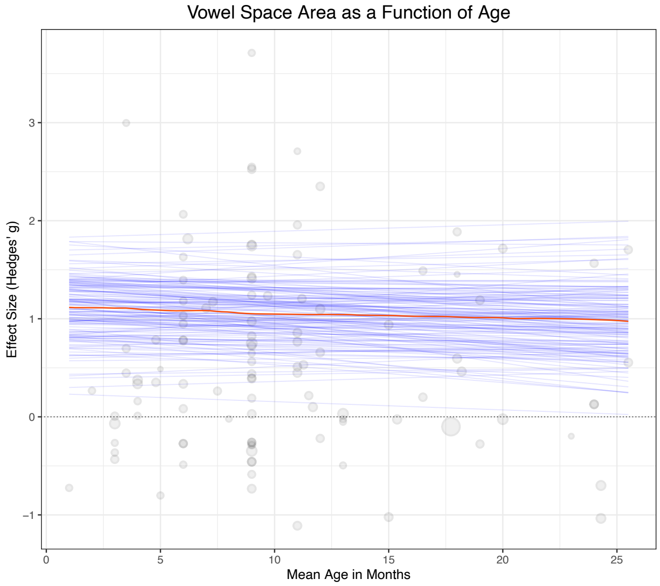
3.4.4. Vowel space as a function of task and environment

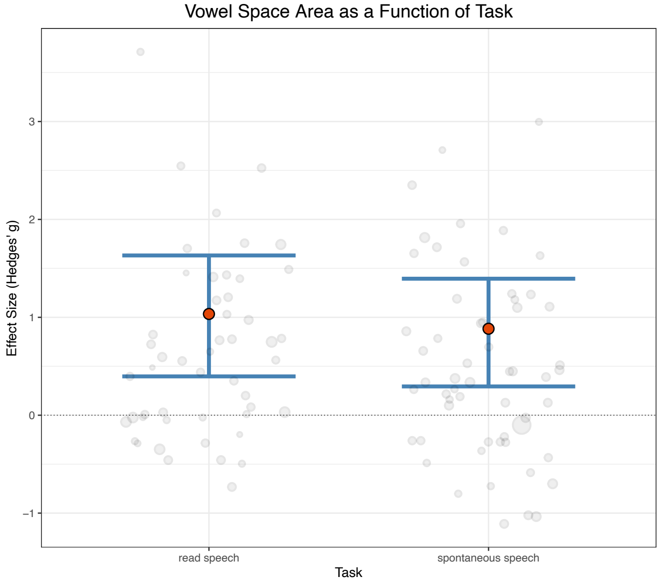
As shown in the middle-right plot in Figure 5, caregivers did not seem to credibly produce a greater vowel space area in the experimental task of producing spontaneous speech (estimate: -0.16 with 95% CrI [-0.49; 0.18], evidence ratio: 3.58, credibility: 0.78). Similarly, as shown in the lower-right-hand plot in Figure 5, recording caregivers with their infants in a naturalistic setting did not appear to exert an effect on the vowel space area of caregivers' IDS (estimate: -0.27 with 95% CrI [-0.76; 0.23], evidence ratio: 4.29, credibility: 0.81).

3.4.5. Publication bias for vowel space area

A sensitivity analysis with a random-effects specification indicated that if moderate publication bias were present in the literature, then the effect size estimate may be closer to null effects. That is, if significant results were fourfold more likely to be published in the literature, the credible interval would include an effect size of 0.1, as shown in in Figure S10.1 in the Supplementary Information. The uncorrected worst-case estimate for the effect size based solely on non-significant studies is 0.20 with 95% CrI [-0.01; 0.42], as shown in in Figure S10.2 in the Supplementary Information.


Vowel Space Area as a Function of Language


Vowel Space Area as a Function of Age


Vowel Space Area as a Function of Task

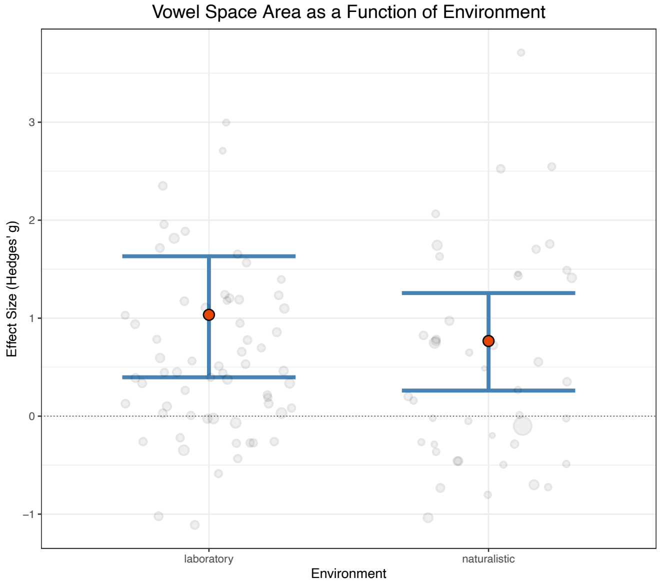

Vowel Space Area as a Function of Environment

Figure 5. Panel showing model estimates for a total of 1702 participants across 33 studies investigating 30 distinct languages. The panel consists of i) A plot of effect size estimates for vowel space area according to language (left). The centres of the error bars (orange points) indicate the posterior effect size estimate for each language pooled across studies. The error bars provide the 95% credible interval and the grey points are the raw effect size data). The size of the points is proportional to the inverse of the standard error of the effect size (i.e., the larger the point, the smaller the standard error). ii) A spaghetti plot showing 100 posterior model predictions for the effect size estimates for vowel space area as a function of age (top right), iii) A plot showing the distribution of effect size estimates across experimental tasks (middle-right). The orange points indicate the posterior effect size estimate for each experimental condition. The error bars provide the 95% credible interval and the grey points are the raw effect size data). The size of the points is proportional to the inverse of the standard error of the effect size (i.e., the larger the point, the smaller the standard error). iv) A plot showing the distribution of effect size estimates across recording environments (bottom-right). The orange points indicate the posterior effect size estimate for each recording condition. The error bars provide the 95% credible interval and the grey points are the raw effect size data). The size of the points is proportional to the inverse of the standard error of the effect size (i.e., the larger the point, the smaller the standard error).

3.5. Articulation rate

Speech production rate is generally measured in one of two ways: articulation rate excludes pause intervals, but speech rate includes them and consequently takes speaker-specific ways of conveying information (e.g., hesitations and pauses) into account (Laver & John, 1994; Tsao et al., 2006a, 2006b). The majority of studies under investigation here (15 out of 17) reported articulation rate as opposed to speech rate. Because both of these measures capture similar acoustic information (i.e., the number of output units per unit of time), we have combined the measures in our meta-analysis. But the distinction between them should be made theoretically because a slower speech rate may signify factors in addition to a slower articulation rate (e.g., the number and duration of silent pauses) (Laver, 1994). Here, we use articulation rate to refer to this combination of measures.

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

3.5.1. Articulation rate across studies

The Bayesian hierarchical intercepts-only model of articulation rate showed an overall estimated difference of *g* = -1.03 with 95% CrI of [-1.53; -0.56] and a between-languages heterogeneity of *g* = 0.38 [0.02; 1.00], a heterogeneity between studies within languages of *g* = 0.80 [0.50; 1.20], as well as a heterogeneity between measurements of *g* = 0.26 [0.04; 0.47]. With a standardized mean difference of this size, this implies that approximately 85% of IDS speech samples will show a slower rate compared to that of ADS speech samples. An overview of how the studies varied with respect to articulation rate estimate is shown in the forest plot in Figure S6.4 in the Supplementary Information. The estimated effect sizes of the studies are distributed primarily on the negative scale, indicating that caregivers on average speak slower in IDS than in ADS; however, due to the relative sparsity of data for this acoustic measure, many of the languages include null effects in their credible intervals.

3.5.2. Articulation rate as a function of language

As shown on the left-hand side of Figure 6, all of the effect size point estimates for the languages under investigation appeared in the negative range (cf. in Table S9.4 in the Supplementary Information for language-specific estimates and credible intervals).

3.5.3. Articulation rate as a function of age

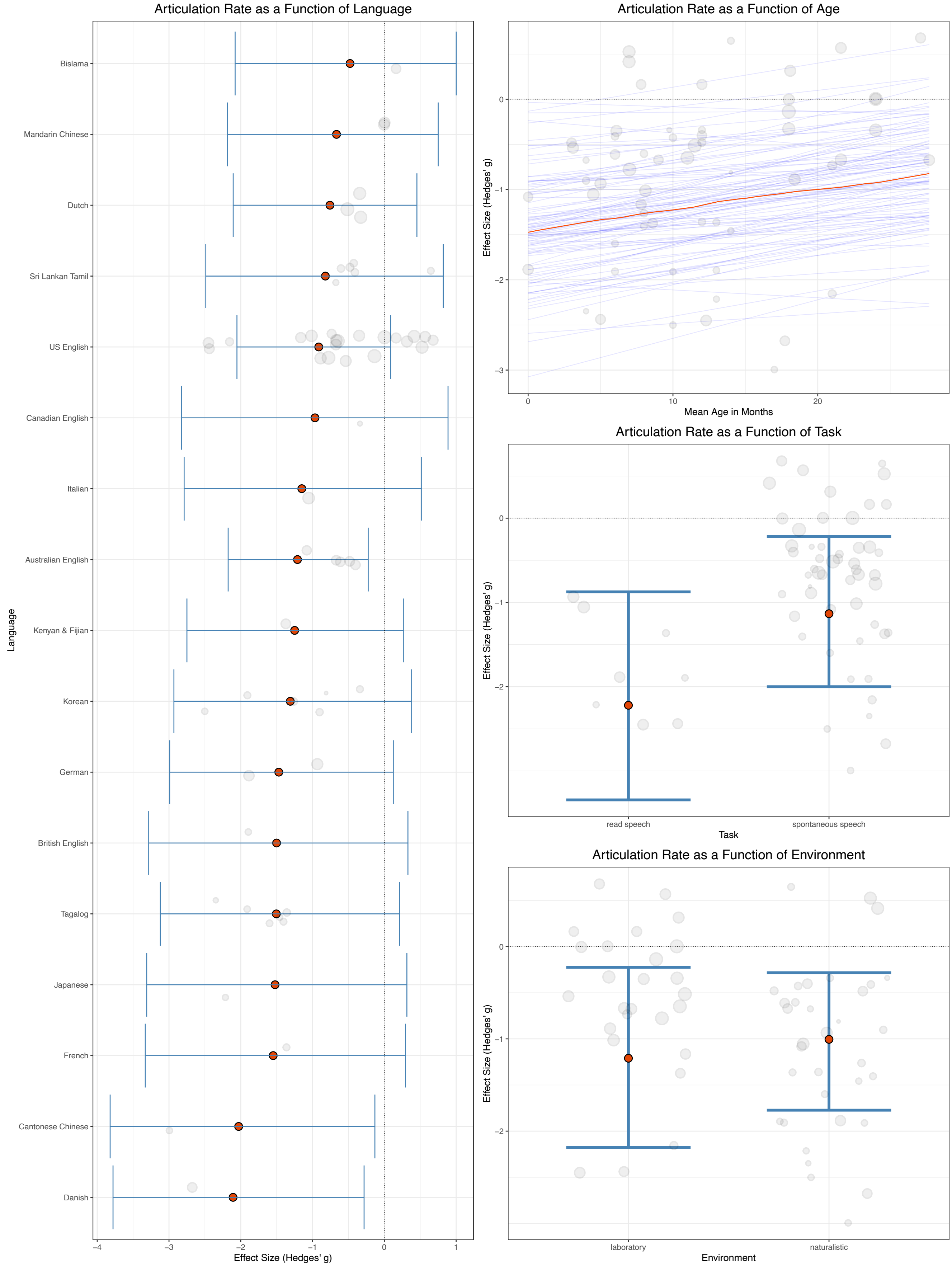
As shown in the top-right of Figure 6, the model indicated a reliable effect of infant age. The estimate for the effect of age is 0.02 with 95% CrI [0.00; 0.05], evidence ratio: 33.33, credibility: 0.97. This result shows that caregivers' articulation rate in IDS becomes more similar to ADS over the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

3.5.4. Articulation rate as a function task and environment

As shown in the middle-right plot in Figure 6, caregivers appeared to speak faster to their infants in spontaneous speech than in read speech (estimate: 0.95 with 95% CrI [0.1; 1.73], evidence ratio: 28.34, credibility: 0.97. Conversely, the lower-right-hand plot in Figure 6 indicates that there is no evidence that recording infants outside of the laboratory affects the articulation rate in caregivers' IDS (estimate: 0.15 with 95% CrI [-0.71; 0.96], evidence ratio: 1.66, credibility: 0.62).s

3.5.5. Publication bias for articulation rate

A sensitivity analysis with a random-effects specification indicated that no amount of publication bias would be able to attenuate the estimate to null, as shown in Figure S10.1 in the Supplementary Information. If moderate publication bias were present in the literature, then the effect size estimate may represent a more moderate effect; the uncorrected worst-case estimate for the effect size based solely on non-significant studies is -0.445 with 95% CrI [-0.757; -0.133], as shown in Figure S10.2 in the Supplementary Information.



Articulation Rate as a Function of Language

Articulation Rate as a Function of Age

Articulation Rate as a Function of Task

Articulation Rate as a Function of Environment

Figure 6. Panel showing model estimates for a total of 976 participants across 17 studies investigating 17 distinct languages. The panel consists of i) A plot of effect size estimates for articulation rate according to language (left). The centres of the error bars (orange points) indicate the posterior effect size estimate for each language pooled across studies. The error bars provide the 95% credible interval and the grey points are the raw effect size data). The size of the points is proportional to the inverse of the standard error of the effect size (i.e., the larger the point, the smaller the standard error). ii) A spaghetti plot showing 100 posterior model predictions for the effect size estimates for articulation rate as a function of age (top right), iii) A plot showing the distribution of effect size estimates across experimental tasks (middle-right). The orange points indicate the posterior effect size estimate for each experimental condition. The error bars provide the 95% credible interval and the grey points are the raw effect size data). The size of the points is proportional to the inverse of the standard error of the effect size (i.e., the larger the point, the smaller the standard error). iv) A plot showing the distribution of effect size estimates across recording environments (bottom-right). The orange points indicate the posterior effect size estimate for each recording condition. The error bars provide the 95% credible interval and the grey points are the raw effect size data). The size of the points is proportional to the inverse of the standard error of the effect size (i.e., the larger the point, the smaller the standard error).

3.6. Vowel duration

The acoustic measure of vowel duration was analyzed in 26 of the 88 studies, and 81 effect sizes were extracted from these studies. We should note that the vowel categories for which data were available differed markedly across studies, with some studies reporting vowel duration only for the articulatory extremes of /i/, /a/ and /u/ (e.g., Lovcevic et al., 2020; Steen & Englund, 2021) and others reporting vowel duration for the full set of vowel phonemes in their language (e.g., Englund, 2018). In this context, a positive Hedges' *g* value signifies a longer vowel duration in IDS compared to that in ADS, and vice versa. The model with age and language as predictors was shown to provide a better account of the data (stacking weight: 0.393) than the model including task and environment (0.154), the model including task (stacking weight: 0.242) and the model including environment (stacking weight: 0.211).

3.6.1. Vowel duration across studies

The Bayesian hierarchical intercepts-only model of vowel duration showed an overall estimated difference of $g = 0.48$ with 95% CrI of [0.08; 0.88], a between-languages heterogeneity of $g = 0.38$ [0.03; 0.92], a heterogeneity between studies within languages of $g = 0.43$ [0.06; 0.85], as well as a between-measures heterogeneity of $g = 0.17$ [0.01; 0.38]. With a standardized mean difference of this size, this implies that approximately 70% of IDS speech samples will show a longer vowel duration to that of ADS speech samples. An overview of how the studies varied with respect to the vowel duration estimate is shown in the forest plot in Figure S6.5 in the Supplementary Information. The majority of the effect size estimates were distributed on the positive scale, indicating that the studies show that caregivers produce vowels with a longer duration in IDS than in ADS.

3.6.2. Vowel duration as a function of language

As shown in Figure 7, most of the effect size estimates for the languages under investigation appeared in the positive range (cf. in Table S9.5 in the Supplementary Information for language-specific estimates and credible intervals); however, there appears to be an influence of language-specific phonological properties, as some languages exhibit substantially longer vowel durations in IDS (e.g., Mandarin Chinese), mixed results (e.g., US English and Japanese), while others indicate no durational differences between the speech styles (e.g., Swedish, Norwegian, and Danish).

3.6.3. Vowel duration as a function of age

As shown in the top-right of Figure 7, the model indicated a moderate effect of infant age. The estimate for the effect of age is -0.02 with 95% CrI [-0.05; 0.01], evidence ratio: 6.48, credibility: 0.87. This suggests that caregivers' vowel durations in IDS became slightly more similar to ADS as infants got older.

3.6.4: Vowel duration as a function of task and environment

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

3.6.5. Publication bias for vowel duration

A sensitivity analysis with a random-effects specification indicated that no amount of publication bias can attenuate the estimate to 0.1, as shown in the sensitivity plot of in Figure S10.1 in the Supplementary Information. The uncorrected worst-case estimate for the effect size based solely on non-significant studies is 0.277 with 95% CrI [0.134; 0.417], as shown in Figure S10.2 in the Supplementary Information.

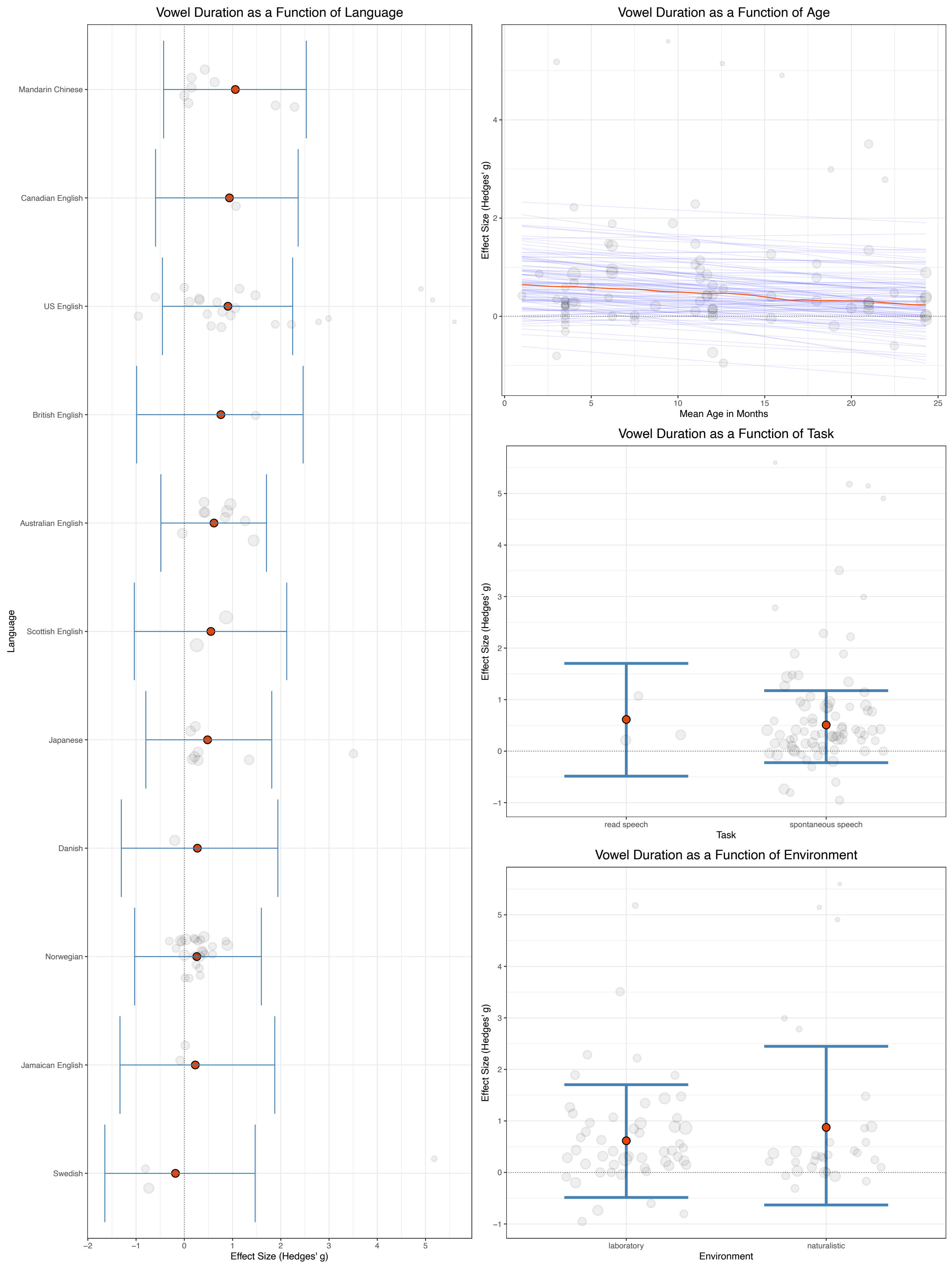

Vowel Duration as a Function of Language
Language
Mandarin Chinese
Canadian English
US English
British English
Australian English
Scottish English
Japanese
Danish
Norwegian
Jamaican English
Swedish
Effect Size (Hedges' g)
−2
−1
0
1
2
3
4
5
Vowel Duration as a Function of Age
Effect Size (Hedges' g)
4
2
0
Mean Age in Months
0
5
10
15
20
25
Vowel Duration as a Function of Task
Effect Size (Hedges' g)
5
4
3
2
1
0
−1
read speech
spontaneous speech
Task
Vowel Duration as a Function of Environment
Effect Size (Hedges' g)
5
4
3
2
1
0
−1
laboratory
naturalistic
Environment

Figure 7. Panel showing model estimates for a total of 1411 participants across 26 studies investigating 11 distinct languages. The panel consists of i) A plot of effect size estimates for vowel duration according to language (left). The centres of the error bars (orange points) indicate the posterior effect size estimate for each language pooled across studies. The error bars provide the 95% credible interval and the grey points are the raw effect size data). The size of the points is proportional to the inverse of the standard error of the effect size (i.e., the larger the point, the smaller the standard error). ii) A spaghetti plot showing 100 posterior model predictions for the effect size estimates for vowel duration as a function of age (top right), iii) A plot showing the distribution of effect size estimates across experimental tasks (middle-right). The orange points indicate the posterior effect size estimate for each experimental condition. The error bars provide the 95% credible interval and the grey points are the raw effect size data). The size of the points is proportional to the inverse of the standard error of the effect size (i.e., the larger the point, the smaller the standard error). iv) A plot showing the distribution of effect size estimates across recording environments (bottom-right). The orange points indicate the posterior effect size estimate for each recording condition. The error bars provide the 95% credible interval and the grey points are the raw effect size data). The size of the points is proportional to the inverse of the standard error of the effect size (i.e., the larger the point, the smaller the standard error).

Discussion

The tendency for caregivers to modify their speech to infants represents a widespread cross-cultural and cross-linguistic phenomenon. The aims of this meta-analysis (see section 1.5) were to examine how the acoustic properties of IDS i) change over the course of early infant development, ii) vary across languages, and iii) differ according to experimental task and recording environment, with an eye towards a better understanding of culturally widespread IDS communicative functions. The results confirmed that across multiple languages and cultures, IDS contains acoustic features that are distinct from ADS, and that different acoustic features operate on varying timescales, as described further in section 4.1. Our analysis of publication bias showed that the pattern of acoustic features in IDS would remain reliable even if a strong bias for significant results existed in the literature (although potentially with the exception of vowel space area, cf. Figure S10.1-10.2). The findings thus provide reliable evidence that caregivers across multiple languages produce IDS with a higher f_o, a higher degree of f_o variability, an expanded vowel space area, a slower articulation rate, and a longer vowel duration, as summarized in Figures 3-7 and Table 1 (cf. also S8.1-8.5 in the Supplementary Information). The analyses, however, also suggested a high degree of unexplained between-study and between-language heterogeneity, as discussed further in section 4.2. Our analyses of moderators indicated that f_o, articulation rate, and vowel duration became more similar to ADS over the course of infants' early development, while vowel space area and f_o variability remained stable, at least up to 25 and 36 months of age, respectively. Our analysis of the effect of experimental task revealed that spontaneous speech displayed greater differences in f_o, articulation rate, and f_o variability between ADS and IDS, compared to read speech. Recording environment likewise showed a reliable influence on the estimates for f_o.

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

4.1. Changes in IDS acoustic features and their relation to functions

The tendency for some of the acoustic features of IDS to change over the course of early development may be due to a form-functions relationship between caregivers' acoustic production patterns and infants' attentional allocation to certain aspects of the speech stream (Kitamura & Notley, 2009; McRoberts et al., 2009; Panneton et al., 2006; Segal & Newman, 2015). For example, the increase in articulation rate and parallel decrease in vowel duration during development may reflect caregivers' sensitivity to infants' improved processing of the speech stream. Articulation rate exhibits robustness across languages (cf., Figure 6), with a universal tendency for caregivers to slow down their speech to infants. Slowed IDS likely eases the cognitive load involved in young infants' speech and language processing (e.g., Christiansen & Chater, 2016; Saffran & Kirkham, 2018; Werker & Tees, 1999). Similarly, the decrease in the utterance-global measure of f_0 in IDS may be a consequence of infants' changing preferences to attend to this acoustic feature in the speech stream (Panneton et al., 2006). Younger infants have been shown to prefer to attend to the positive affect of IDS (Kitamura & Burnham, 1998; Singh et al., 2002), while older infants prefer aspects of the speech stream that provide less positive affect and more linguistically relevant information (Kitamura & Notley, 2009; McRoberts et al., 2009; Segal & Newman, 2015). Vocal pitch exhibited a high degree of robustness across languages (cf., Figure 3), supporting the notion that it is a highly salient property of IDS (e.g., Fernald, 1989; Stern et al., 1983) and that caregivers adjust IDS acoustic properties in ways that suit infants' developmental needs (Fusaroli et al., 2019; Smith & Trainor, 2008). Similarly, the cross-linguistic tendency for the acoustic properties of f_0 variability and vowel space area to remain stable throughout early infancy (cf. Figure S6.6) suggests ongoing developmental relevance (Peter et al., 2016; Song et al., 2010). We should note, however, that vowel space area exhibited cross-linguistic variation (cf. Figure 5), with some of the studies reporting reduced vowel separability in IDS (e.g., Benders, 2013; Englund & Behne, 2005; Rattanasone et al., 2013; Steen & Englund, 2021). Both acoustic features have been implicated in facilitating language development (Fernald & Kuhl, 1987; Hartman et al., 2017; Liu et al., 2003; Spinelli et al., 2017), but whether the benefits of IDS derive mainly from its capacity to direct infants' attention or to emphasize linguistic aspects of the speech stream (or both) remains as an important open question. We should also note that although infant age appears to exert an effect on some of the acoustic measures, the amount of available data across different age ranges varies, ranging from 0-25 months for vowel duration to 0-36 for f_o and f_o variability (cf. Figure S6.6 in Supplementary Information). These results highlight the need for an expansion in the availability of data with a high density of observations across many different age ranges, as discussed further in section 4.3.

Computational evidence indicates that vowel space expansion can aid speech intelligibility (De Boer & Kuhl, 2003; Eaves et al., 2016; McMurray et al., 2009; Vallabha et al., 2007), but beyond considerations of the information content in the speech signal (Golinkoff et al., 2015; Kalashnikova et al., 2017), the benefits may simply be a product of the social qualities of IDS, which facilitate learning through increased infant attention (ManyBabies Consortium, 2020; Werker & McLeod, 1989) and social motivation (Goldstein & Schwade, 2008; Singh et al., 2002). The question of how specific acoustic properties in IDS may facilitate aspects of infant development could be pursued with more detailed theory-driven studies of languages with distinct linguistic systems, as discussed further in section 4.3.

4.2. Unexplained variability across studies and languages

Our meta-analytic models revealed a substantial amount of between-study heterogeneity for each of the acoustic features, especially among the studies reporting measures of f_o, f_o variability and articulation rate (cf. Table 1). Some between-study heterogeneity is expected simply from random sampling error and the mathematics of estimating an effect across a large number of studies (Mikolajewicz & Komarova, 2019; Song et al., 2001). But some of this unexplained variance may derive from the inclusion of studies that differ from one another in meaningful ways, such as in study designs, population sample characteristics, cross-linguistic diversity, and experimental methodologies (Ruppar, 2020). For example, our results indicated larger differences between the speech styles in f_o, f_o variability, and articulation rate for studies recording parents' spontaneous speech as opposed to read speech (cf. Figures 3, 4, and 6). Without a complete characterization of the sources of this unexplained heterogeneity, factors influencing the generalizability of the effects remain undetermined and therefore constitute an important avenue for future research.

One source of heterogeneity could be the variability induced by cross-linguistic differences in IDS. The acoustic features of IDS were shown to vary across languages, many of which relied on a small number of datapoints and studies, and therefore exhibited substantial uncertainty (cf. S8 in Supplementary Information). Part of this heterogeneity and cross-linguistic uncertainty may also depend on the variability caused by subtle differences in phonological systems across languages. For example, although our results suggest a strong cross-linguistic tendency for caregivers to produce IDS with an overall slower articulation rate, Church et al. (2005) found that the difference in articulation rate between Canadian English ADS and IDS to 8.5- and 11-month-old infants disappeared when utterance-final syllables were excluded, due to the phonological tendency for utterance-final syllables to be lengthened in Canadian English (cf. Martin et al., 2016 for similar results for Japanese). Similarly, substantial differences in the number and category of vowels included in our analysis of vowel duration may influence the generalizability of results in languages with other types of vowel inventories and phonological systems. Determining the influence of subtle cross-linguistic differences, such as prosodic phonology, as well as vowel inventories and phonemes, will be a fruitful area for future investigations. Although we were unable to accommodate these types of subtle phonological differences between languages in our analyses, these sources of variability highlight the need for fine-grained, theory-driven comparisons of the acoustic properties of IDS across different languages and population characteristics (e.g., gender and ethnicity) as well as careful consideration of the causal mechanisms involved (Christiansen et al., 2022; Deffner et al., 2021; Trecca et al., 2021).

Another source of the between-study heterogeneity may come from intra-study participant characteristics. Low sample sizes and tight experimental controls characteristic of infant research may result in outcomes that are idiosyncratic to particular study conditions (Song et al., 2001). Between-study differences in participant characteristics, such as gender and kinship, are thus likely to function as potential sources of unexplained heterogeneity. For example, the high prevalence of post-partum depression (Gavin et al., 2005; Gelaye et al., 2016) and its attested effects on the prosodic properties of IDS (Kaplan et al., 2001; Lam-Cassettari & Kohlhoff, 2020; Porritt et al., 2014) may affect the generalizability of the current results to these population samples. The developmental status of the infant, moreover, may also function as a potential source of heterogeneity in IDS properties, as caregivers have been shown to respond differently according to the developmental status of the infant (Fusaroli et al., 2019; Nguyen et al., 2022; Woolard et al., 2022). Future research exploring the effects of diverse speaker characteristics, such as depression, kinship, gender and infants' developmental status, would provide important insights into factors affecting the acoustic properties of IDS.

In order to allow for more fine-grained temporal analyses of how acoustic features of IDS manifest themselves across early infancy, and to further explore sources of between-study variability, we encourage researchers to share participant-level data in open repositories. A cumulative approach to improving the external validity of studies can also be carried out by conducting experiments across multiple laboratories (e.g., ManyBabies Consortium, 2020), affording the exploration of within-lab and between-lab variability. Because logistical constraints may hinder multi-laboratory approaches, we argue that providing access to participant-level data may represent the easiest, most practical alternative.

Despite the finding of substantial between-study heterogeneity, we should emphasize that the studies exhibited consistency with each other; that is, the credible intervals for the results of individual studies showed substantial overlap (cf., Figures S6.1-6.5). Moreover, our meta-analytic models included random effects by study to address the dependency among effect sizes as well as predictor variables to explain the heterogeneity between studies. In the following section, we provide a series of recommendations that will enable a better understanding of the factors moderating the acoustic properties of caregivers' IDS.

4.3. Recommendations for Future Research

While solid progress has been made toward examining a wide variety of relevant aspects of IDS, we have identified various shortcomings that should be addressed in future investigations.

In the following, we provide several suggestions. First, with the continued rise of day-long recordings (e.g., Xu et al., 2009) and open archives of acoustic and phonetically transcribed data (e.g., MacWhinney, 2014), as well as the continued development of techniques to automatically assess and code large amounts of audio data (e.g., Cychosz et al., 2021; Räsänen et al., 2021), future research can provide an expansion in the availability of cross-linguistic data and a high density of observations for each participant (e.g., Le Franc et al., 2018). These technological developments will allow for a more fine-grained resolution and comparison of how IDS differs across individuals, languages and infant ages. Second, as noted above, to further explore the functions and learnability afforded by IDS, more theory-driven comparisons across distinct linguistic systems are needed (Christiansen et al., 2022; Trecca et al., 2021), as well as testable predictions from computational models disentangling different theoretical accounts. For example, computational models that explore the supposed learnability afforded by the acoustic properties of IDS constitute fruitful future avenues of research (e.g., Adriaans & Swingley, 2012; Ludusan et al., 2021; Vallabha et al., 2007) as do computational models of stimulus-driven attention and prominence of IDS (Räsänen et al., 2018) and other sensory inputs more generally (Kidd et al., 2012, 2014). Assessing these models on data from a broad range of cultural, linguistic and sociodemographic settings would provide a more robust assessment of theoretical limitations and provide fuel for further theoretical development. Finally, adapting speech to a listener is not a unilateral phenomenon. We want to highlight the importance of considering the mutual feedback loops between infant and caregiver, with infants being an important source of information regarding which sort of signal would be most beneficial for their developmental progress (Goldstein & Schwade, 2008; Ko et al., 2016; Murray & Trevarthen, 1986; Nguyen et al., 2022; Warlaumont et al., 2014). This is especially important given the substantial variability in developmental trajectories across individuals. Studies investigating the importance of the bidirectional process of adaptation between infants' communicative signals and caregiver responsiveness on a turn-by-turn basis comprise another fruitful avenue of future work that can deliver new accounts, predictions, and data from both interactants' viewpoints (Englund & Behne, 2005; Fusaroli et al., 2019; Golinkoff et al., 2015; Kalashnikova & Burnham, 2018; Lam & Kitamura, 2010, 2012; Nguyen et al., 2022).

5 Conclusion

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

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

6 List of References

Adriaans, F., & Swingley, D. (2012). *Distributional learning of vowel categories is supported by prosody in infant-directed speech*. *34*(34). https://escholarship.org/uc/item/7pt904fz

Amano, S., Nakatani, T., & Kondo, T. (2006). Fundamental frequency of infants' and parents' utterances in longitudinal recordings. *The Journal of the Acoustical Society of America*, *119*(3), 1636–1647. https://doi.org/10.1121/1.2161443

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Bainbridge, C. M., Bertolo, M., Youngers, J., Atwood, S., Yurdum, L., Simson, J., Lopez, K., Xing, F., Martin, A., & Mehr, S. A. (2021). Infants relax in response to unfamiliar foreign lullabies. *Nature Human Behaviour*, *5*(2), 256–264. https://doi.org/10.1038/s41562-020-00963-z

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

Bryant, G. A. (2022). Vocal communication across cultures: Theoretical and methodological issues. *Philosophical Transactions of the Royal Society B*, *377*(1841), 20200387. https://doi.org/10.1098/rstb.2020.0387

Bryant, G. A., & Barrett, H. C. (2007). Recognizing intentions in infant-directed speech: Evidence for universals. *Psychological Science*, *18*(8), 746–751. https://doi.org/10.1111/j.1467-9280.2007.01970.x

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

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., 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

Casillas, M., Brown, P., & Levinson, S. C. (2020). Early language experience in a Tseltal Mayan village. *Child Development*, *91*(5), 1819–1835. https://doi.org/10.1111/cdev.13349

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*, E62. https://doi.org/10.1017/S0140525X1500031X

Christiansen, M. H., Kallens, P. C., & Trecca, F. (2022). Towards A Comparative Approach to Language Acquisition. *Current Directions in Psychological Science*, *31*, 131–138. https://doi.org/10.1177/09637214211049229

Church, R., Bernhardt, B., Pichora-Fuller, K., & Shi, R. (2005). Infant-directed speech: Final syllable lengthening and rate of speech. *Canadian Acoustics*, *33*(4), 13–19.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61*(5), 1584–1595. https://doi.org/10.1111/j.1467-8624.1990.tb02885.x

Cristia, A. (2010). Phonetic enhancement of sibilants in infant-directed speech. *The Journal of the Acoustical Society of America*, *128*(1), 424–434. https://doi.org/10.1121/1.3436529

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

Cychosz, M., Cristia, A., Bergelson, E., Casillas, M., Baudet, G., Warlaumont, A. S., Scaff, C., Yankowitz, L., & Seidl, A. (2021). Vocal development in a large-scale crosslinguistic corpus. *Developmental Science*, e13090. https://doi.org/10.1111/desc.13090

De Boer, B., & Kuhl, P. K. (2003). Investigating the role of infant-directed speech with a computer model. *Acoustics Research Letters Online*, *4*(4), 129–134. https://doi.org/10.1121/1.1613311

Deffner, D., Rohrer, J. M., & McElreath, R. (2021). A Causal Framework for Cross-Cultural Generalizability. *Advances in Methods and Practices in Psychological Science*, *5*(3). https://doi.org/10.1177/25152459221106366

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).

Dunst, C., Gorman, E., & Hamby, D. (2012). Preference for infant-directed speech in preverbal young children. *Center for Early Literacy Learning*, *5*(1), 1–13.

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

Englund, K. T. (2018). Hypoarticulation in infant-directed speech. *Applied Psycholinguistics*, *39*(1), 67–87.

Englund, K. T., & Behne, D. M. (2005). Infant directed speech in natural interaction—Norwegian vowel quantity and quality. *Journal of Psycholinguistic Research*, *34*(3), 259–280. https://doi.org/10.1007/s10936-005-3640-7

Ferguson, C. A. (1964). Baby talk in six languages. *American Anthropologist*, *66*(6_PART2), 103–114.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, 1497–1510. https://doi.org/10.2307/1130938

Fernald, A. (1992). Meaningful melodies in mothers' speech to infants. In *Nonverbal vocal communication: Comparative and developmental approaches* (pp. 262–282). Cambridge University Press.

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, *10*(3), 279–293.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology*, *27*(2), 209.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, *20*(1), 104–113. https://doi.org/10.1037/0012-1649.20.1.104

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

Fernández-Castilla, B., Jamshidi, L., Declercq, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. (2020). The application of meta-analytic (multi-level) models with multiple random effects: A systematic review. *Behavior Research Methods*, *52*(5). https://doi.org/10.3758/s13428-020-01373-9

Floccia, C., Keren-Portnoy, T., DePaolis, R., Duffy, H., Delle Luche, C., Durrant, S., White, L., Goslin, J., & Vihman, M. (2016). British English infants segment words only with exaggerated infant-directed speech stimuli. *Cognition*, *148*, 1–9. https://doi.org/10.1016/j.cognition.2015.12.004

Fusaroli, R., Lambrechts, A., Bang, D., Bowler, D. M., & Gaigg, S. B. (2017). Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis. *Autism Research*, *10*(3), 384–407.

Fusaroli, R., Weed, E., Fein, D., & Naigles, L. (2019). Hearing me hearing you: Reciprocal effects between child and parent language in autism and typical development. *Cognition*, *183*, 1–18. https://doi.org/10.1016/j.cognition.2018.10.022

Gavin, N. I., Gaynes, B. N., Lohr, K. N., Meltzer-Brody, S., Gartlehner, G., & Swinson, T. (2005). Perinatal depression: A systematic review of prevalence and incidence. *Obstetrics & Gynecology*, *106*(5 Part 1), 1071–1083. https://doi.org/10.1097/01.AOG.0000183597.31630.db

Gelaye, B., Rondon, M. B., Araya, R., & Williams, M. A. (2016). Epidemiology of maternal depression, risk factors, and child outcomes in low-income and middle-income countries. *The Lancet Psychiatry*, *3*(10), 973–982.

Gelman, A., Simpson, D., & Betancourt, M. (2017). The prior can often only be understood in the context of the likelihood. *Entropy*, *19*(10), 555.

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), 1–10. https://doi.org/10.1038/s41598-017-13883-2

Gleitman, L. R., Newport, E. L., & Gleitman, H. (1984). The current status of the motherese hypothesis. *Journal of Child Language*, *11*(1), 43–79. https://doi.org/10.1017/S0305000900005584

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19*(5), 515–523. https://doi.org/10.1111/j.1467-9280.2008.02117.x

Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby) talk to me: The social context of infant-directed speech and its effects on early language acquisition. *Current Directions in Psychological Science*, *24*(5), 339–344. https://doi.org/10.1177/0963721415595345

Grieser, D. L., & Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology*, *24*(1), 14–20. https://doi.org/10.1037/0012-1649.24.1.14

Han, M., De Jong, N., & Kager, R. (2020). Pitch properties of infant-directed speech specific to word-learning contexts: A cross-linguistic investigation of Mandarin Chinese and Dutch. *Journal of Child Language*, *47*(1), 85–111. https://doi.org/10.1017/S0305000919000813

Hartman, K. M. (2013). Patterns and possible influences of maternal vowel clarification on child language development. *ProQuest Dissertations Publishing*.

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

Haspelmath, M. (2009). The typological database of the World Atlas of Language Structures. In *The Use of Databases in Cross-Linguistic Studies* (pp. 283–300). De Gruyter Mouton.

Hayashi, A., Tamekawa, Y., & Kiritani, S. (2001). Developmental change in auditory preferences for speech stimuli in Japanese infants. *Journal of Speech, Language, and Hearing Research*, *44*(6), 1189–1200. https://doi.org/10.1044/1092-4388

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

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic press.

Hedges, L. V., & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological Methods*, *3*(4), 486.

Hilton, C. B., Moser, C. J., Bertolo, M., Lee-Rubin, H., Amir, D., Bainbridge, C. M., Simson, J., Knox, D., Glowacki, L., Galbarczyk, A., Jasienska, G., Ross, C. T., Neff, M. B., Martin, A., Cirelli, L. K., Trehub, S. E., Song, J., Kim, M., Schachner, A., … Mehr, S. A. (2022). Acoustic regularities in infant-directed speech and song across cultures. *Nature Human Behaviour*, 1–12. https://doi.org/10.1038/s41562-022-01410-x

Huettig, F., & Guerra, E. (2019). Effects of speech rate, preview time of visual context, and participant instructions reveal strong limits on prediction in language processing. *Brain Research*, *1706*, 196–208. https://doi.org/10.1016/j.brainres.2018.11.013

Jin, Z., Zhou, X., & He, J. (2015). Statistical methods for dealing with publication bias in meta-analysis. *Statistics in Medicine*, *34*(2), 343–360. https://doi.org/10.1002/sim.6342

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*, 170306. https://doi.org/10.1098/rsos.170306

Kaplan, P. S., Bachorowski, J., Smoski, M. J., & Zinser, M. (2001). Role of clinical diagnosis and medication use in effects of maternal depression on infant-directed speech. *Infancy*, *2*(4), 537–548. https://doi.org/10.1207/S15327078IN0204_08

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PloS One*, *7*(5), e36399. https://doi.org/10.1371/journal.pone.0036399

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The Goldilocks effect in infant auditory attention. *Child Development*, *85*(5), 1795–1804. https://doi.org/10.1111/cdev.12263

Kitamura, C., & Burnham, D. (1998). *Acoustic and affective qualities of IDS in English*. Fifth international conference on spoken language processing. http://www.isca-speech.org/archive.

Kitamura, C., & Lam, C. (2009). Age-specific preferences for infant-directed affective intent. *Infancy*, *14*(1), 77–100. https://doi.org/10.1080/15250000802569777

Kitamura, C., & Notley, A. (2009). The shift in infant preferences for vowel duration and pitch contour between 6 and 10 months of age. *Developmental Science*, *12*(5), 706–714. https://doi.org/10.1111/j.1467-7687.2009.00818.x

Knoll, M. A., & Costall, A. (2015). Characterising F (0) contour shape in infant-and foreigner-directed speech. *Speech Communication*, *66*, 231–243. https://doi.org/10.1016/j.specom.2014.10.007

Ko, E.-S., Seidl, A., Cristia, A., Reimchen, M., & Soderstrom, M. (2016). Entrainment of prosody in the interaction of mothers with their young children. *Journal of Child Language*, *43*(2), 284–309. https://doi.org/10.1017/S0305000915000410

Kondaurova, M. V., & Bergeson, T. R. (2011). The effects of age and infant hearing status on maternal use of prosodic cues for clause boundaries in speech. *Journal of Speech, Language, and Hearing Research*, *54*(3), 740–754. https://doi.org/10.1044/1092-4388(2010/09-0225)

Kondaurova, M. V., Bergeson, T. R., & Xu, H. (2013). Age-Related Changes in Prosodic Features of Maternal Speech to Prelingually Deaf Infants with Cochlear Implants. *Infancy*, *18*(5), 825–848. https://doi.org/10.1111/infa.12010

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

Lam, C., & Kitamura, C. (2006). *Developmental trends in infant preferences for affective intent in mothers' speech*. 100–105.

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

Lam-Cassettari, C., & Kohlhoff, J. (2020). Effect of maternal depression on infant-directed speech to prelinguistic infants: Implications for language development. *PLoS One*, *15*(7), e0236787.

Laver, J., & John, L. (1994). *Principles of phonetics*. Cambridge university press.

Le Franc, A., Riebling, E., Karadayi, J., Wang, Y., Scaff, C., Metze, F., & Cristia, A. (2018). *The ACLEW DiViMe: An Easy-to-use Diarization Tool.* 1383–1387.

Lee, C. S., Kitamura, C., Burnham, D., & McAngus Todd, N. P. (2014). On the rhythm of infant-versus adult-directed speech in Australian English. *The Journal of the Acoustical Society of America*, *136*(1), 357–365. https://doi.org/10.1121/1.4883479

Lemoine, N. P. (2019). Moving beyond noninformative priors: Why and how to choose weakly informative priors in Bayesian analyses. *Oikos*, *128*(7), 912–928.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, 839–862. https://doi.org/10.2307/411991

Liu, H., Kuhl, P. K., & Tsao, F. (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

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

Lüdecke, D. (2019). R Package 'esc': Effect Size Computation for Meta Analysis. *R Package Version 0.5*. https://strengejacke.github.io/esc/

Ludusan, B., Mazuka, R., & Dupoux, E. (2021). Does Infant-Directed Speech Help Phonetic Learning? A Machine Learning Investigation. *Cognitive Science*, *45*(5), e12946.

MacWhinney, B. (2014). *The CHILDES project: Tools for analyzing talk, Volume II: The database*. Psychology Press. https://doi.org/10.4324/9781315805641

ManyBabies Consortium. (2020). Quantifying sources of variability in infancy research using the infant-directed-speech preference. *Advances in Methods and Practices in Psychological Science*, *3*(1), 24–52. https://doi.org/10.1177/2515245919900809

Martin, A., Igarashi, Y., Jincho, N., & Mazuka, R. (2016). Utterances in infant-directed speech are shorter, not slower. *Cognition*, *156*, 52–59. https://doi.org/10.1016/j.cognition.2016.07.015

Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., & Cristia, A. (2015). Mothers speak less clearly to infants than to adults: A comprehensive test of the hyperarticulation hypothesis. *Psychological Science*, *26*(3), 341–347. https://doi.org/10.1177/0956797614562453

Mathur, M. B., & VanderWeele, T. J. (2020). Sensitivity analysis for publication bias in meta-analyses. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *69*(5), 1091–1119.

McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: Insights from a computational approach. *Developmental Science*, *12*(3), 369–378. https://doi.org/10.1111/j.1467-7687.2009.00822.x

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

McRoberts, G. W., McDonough, C., & Lakusta, L. (2009). The role of verbal repetition in the development of infant speech preferences from 4 to 14 months of age. *Infancy*, *14*(2), 162–194. https://doi.org/10.1080/15250000802707062

McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in meta-analysis: An evaluation of selection methods and some cautionary notes. *Perspectives on Psychological Science*, *11*(5), 730–749. https://doi.org/10.1177/1745691616662243

Mikolajewicz, N., & Komarova, S. V. (2019). Meta-Analytic Methodology for Basic Research: A Practical Guide. *Frontiers in Physiology*, *10*, 203. https://doi.org/10.3389/fphys.2019.00203

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

Morris, S. B. (2000). Distribution of the standardized mean change effect size for meta-analysis on repeated measures. *British Journal of Mathematical and Statistical Psychology*, *53*(1), 17–29. https://doi.org/10.1348/000711000159150

Murray, L., & Trevarthen, C. (1986). The infant's role in mother–infant communications. *Journal of Child Language*, *13*(1), 15–29. https://doi.org/10.1017/S0305000900000271

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

Newman, R. S., & Hussain, I. (2006). Changes in preference for infant-directed speech in low and moderate noise by 4.5-to 13-month-olds. *Infancy*, *10*(1), 61–76. https://doi.org/10.1207/s15327078in1001_4

Nguyen, V., Versyp, O., Cox, C. M. M., & Fusaroli, R. (2022). A systematic review and Bayesian meta-analysis of the development of turn taking in adult-child vocal interactions. *Child Development*. https://doi.org/10.1111/cdev.13754

Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology*, *162*, 31–38. https://doi.org/10.1016/j.jecp.2017.04.017

Niwano, K., & Sugai, K. (2002). Intonation contour of Japanese maternal infant-directed speech and infant vocal response. *The Japanese Journal of Special Education*, *39*(6), 59–68.

Outters, V., Schreiner, M. S., Behne, T., & Mani, N. (2020). Maternal input and infants' response to infant-directed speech. *Infancy*, *25*(4), 478–499. https://doi.org/10.1111/infa.12334

Owren, M. J., & Rendall, D. (2001). Sound on the rebound: Bringing form and function back to the forefront in understanding nonhuman primate vocal signaling. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews*, *10*(2), 58–71.

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. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*, *88*, 105918. https://doi.org/10.1016/j.ijsu.2021.105906

Panneton, R., Kitamura, C., Mattock, K., & Burnham, D. (2006). Slow speech enhances younger but not older infants' perception of vocal emotion. *Research in Human Development*, *3*(1), 7–19. https://doi.org/10.1207/s15427617rhd0301_2

Papoušek, M., Papoušek, H., & Symmes, D. (1991). The meanings of melodies in motherese in tone and stress languages. *Infant Behavior and Development*, *14*(4), 415–440. https://doi.org/10.1016/0163-6383(91)90031-M

Parola, A., Simonsen, A., Bliksted, V., & Fusaroli, R. (2020). Voice patterns in schizophrenia: A systematic review and Bayesian meta-analysis. *Schizophrenia Research*, *216*, 24–40.

Pegg, J. E., Werker, J. F., & McLeod, P. J. (1992). Preference for infant-directed over adult-directed speech: Evidence from 7-week-old infants. *Infant Behavior and Development*, *15*(3), 325–345. https://doi.org/10.1016/0163-6383(92)80003-D

Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn locally, think globally: Exemplar variability supports higher-order generalization and word learning. *Psychological Science*, *21*(12), 1894–1902. https://doi.org/10.1177/0956797610389189

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), 1–14. https://doi.org/10.1038/srep34273

Porritt, L. L., Zinser, M. C., Bachorowski, J.-A., & Kaplan, P. S. (2014). Depression diagnoses and fundamental frequency-based acoustic cues in maternal infant-directed speech. *Language Learning and Development*, *10*(1), 51–67.

Pustejovsky, J. E., & Rodgers, M. A. (2019). Testing for funnel plot asymmetry of standardized mean differences. *Research Synthesis Methods*, *10*(1), 57–71. https://doi.org/10.1002/jrsm.1332

R Core Team. (2020). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.* [Computer software].

Raneri, D. P. (2015). Infant-directed speech: Maternal pitch variability, rate of speech, and child language outcomes. *ProQuest Dissertations Publishing*, 1–56.

Räsänen, O., Kakouros, S., & Soderstrom, M. (2018). Is infant-directed speech interesting because it is surprising?– Linking properties of IDS to statistical learning and attention at the prosodic level. *Cognition*, *178*, 193–206. https://doi.org/10.1016/j.cognition.2018.05.015

Räsänen, O., Seshadri, S., Lavechin, M, Cristia, A., & Casillas, M. (2021). An open-source tool for automatic measurement of phoneme, syllable, and word counts from child-centered daylong recordings. *Behavior Research Methods*, *53*(2), 818–835. https://doi.org/10.3758/s13428-020-01460-x

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

Raudenbush, S. W., & Bryk, A. S. (1985). Empirical Bayes meta-analysis. *Journal of Educational Statistics*, *10*(2), 75–98. https://doi.org/10.3102/10769986010002075

Rohatgi, A. (2014). WebPlotDigitizer user manual version 3.4. *URL Http://Arohatgi. Info/WebPlotDigitizer/App*, 1–18.

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.

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*(2), 339–349. https://doi.org/10.1111/j.1467-7687.2008.00786.x

Rost, G. C., & McMurray, B. (2010). Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning. *Infancy*, *15*(6), 608–635. https://doi.org/10.1111/j.1532-7078.2010.00033.x

RStudio Team. (2020). *RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.* [Computer software].

Ruppar, T. (2020). Meta-analysis: How to quantify and explain heterogeneity? *European Journal of Cardiovascular Nursing*, *19*(7), 646–652. https://doi.org/10.1177/1474515120944014

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology*, *69*, 181–203. https://doi.org/10.1146/annurev-psych-122216-011805

Segal, J., & Newman, R. S. (2015). Infant preferences for structural and prosodic properties of infant-directed speech in the second year of life. *Infancy*, *20*(3), 339–351. https://doi.org/10.1111/infa.12077

Seidl, A., & Cristià, A. (2008). Developmental changes in the weighting of prosodic cues. *Developmental Science*, *11*(4), 596–606. https://doi.org/10.1111/j.1467-7687.2008.00704.x

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

Shute, B., & Wheldall, K. (1999). Fundamental frequency and temporal modifications in the speech of British fathers to their children. *Educational Psychology*, *19*(2), 221–233. https://doi.org/10.1080/0144341990190208

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

Smith, N. A., & Trainor, L. J. (2008). Infant-directed speech is modulated by infant feedback. *Infancy*, *13*(4), 410–420. https://doi.org/10.1080/15250000802188719

Soderstrom, M., Seidl, A., Nelson, D. G. K., & Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language*, *49*(2), 249–267. https://doi.org/10.1016/S0749-596X(03)00024-X

Song, 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

Song, F., Sheldon, T. A., Sutton, A. J., Abrams, K. R., & Jones, D. R. (2001). Methods for Exploring Heterogeneity in Meta-Analysis. *Evaluation & the Health Professions*, *24*(2), 126–151. https://doi.org/10.1177/016327870102400203

Spinelli, M., Fasolo, M., & Mesman, J. (2017). Does prosody make the difference? A meta-analysis on relations between prosodic aspects of infant-directed speech and infant outcomes. *Developmental Review*, *44*, 1–18. https://doi.org/10.1016/j.dr.2016.12.001

Steen, V. B., & Englund, N. (2021). Child-directed Speech in a Norwegian Kindergarten Setting. *Scandinavian Journal of Educational Research*, 1–14. https://doi.org/10.1080/00313831.2021.1897873

Stern, D. N., Spieker, S., Barnett, R., & 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

Trecca, F., Tylén, K., Højen, A., & Christiansen, M. H. (2021). Danish as a Window Onto Language Processing and Learning. *Language Learning*, *71*, 799–833. https://doi.org/10.1111/lang.12450

Tsao, Y.-C., Weismer, G., & Iqbal, K. (2006a). Interspeaker variation in habitual speaking rate: Additional evidence. *Journal of Speech, Language, and Hearing Research*, *49*(5), 1156–1164. https://doi.org/10.1044/1092-4388(2006/083)

Tsao, Y.-C., Weismer, G., & Iqbal, K. (2006b). The effect of intertalker speech rate variation on acoustic vowel space. *The Journal of the Acoustical Society of America*, *119*(2), 1074–1082. https://doi.org/10.1121/1.2149774

Tsuji, S., Bergmann, C., Lewis, M., Braginsky, M., Piccinini, P., Frank, M. C., & Cristia, A. (2017). *MetaLab: A Repository for Meta-Analyses on Language Development, and More.* 2038–2039.

Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences*, *104*(33), 13273–13278. https://doi.org/10.1073/pnas.0705369104

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*, 1–67. https://doi.org/10.18637/jss.v045.i03

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Vosoughi, S., & Roy, D. K. (2012). A longitudinal study of prosodic exaggeration in child-directed speech. *Proceedings of the Speech Prosody, 6th International Conference*.

Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2014). A social feedback loop for speech development and its reduction in autism. *Psychological Science*, *25*(7), 1314–1324. https://doi.org/10.1177/0956797614531023

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

Werker, J. F., & McLeod, P. J. (1989). Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology*, *43*(2), 230. https://doi.org/10.1037/h0084224

Werker, J. F., & Tees, R. C. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology*, *50*(1), 509–535. https://doi.org/10.1146/annurev.psych.50.1.509

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

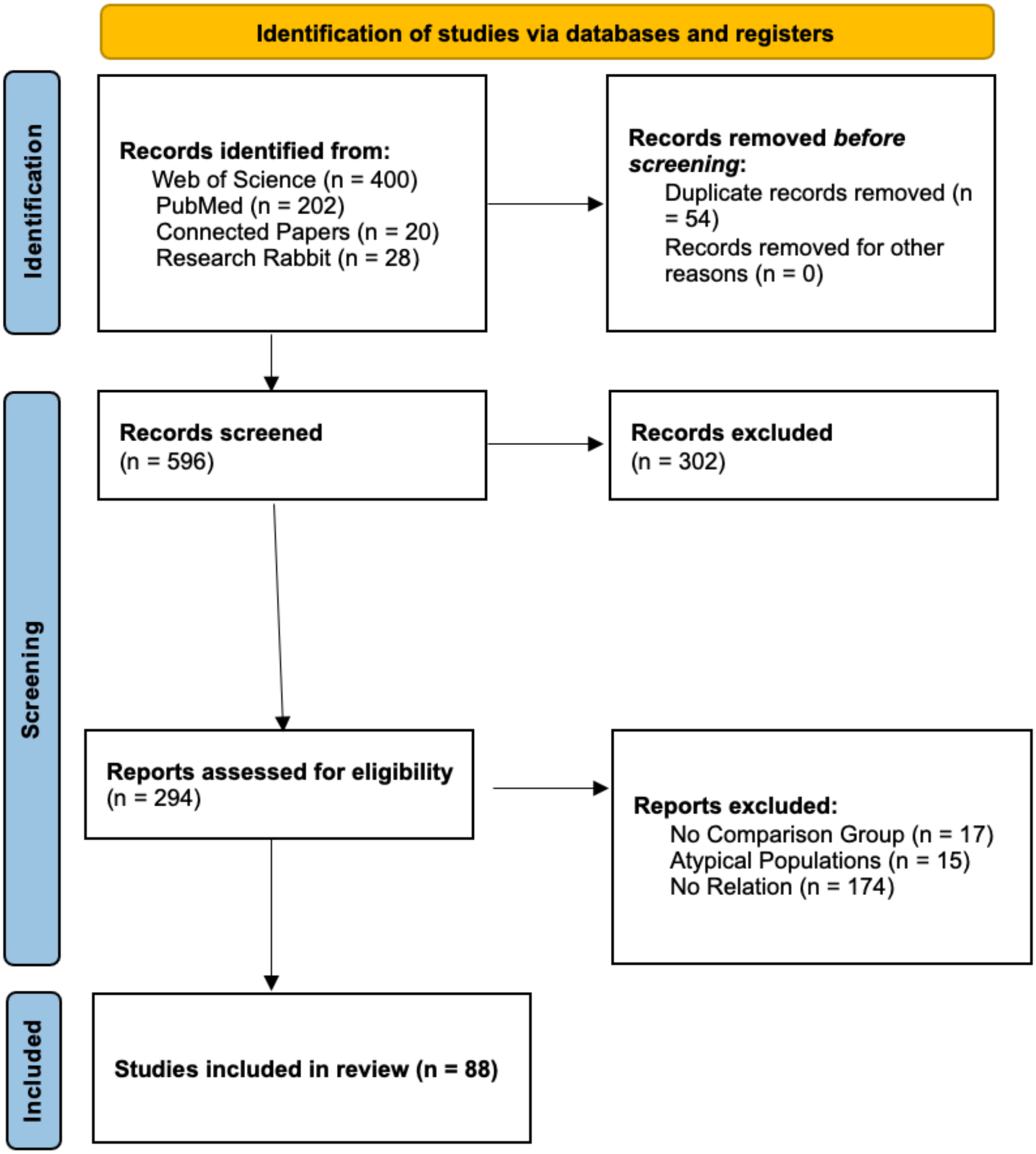
Woolard, A., Lane, A. E., Campbell, L. E., Whalen, O. M., Swaab, L., Karayanidis, F., Barker, D., Murphy, V., & Benders, T. (2022). Infant and Child-Directed Speech Used with Infants and Children at Risk or Diagnosed with Autism Spectrum Disorder: A Scoping Review. *Review Journal of Autism and Developmental Disorders*, *9*, 290–306. https://doi.org/10.1007/s40489-021-00253-y

Xu, D., Yapanel, U., & Gray, S. (2009). Reliability of the LENA Language Environment Analysis System in young children's natural home environment. *Boulder, CO: Lena Foundation*, 1–16.

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

Supplementary Information

S1: Details about the Systematic Search



Supplementary Figure 1.1: PRISMA chart of the systematic review process

Topic	No.	Item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Title
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist	
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Introduction
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Introduction
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Methods
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Methods
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Methods
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Methods
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Methods

Topic	No.	Item	Location where item is reported
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Methods
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Methods
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Methods
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	Methods
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item 5)).	Methods
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	Methods
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Methods
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Methods
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	Methods
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	Methods
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Methods

Topic	No.	Item	Location where item is reported
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Methods
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	S1.1
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	NA
Study characteristics	17	Cite each included study and present its characteristics.	S2
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	S1.1
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Results
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Results
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	Results
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Results
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	Results
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	Methods
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Results
DISCUSSION			

Topic	No.	Item	Location where item is reported
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Discussion
	23b	Discuss any limitations of the evidence included in the review.	Discussion
	23c	Discuss any limitations of the review processes used.	Discussion
	23d	Discuss implications of the results for practice, policy, and future research.	Discussion
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	NA
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Methods
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	NA
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	Acknowledgements
Competing interests	26	Declare any competing interests of review authors.	Competing Interests Statement
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Data & Code availability statement

Supplementary Table 1.2: PRISMA Main Checklist

Topic	No.	Item	Reported?
TITLE			
Title	1	Identify the report as a systematic review.	Yes
BACKGROUND			
Objectives	2	Provide an explicit statement of the main objective(s) or question(s) the review addresses.	Yes
METHODS			
Eligibility criteria	3	Specify the inclusion and exclusion criteria for the review.	Yes
Information sources	4	Specify the information sources (e.g. databases, registers) used to identify studies and the date when each was last searched.	Yes
Risk of bias	5	Specify the methods used to assess risk of bias in the included studies.	Yes
Synthesis of results	6	Specify the methods used to present and synthesize results.	Yes
RESULTS			
Included studies	7	Give the total number of included studies and participants and summarise relevant characteristics of studies.	Yes
Synthesis of results	8	Present results for main outcomes, preferably indicating the number of included studies and participants for each. If meta-analysis was done, report the summary estimate and confidence/credible interval. If comparing groups, indicate the direction of the effect (i.e. which group is favoured).	Yes
DISCUSSION			
Limitations of evidence	9	Provide a brief summary of the limitations of the evidence included in the review (e.g. study risk of bias, inconsistency and imprecision).	Yes
Interpretation	10	Provide a general interpretation of the results and important implications.	Yes
OTHER			
Funding	11	Specify the primary source of funding for the review.	Yes
Registration	12	Provide the register name and registration number.	NA

Supplementary Table 1.3: PRISMA Abstract Checklist

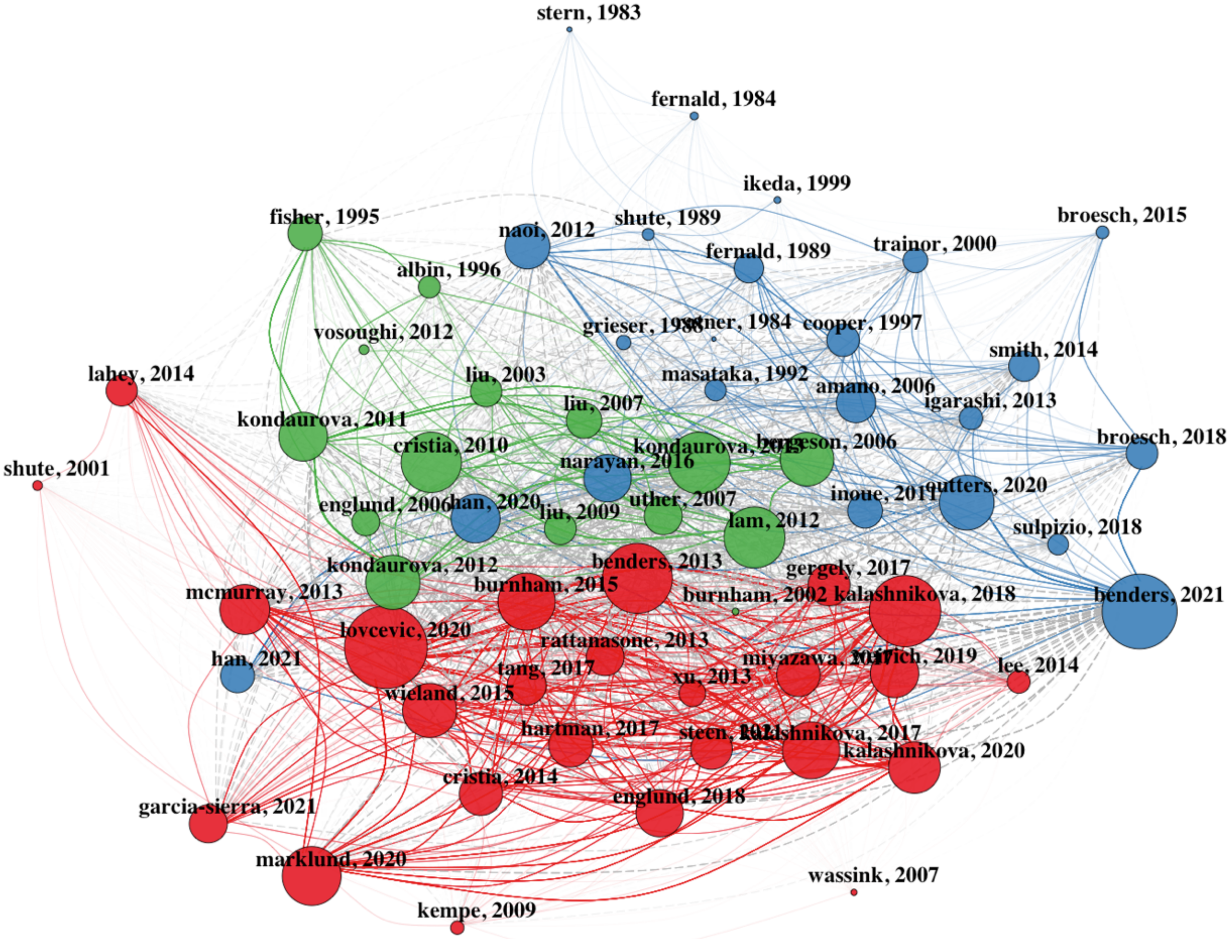
S1.1: Risk of Bias Assessment

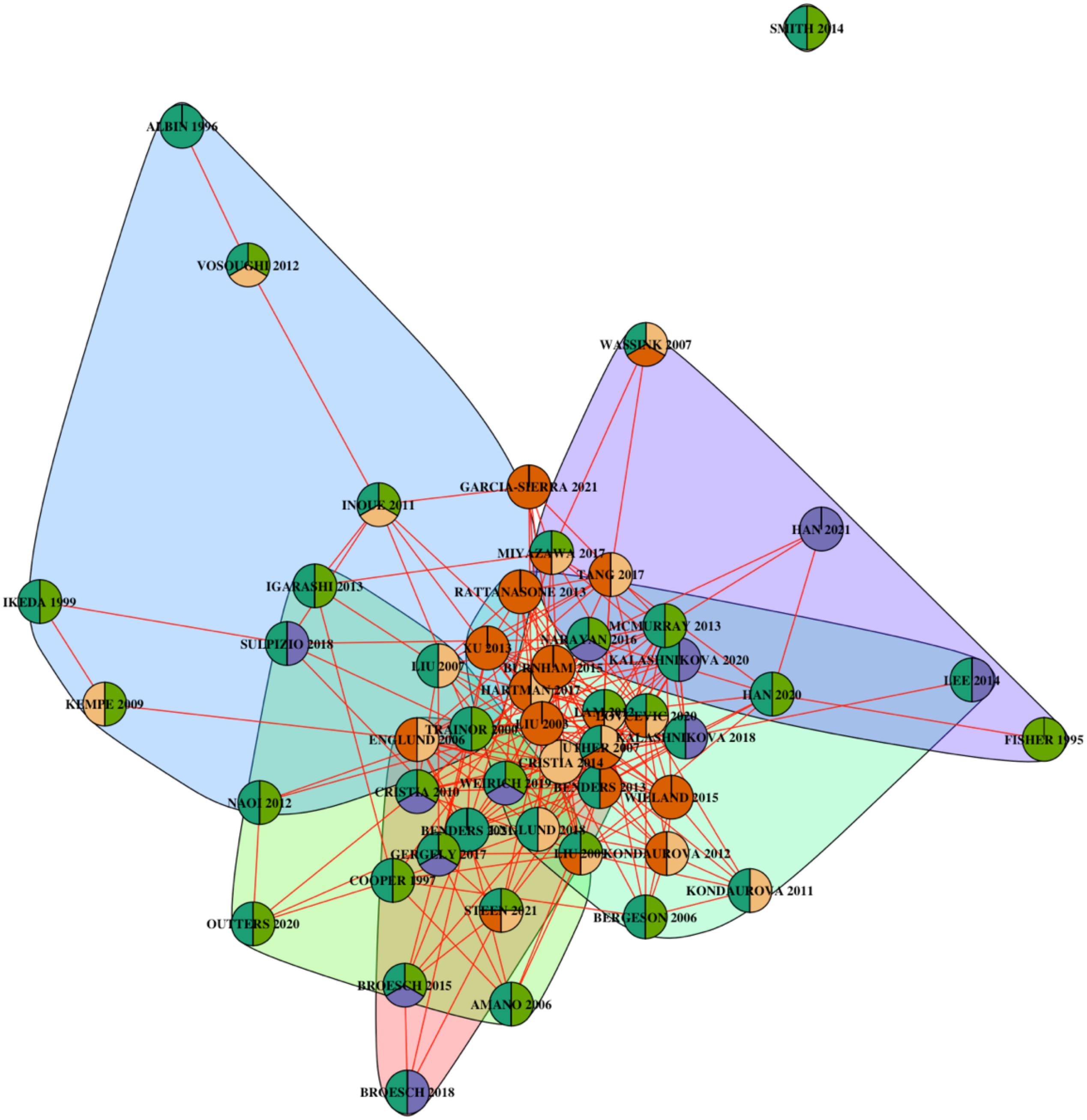
Despite being less prone to bias than more subjective literature overviews, systematic searches and meta-analyses cannot completely avoid bias. In this section, we discuss some of the potential biases in our systematic search and selection process. Firstly, *our choice of search terms may select a biased subset of the literature*. In order to counteract this potential source of bias, we aimed to make our list of search terms as inclusive as possible. We conducted initial searches after carefully reading through relevant papers and included additional terms before conducting the final systematic search. We also performed forward and backward searches of the literature using cutting-edge bibliography tools (Research Rabbit and Connected Papers) to expand the scope of our search to relevant studies that were not found initially. Secondly, the *published literature itself may represent a biased subset of the literature available* on the acoustic features of IDS, to a greater extent reporting outcome measures for which manipulation created a significant effect. In order to counteract the effects of this publication bias in the meta-analysis, we carried out the following. i) In the literature search we included both published and grey literature, such as pre-prints, conference proceedings, etc. and informally solicited literature suggestions on twitter and from experts. ii) In the manuscript, we actively encourage researchers with unpublished and published work to submit their experimental results to an open repository with the data from our meta-analysis (MetaLab: https://langcog.github.io/metalab/). iii) In the meta-analysis, we assess the extent to which the meta-analytic estimates change under different assumptions of publication bias in the literature, by conducting quantitative sensitivity analyses. We should also note that a related source of bias may manifest itself in *the study selection process where authors exclude papers not conforming to their hypotheses*. However, because this project concerns estimates of acoustic features above and beyond statistical significance tests, we had no specific directional hypotheses to test and strong incentives to include as much data as possible. Thirdly, *bias might arise as a function of the reporting of estimates* (e.g., studies with missing estimates of uncertainty may be of systematically lower quality than other studies). Because most of the papers with missing data were older papers, we chose not to contact the original authors to provide the missing data because we know from previous work that answers are extremely unlikely. We instead decided to impute the missing measures of uncertainty, as outlined in Section 2.1. This imputation process was shown not to bias any of the results, as shown in Section S3, and better counteracts this potential source of bias than simply excluding studies not reporting measures of uncertainty.

In general, although no analysis can remain completely unbiased, we hope this project can serve as a first step towards a cumulative self-correcting enterprise. Accordingly, we make our data openly available as a Community Augmented Meta-Analysis on the MetaLab website[151]. This makes it possible to critique, integrate and update our selection of studies in a straightforward manner.

S2: Citation Networks

The upper network shows the co-citation coupling strength (i.e., the number of times two studies are cited together by a new article as well as bibliographic similarity) for only the journal articles of the final sample of cited studies. The colour and thickness of the lines represent clusters of strong citation links. The lower direct-citation network shows which studies cite each other. The colours represent clusters of strong direct-citation links. In the below direct-citation plot, the colours of the nodes represent the acoustic measures under investigation; specifically, dark green is f_0, light green is f_0 variability, orange is vowel space area, purple is articulation rate, and light orange is vowel duration.





Supplementary Figure 2.1: Coupling (upper) and Direct-Citation (lower) Networks of Studies on IDS.

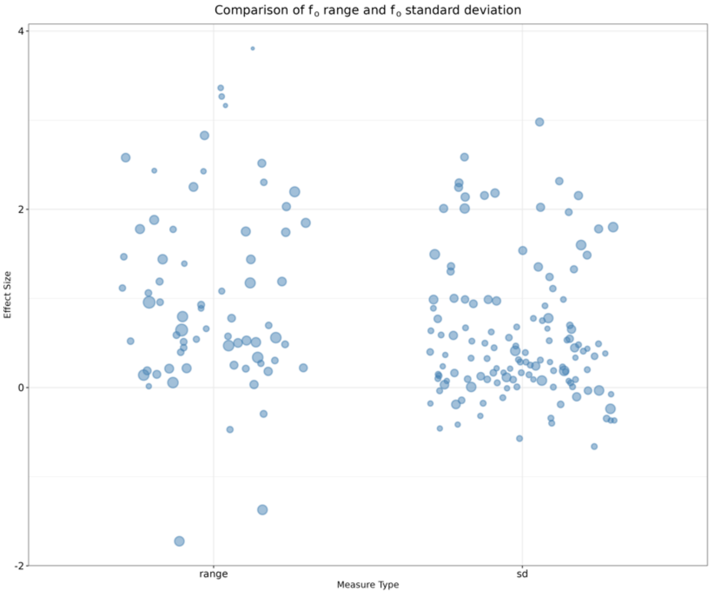
S3: Imputation Process

In order to incorporate the statistical uncertainty associated with the partially stochastic nature of this imputation process (Azur et al., 2011; Sterne et al., 2009), we constructed 20 datasets with sample size, mean values for each acoustic variable, and existing standard deviation values as predictors. The standard deviation values of the imputed datasets were checked for similarity to the reported standard deviations and post-processed to include only values within the range of the existing standard deviation values. In order to check that this process of multiple imputation did not bias the estimation of the overall effect size for each acoustic measure, we compared the estimates of the intercepts-only models for the imputed and non-imputed datasets, as shown in Supplementary Table 3.1 below. There does not appear to be evidence of bias, as the effect size estimate of the models with the imputed datasets lies within the credible interval of the non-imputed datasets in each case.

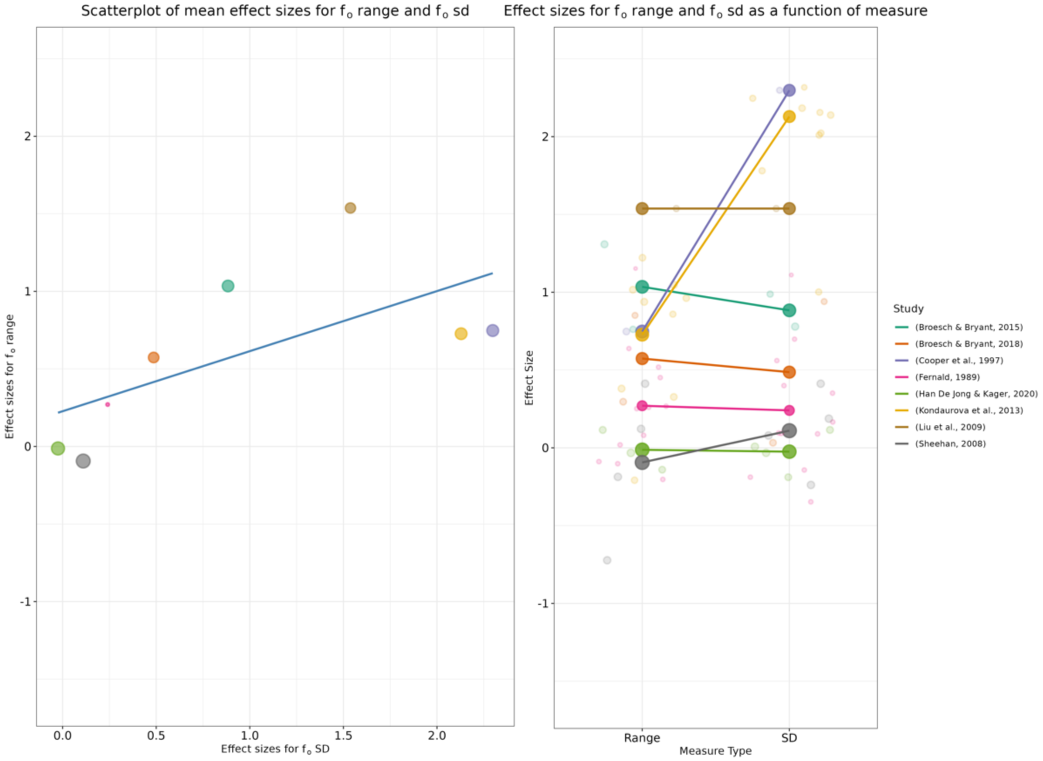
Acoustic Measure	**Intercept Estimate *Without* Imputation (n = total observations)**	**Effect Size Estimate *With* Imputation (n = total observations)**
f_0	1.09 [0.83; 1.34] (n = 250)	1.17 [0.86; 1.45] (n = 262)
f_0 Variability	0.76 [0.49; 1.00] (n = 208)	0.69 [0.44; 1.92] (n = 223)
Vowel Space Area	0.49 [-0.08; 1.09] (n = 51)	0.66 [0.34; 0.98] (n = 107)
Articulation Rate	-0.91 [-1.42; -0.42] (n = 56)	-1.05 [-1.53; -0.60] (n = 60)
Vowel Duration	0.47 [0.02; 0.91] (n = 72)	0.48 [0.08; 0.88] (n = 82)

Supplementary Table 3.1: An overview of the extent to which imputation has influenced the overall estimation of effect sizes for each acoustic measure

S4: Comparison of f_0 range and f_0 standard deviation



Supplementary Figure 4.1: A plot showing the distribution of effect sizes for fo range and fo standard deviation. The similar distributions speak in favor of our choice to combine the measures into one measure of f_0 variability.



Supplementary Figure 4.2: A correlation scatter plot showing the distribution of average effect sizes for the studies reporting both measures (left) and a plot of the effect sizes as a function of measure (right). A Bayesian multivariate model with range and standard deviation as separate outcomes shows a strong

correlation between the two measures 0.73 [0.38; 0.98] (without Kondaurova et al., 2013), as these authors report f_0 range in semitones and f_0 standard deviation in Hz). This estimate is based on a total of 573 participants across the 8 studies.

S5: Choice of Priors, Prior and Posterior Predictive Checks, Prior-Posterior Update Plots, Prior Robustness Checks

S5.1: Choice of Priors

We chose weakly informative priors in order to ensure that their influence on the meta-analytic estimates was small and to discount extreme effect sizes as unlikely[152, 153]

(cf. Lemoine, 2019; Gelman, Simpson & Betancourt, 2017); for the overall effect size, we chose a Gaussian distribution with a mean of 0 and standard deviation of 2.5 based on our prior expectations for effect sizes. This prior implies that we expect approximately 95% of the effect size distribution to be between -5 and 5. For the slope of the model, we encoded our expectations with a Gaussian prior with a mean of 0 and a standard deviation of 1, which implies that we expect the vast majority of values for the coefficient of the effect size difference between ADS and IDS to be between -2 and 2. For the heterogeneity of the effects (i.e., the standard deviation of random effects), we chose a positive truncated normal distribution with a mean of 1 and standard deviation of 1. For the degrees of freedom parameter, ν, of the Student's t-distribution, a gamma distribution with a shape parameter of 2 and a scale parameter of 0.1 was chosen. This ensures that the model remains robust to the influence of outliers. Prior predictive checks were performed to ensure that model predictions for plausible values of effect sizes would only exclude implausibly high or low values on the basis of the priors[152].

The models were fitted with Hamiltonian Monte Carlo samplers with 2 parallel chains with 5,000 iterations each, an adapt delta of 0.99 and a maximum tree depth of 20 in order to ensure no divergence in the estimation process. The quality of the models was assessed by i) ensuring Rhat statistics to be lower than 1.1, ii) carrying out prior and posterior predictive checks, iii) plotting prior against posterior estimates and assessing whether the posteriors had lower variance than the priors, iv) ensuring no divergences in the process of estimation, v) checking that the number of effective bulk and tail samples was above 200, vi) conducting prior sensitivity analyses.

For the intercepts-only and full models, we used the following *brms*[142] formula and priors, with a student t likelihood for all of the effect sizes measures, as shown below:

Intercepts Model Structure: *Effect_Size | se(Effect_Size_se) ~ 1 + (1 | Lang/StudySite/measure)*

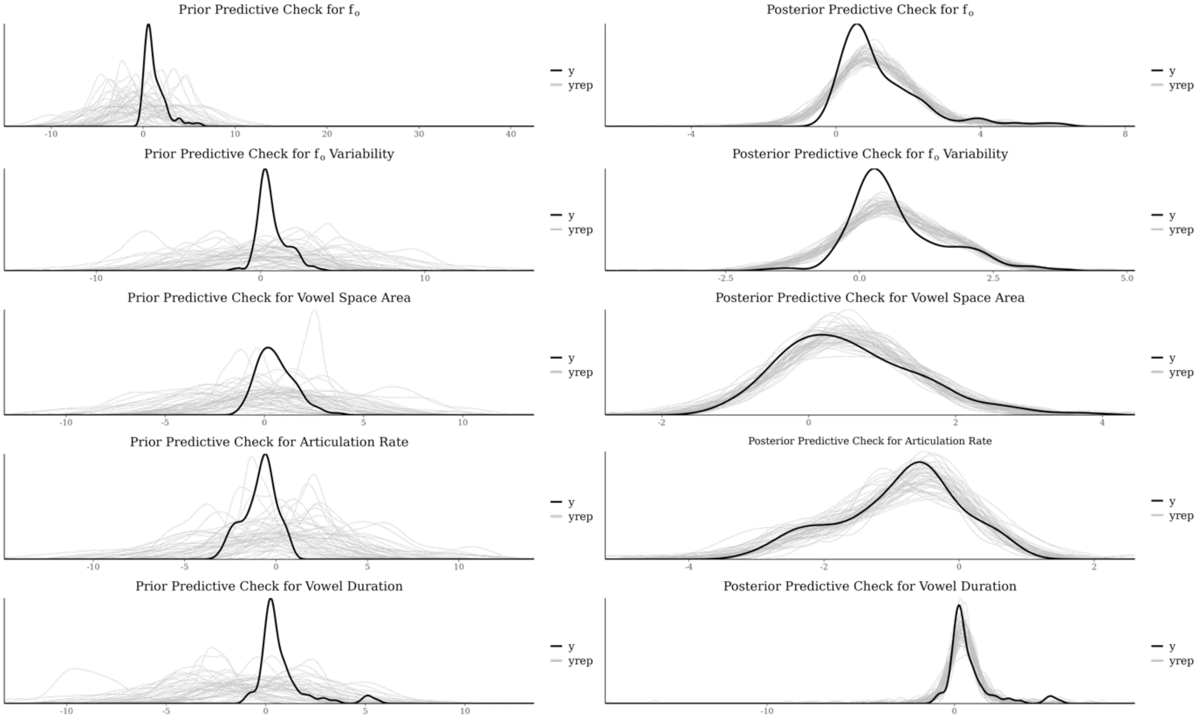
Full Model Structure: *Effect_Size | se(Effect_Size_se) ~ 1 + Age + Lang + Environment + Task + (1 | Lang/StudySite/measure))*

Models	Intercepts	Slopes	SD	DoF
Intercepts Model	N(0,2.5)	-	N(1,1)	G(2,.1)
Full Moderators Model	N(0,2.5)	N(0,1) for Task & Environment N(0.05) for Age	N(1,1)	G(2,.1)

Supplementary Table 5.1: Priors for the parameters in the intercepts-only model and full model with all moderators. N() refers to a normal distribution, G() indicates a gamma distribution, lkj() refers to the Lewandowski-Kurowicka-Joe distribution. DoF refers to Degrees of Freedom parameter (or ν).

S5.2: Prior & Posterior Predictive Checks

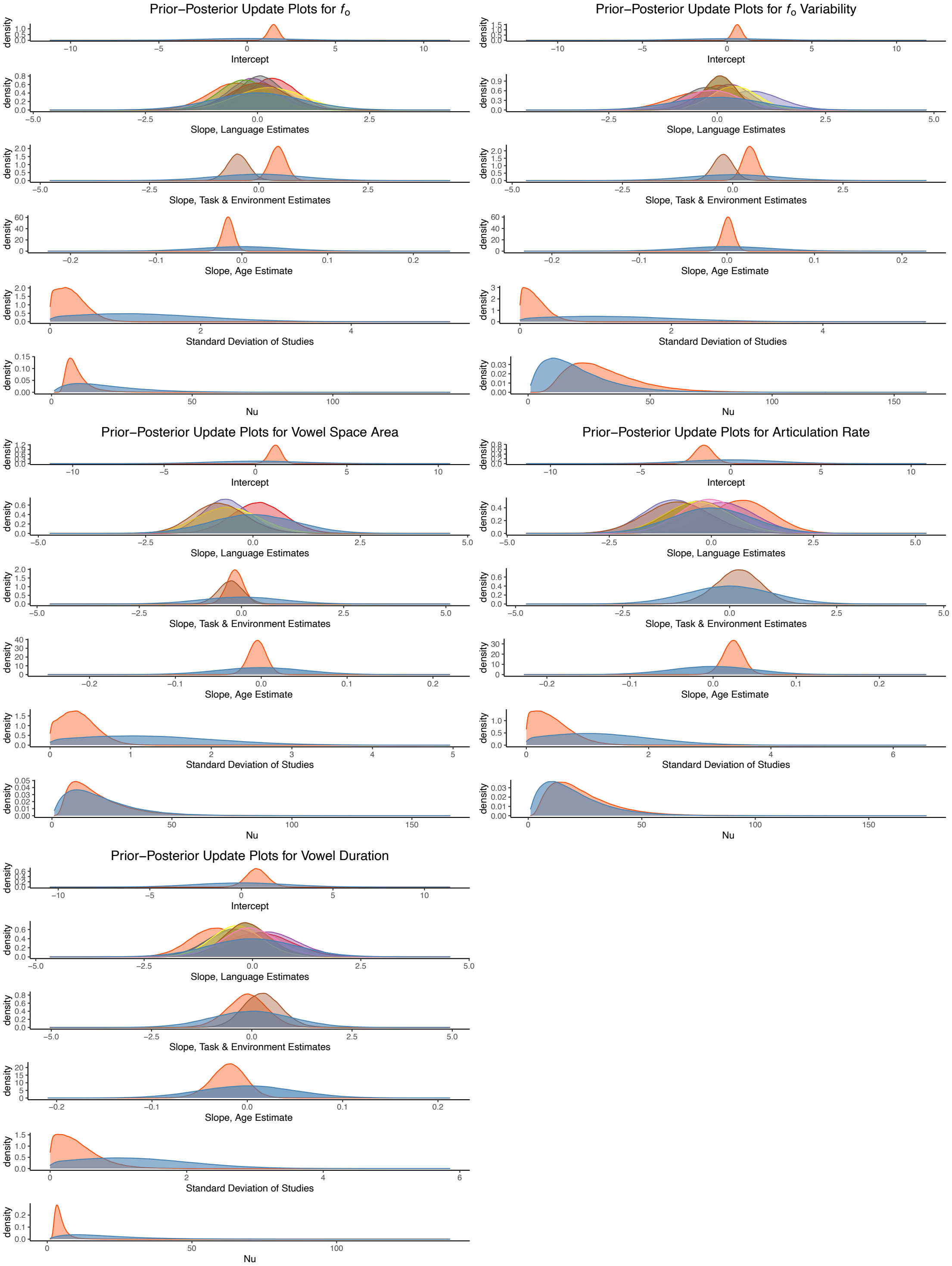
As noted above, we performed quality checks of the models by carrying out prior and posterior predictive checks. The below prior predictive checks (on the left) indicate that our priors predict values within the order of magnitude of the distribution. The posterior predictive checks (on the right) indicate that the models have captured the distributions of data for each of the acoustic measures. These plots provide reassurance that our models capture relevant aspects of the overall distributions of dependent variables.



Supplementary Figure 5.2.1: Plot of the prior and posterior predictive checks (grey) and observed meta-analytic data (black) for the acoustic measures.

S5.3: Prior-Posterior Update Plots

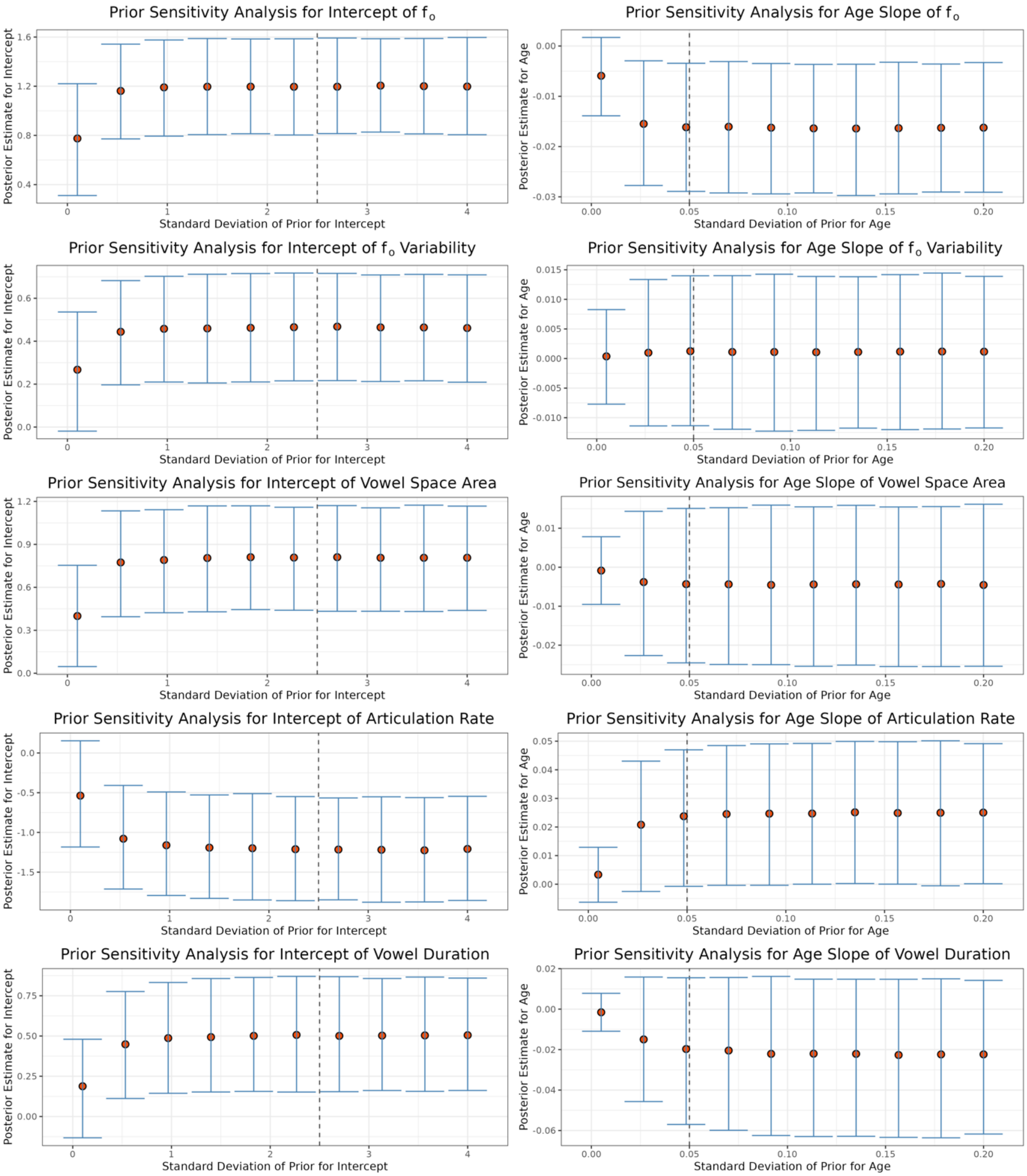
A second quality check of the models was carried out by plotting the prior distributions against the posterior estimates of the model. As shown in the below plots, the posteriors exhibit lower variance than the priors. These plots thus indicate that the models have learned from the data and provide additional reassurance that our models have captured relevant information.



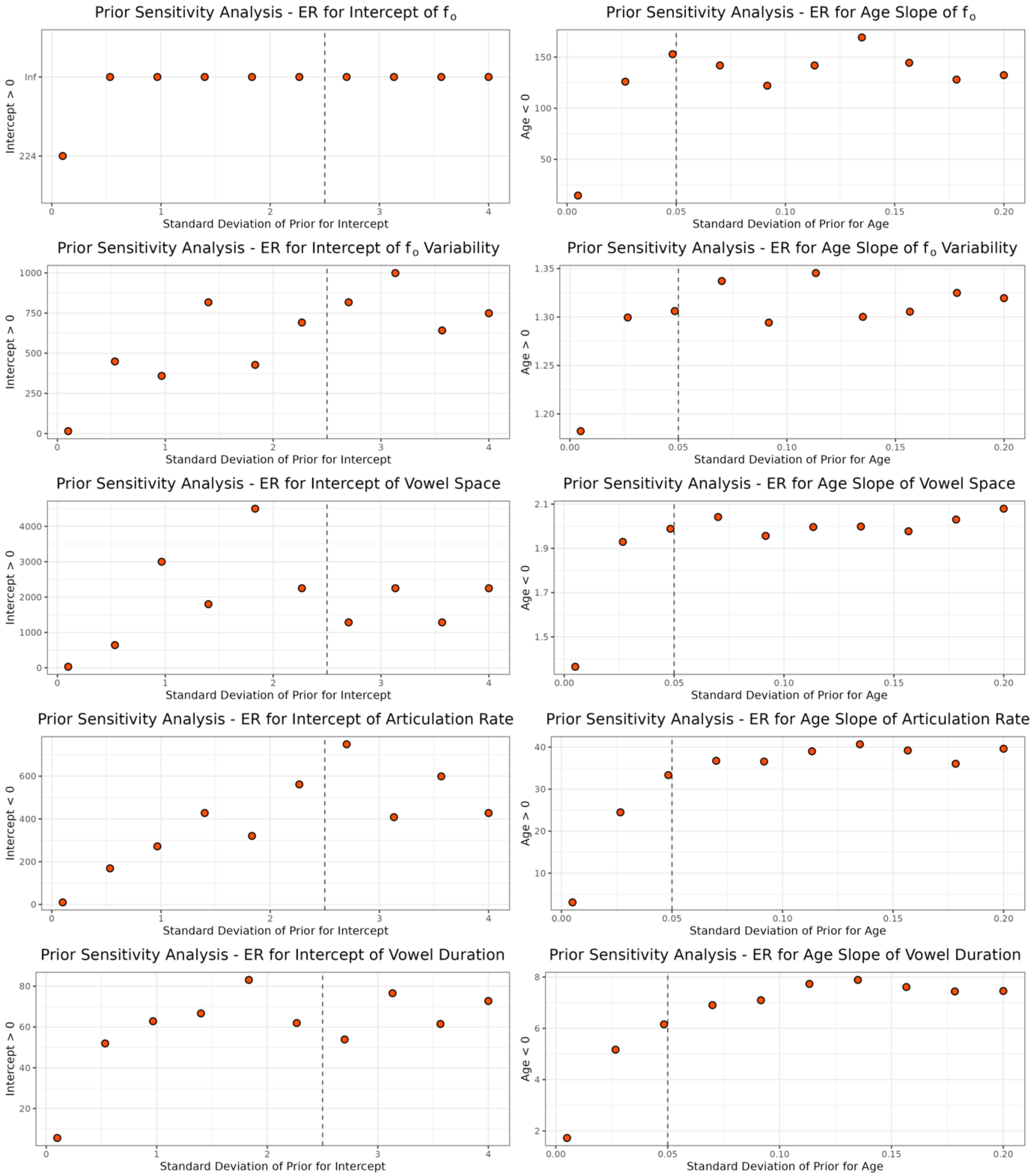
Supplementary Figure 5.3.1: Panel of prior-posterior update plots for the intercept, slope, standard deviation, and nu for each of the acoustic variables under investigation. The prior distributions are represented in blue. In the plots of task and environment, task is represented by orange and environment is represented by brown.

S5.4: Prior Sensitivity Analysis for Intercept & Slope

A third quality check of the models was performed by assessing the extent to which the uncertainty of our priors affected posterior estimates. Because the posterior estimates (on the y-axis) exhibit stability at our choices of priors (i.e., the dashed vertical line), these plots provide reassurance that our choice of priors did not unduly affect model estimations.



Supplementary Figure 5.4.1: Panel of plots showing how the intercept and age estimates for each acoustic variable change with different standard deviations for the priors. The vertical dashed line indicates the standard deviation of the prior chosen for the models. The centres of the error bars (orange points) indicate the posterior estimates for the intercept (left column) and age predictor (right column). The total sample sizes across studies for each of the estimates were 3401, 3006, 1702, 976, 1411 participants for f_o, f_o variability, vowel space area, articulation rate, vowel duration, respectively.



Prior Sensitivity Analysis - ER for Intercept of f_o

Prior Sensitivity Analysis - ER for Age Slope of f_o

Prior Sensitivity Analysis - ER for Intercept of f_o Variability

Prior Sensitivity Analysis - ER for Age Slope of f_o Variability

Prior Sensitivity Analysis - ER for Intercept of Vowel Space

Prior Sensitivity Analysis - ER for Age Slope of Vowel Space

Prior Sensitivity Analysis - ER for Intercept of Articulation Rate

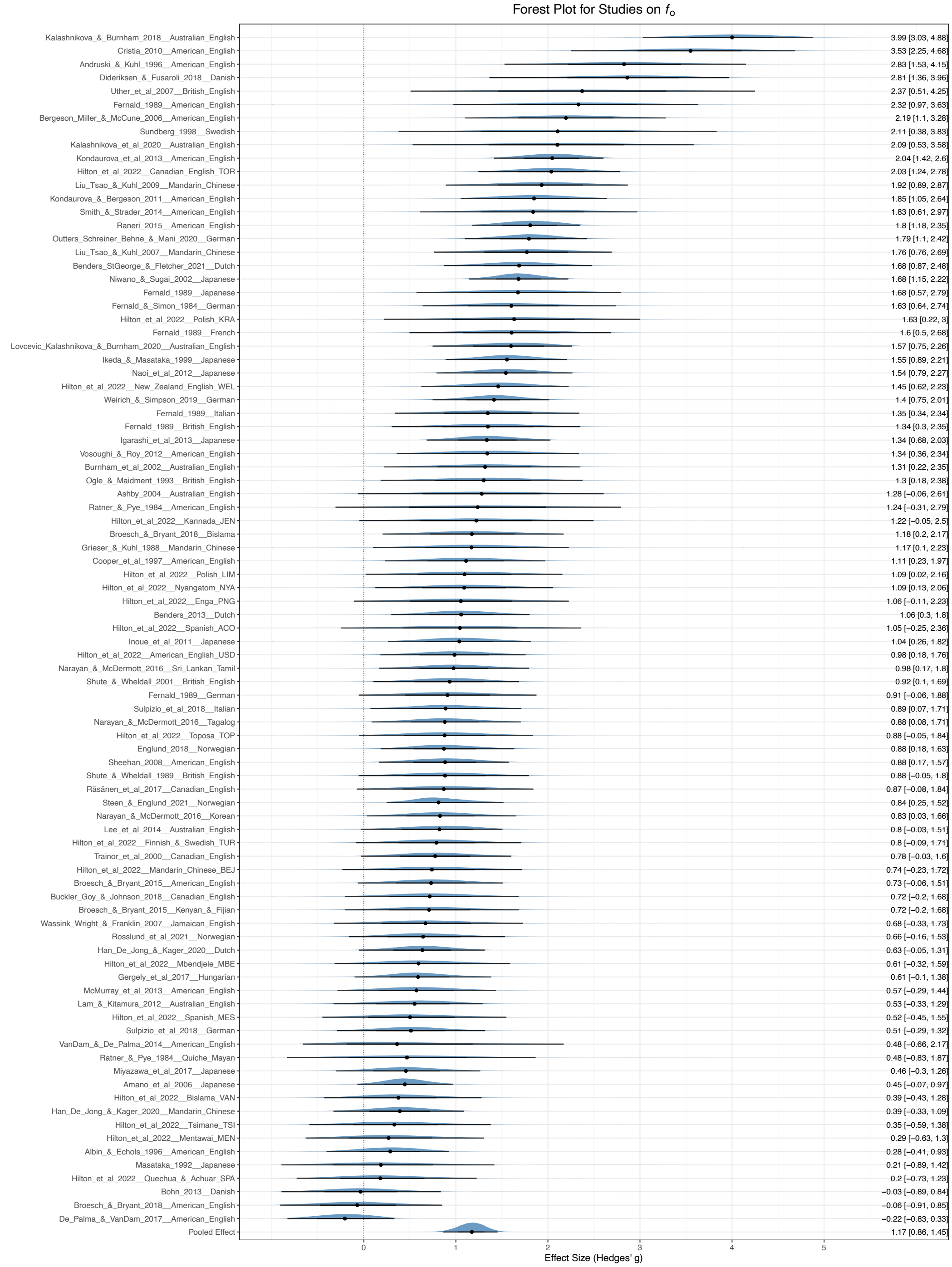
Prior Sensitivity Analysis - ER for Age Slope of Articulation Rate

Prior Sensitivity Analysis - ER for Intercept of Vowel Duration

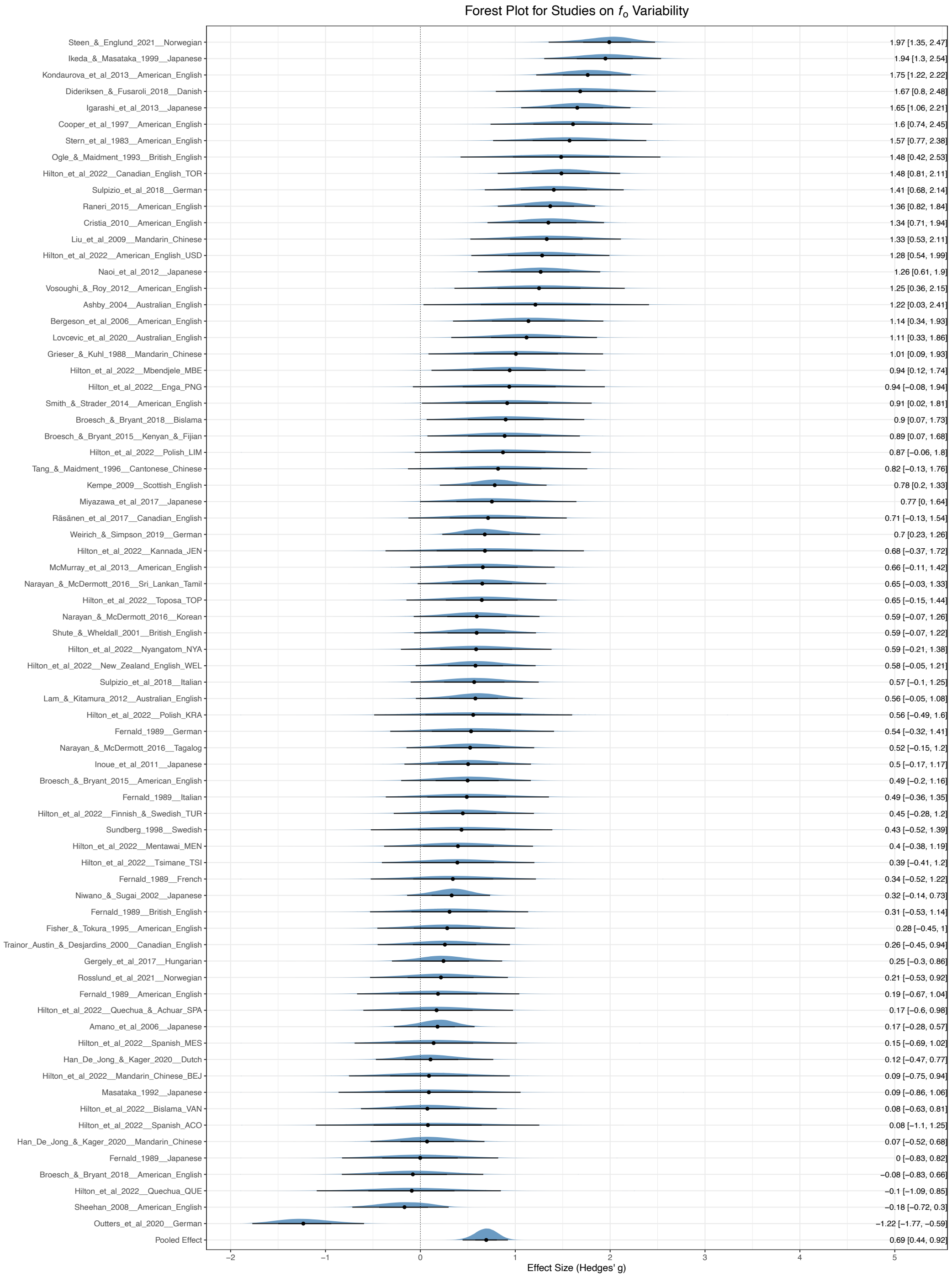
Prior Sensitivity Analysis - ER for Age Slope of Vowel Duration

Supplementary Figure 5.4.2: Panel of plots showing how the evidence ratio (ER) for the intercept and age estimates for each acoustic variable change with different standard deviations for the priors. The vertical dashed line indicates the standard deviation of the prior chosen for the models.

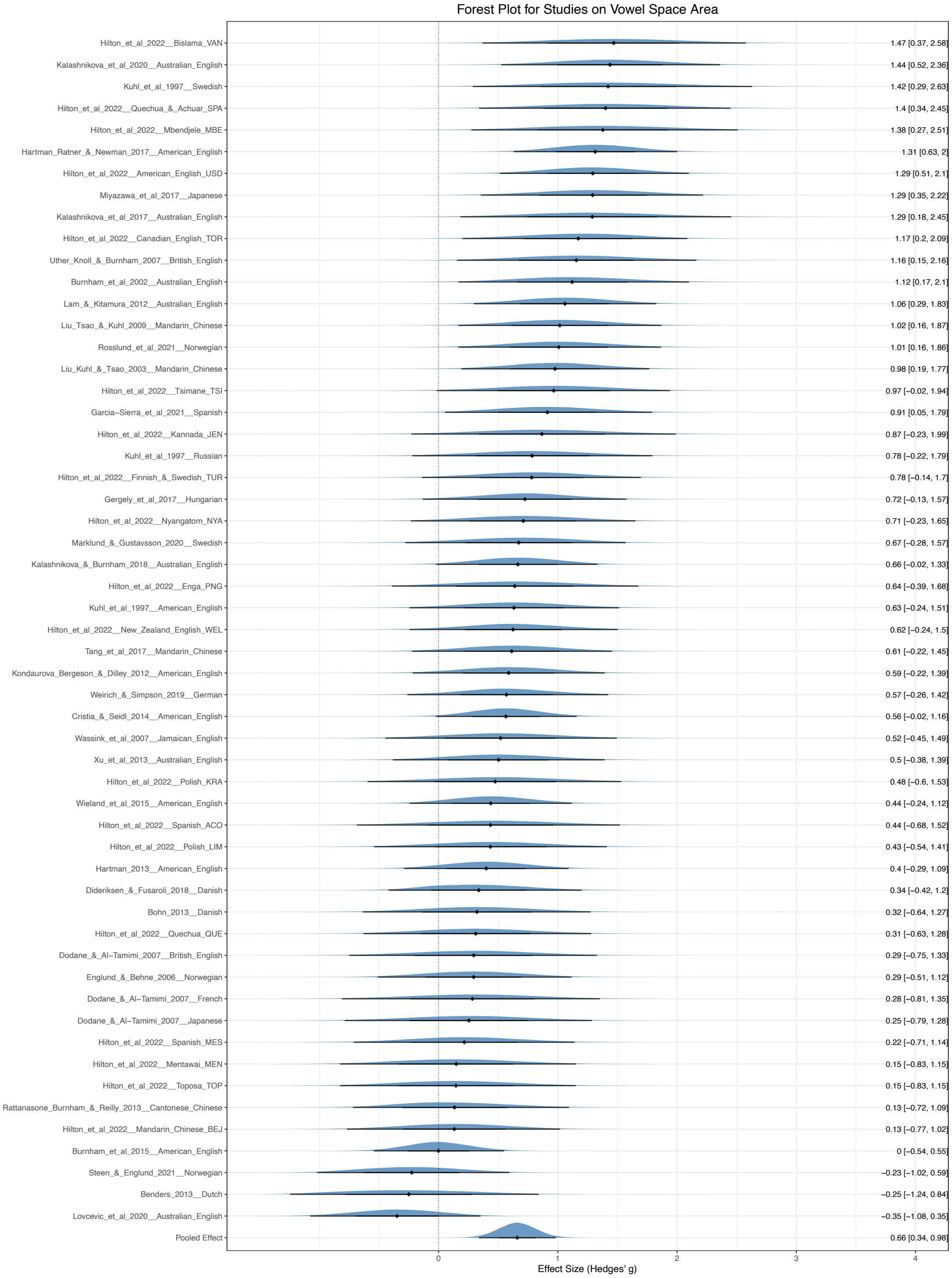
S6: Forest Plots & Overview Plots



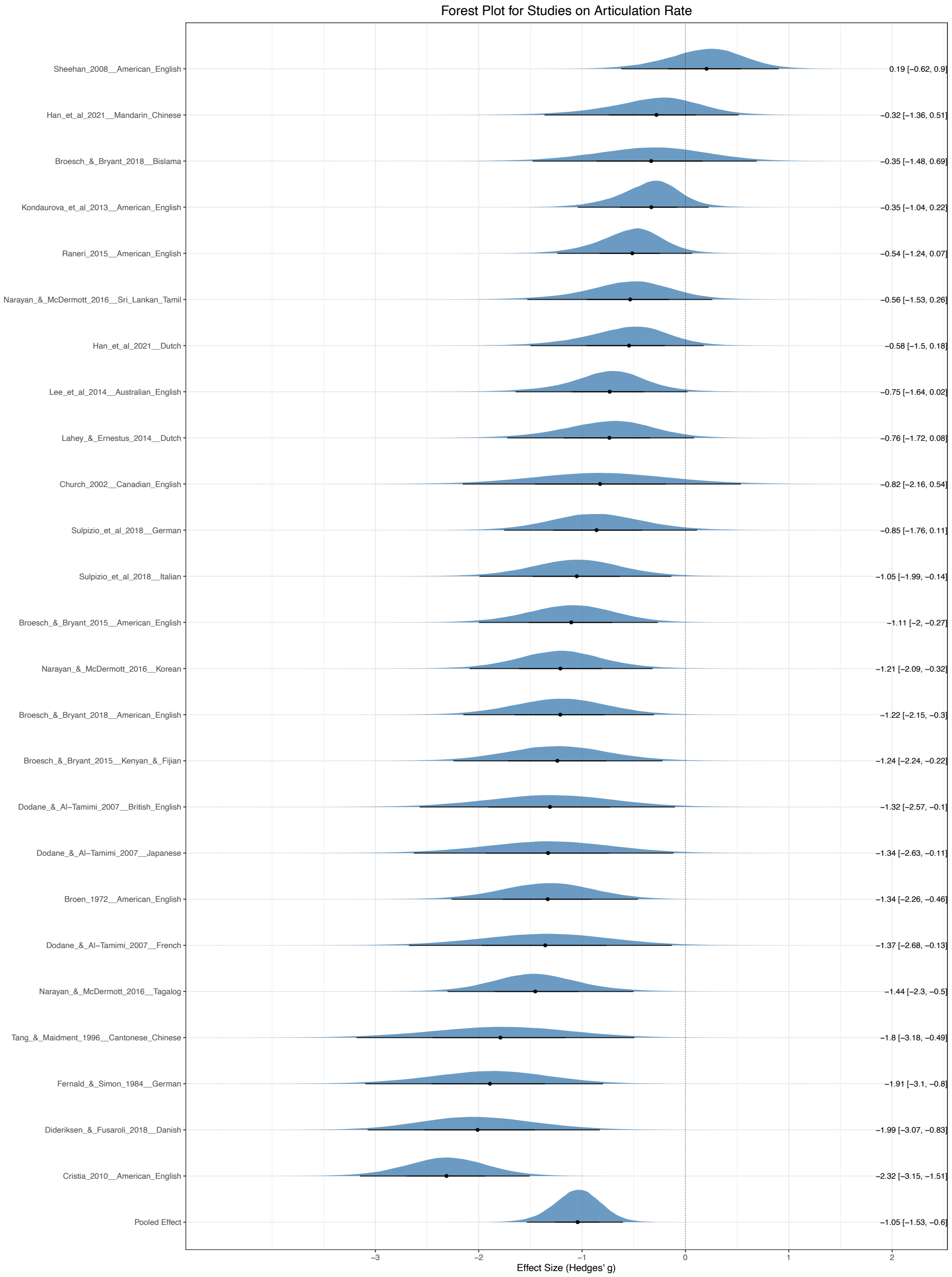
Supplementary Figure 6.1. Forest Plot for fo estimates according to study and language. The estimates are based on a total of 3401 participants across 60 studies investigating 33 distinct languages. The shaded areas indicate the posterior probability density of each estimate. The numbers to the right provide the estimated mean effect size (Hedges' *g*) and upper and lower 95% credible intervals). The estimates within each study are broken down according to language; this is especially evident for studies[1, 54], from which data on a diverse set of languages exist.



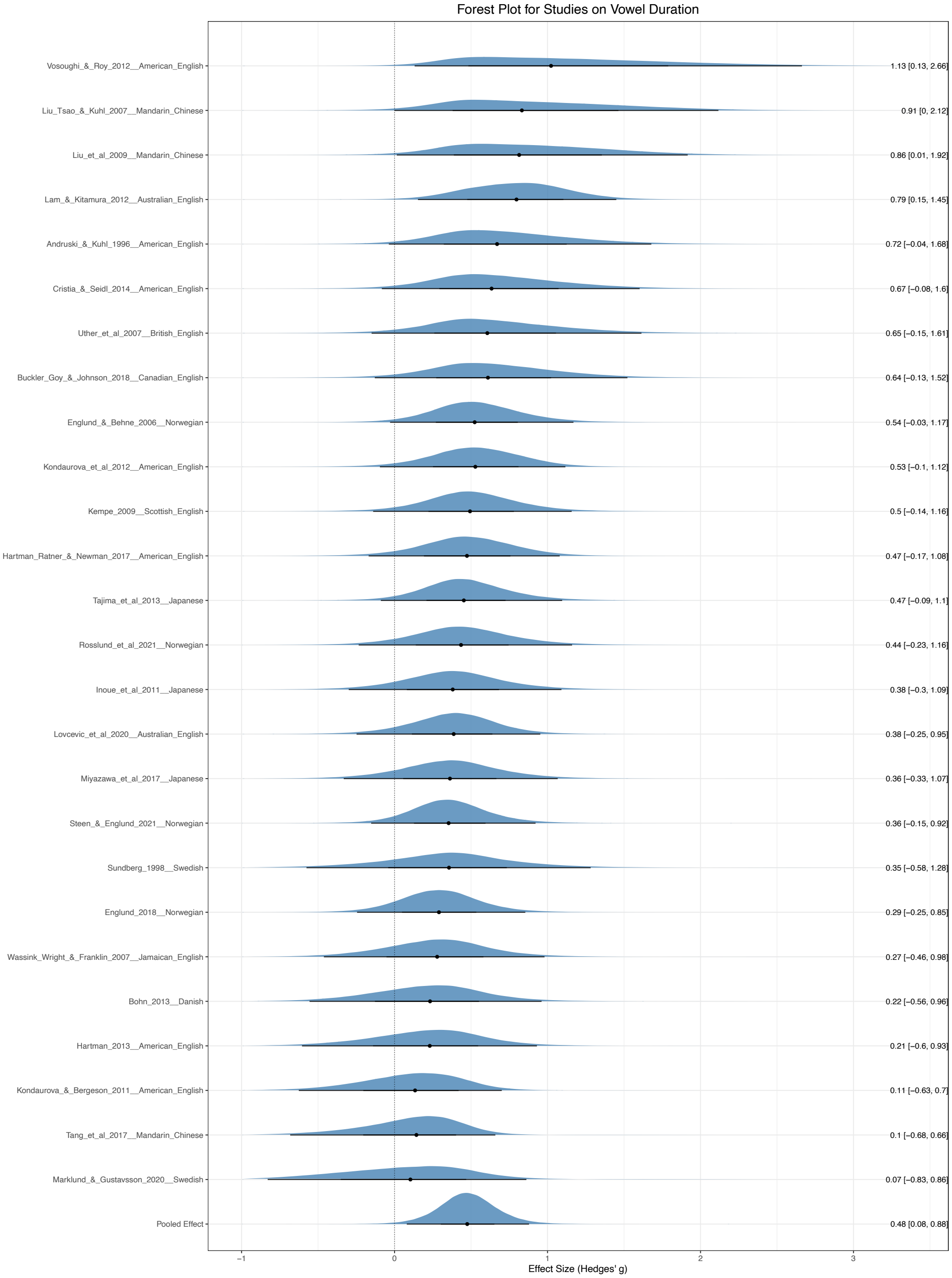
Supplementary Figure 6.2. Forest Plot for fo variability estimates according to study. The estimates are based on a total of 3006 participants across 44 studies investigating 34 distinct languages. The shaded areas indicate the posterior probability density of each estimate. The numbers to the right provide the estimated mean effect size (Hedges' *g*) and upper and lower 95% credible intervals).



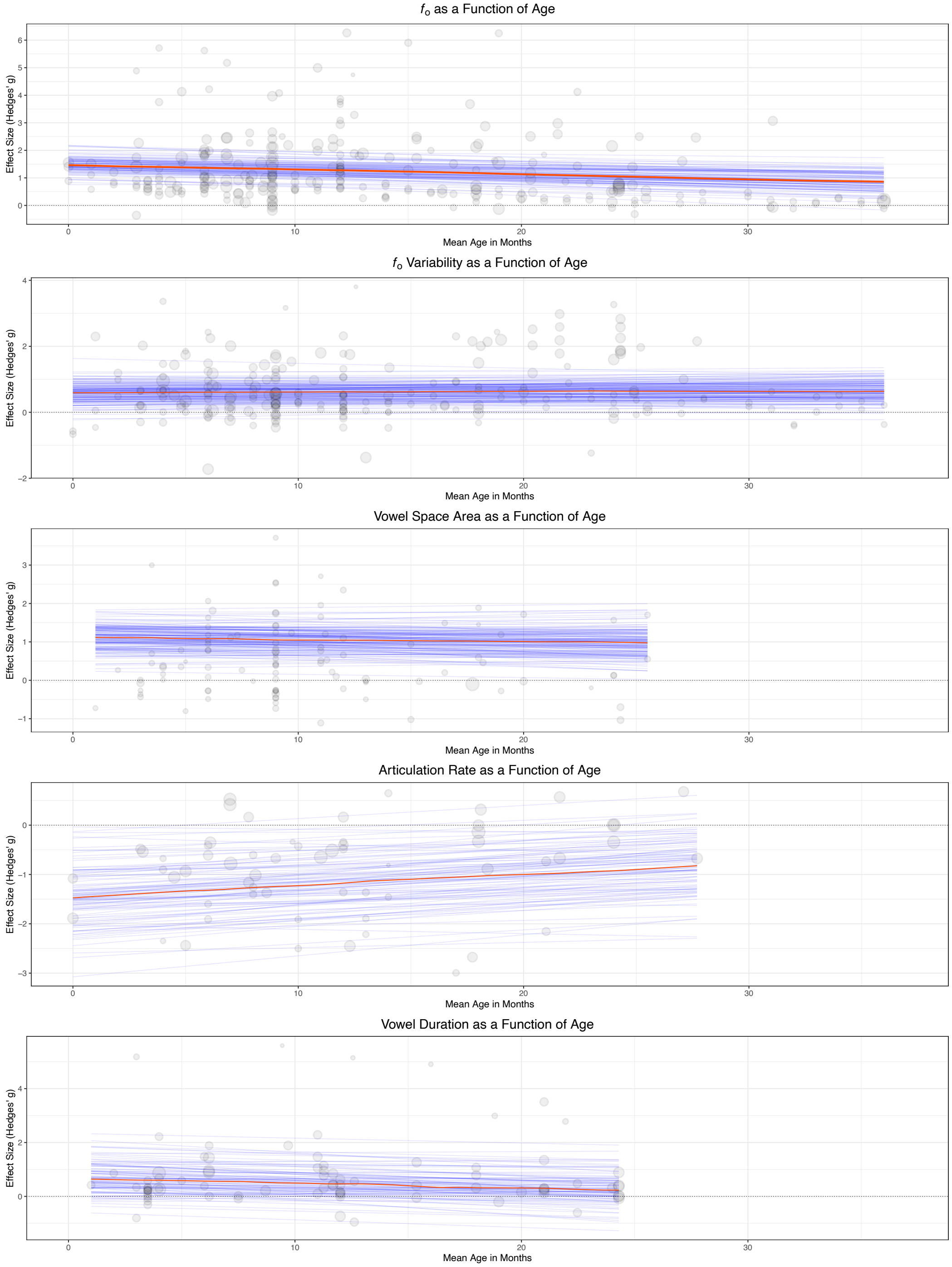
Supplementary Figure 6.3. Forest Plot for vowel space area estimates according to study. The estimates are based on a total of 1702 participants across 33 studies investigating 30 distinct languages. The shaded areas indicate the posterior probability density of each estimate. The numbers to the right provide the estimated mean effect size (Hedges' *g*) and upper and lower 95% credible intervals). The estimates within each study are broken down according to language; this is especially evident for studies[1, 54], from which data on a diverse set of languages exist.



Supplementary Figure 6.4. Forest plot for articulation rate estimates according to study. The estimates are based on a total of 976 participants across 17 studies investigating 17 distinct languages. The shaded areas indicate the posterior probability density of each estimate. The numbers to the right provide the estimated mean effect size (Hedges' *g*) and upper and lower 95% credible intervals).



Supplementary Figure 6.5. Forest plot for vowel duration estimates according to study. The estimates are based on a total of 1411 participants across 26 studies investigating 11 distinct languages. The shaded areas indicate the posterior probability density of each estimate. The numbers to the right provide the estimated mean effect size (Hedges' *g*) and upper and lower 95% credible intervals).

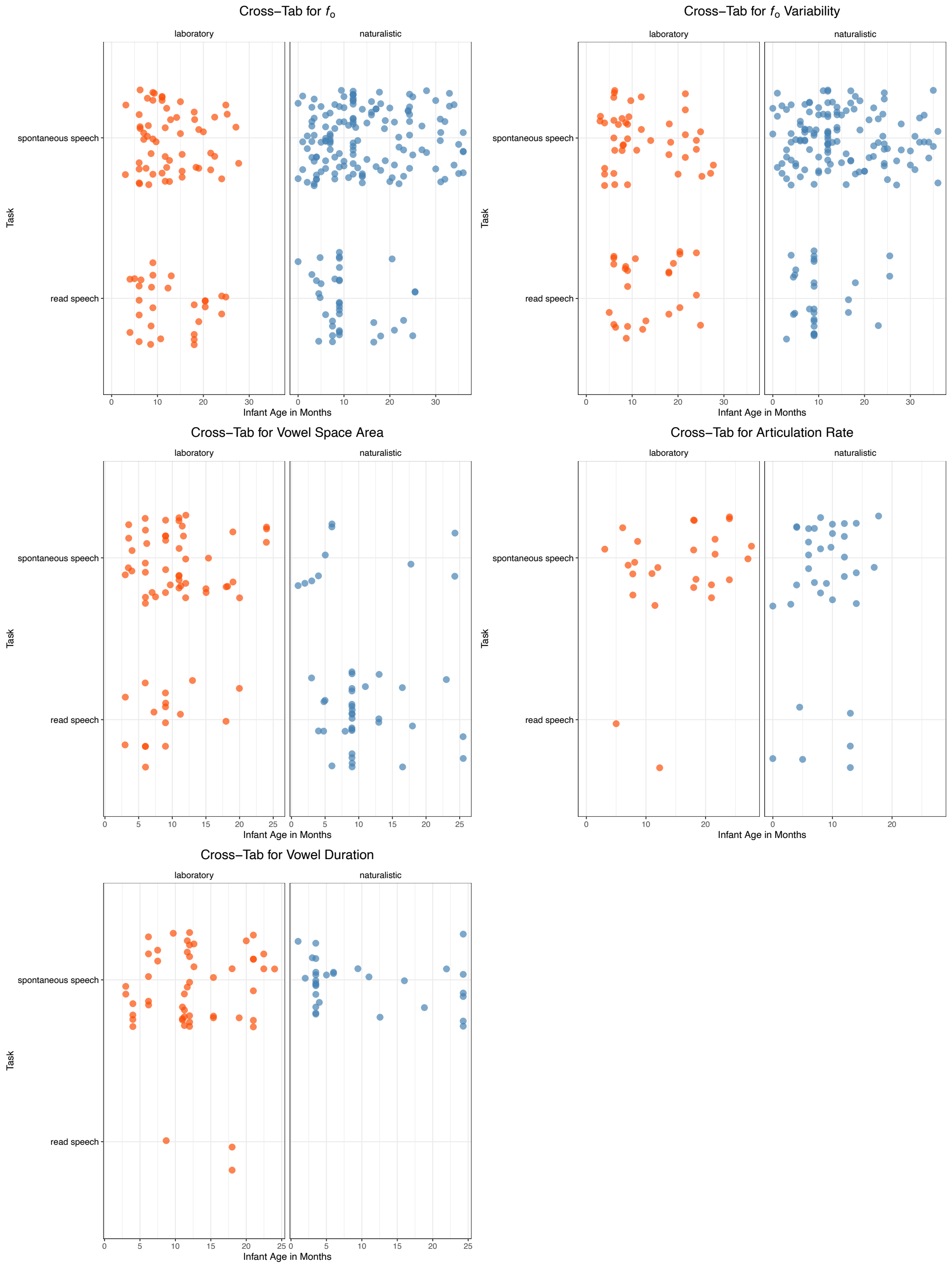


Supplementary Figure 6.6. A panel of plots showing each acoustic measure as a function of infant age. The blue lines reflect 100 posterior model predictions for the effect size estimates for each acoustic variable. As above, the grey points show the raw effect size measures. The size of the points is proportional to the inverse of the standard error of the effect size (i.e., the larger the point, the smaller the standard error). Note that each acoustic measure has an x-axis with different limits based on the data available.

S7: Age Distributions Across Languages

Supplementary Figure 7: Plots showing the age distribution by language of the effect sizes for each of the acoustic variables under investigation.

S8: Cross-Tab for Task & Environment



Supplementary Figure 8: Cross-tab for the predictors of task and environment with age on the x-axis. The colors are purely for aesthetic purposes: blue denotes recordings in a naturalistic environment, orange signifies recordings done in the lab.

S9: Parameter Estimates for Best Models

Model Parameters for f_0	**Estimate**
Standard Deviation of Languages	0.29 [0.01; 0.72]
Standard Deviation of Studies within Languages	0.91 [0.72; 1.14]
Standard Deviation of Measurements	0.07 [0.00; 0.19]
Age (months)	-0.02 [-0.03 0.01]
Australian English	1.29 [0.72; 1.84]
Bislama	0.85 [-0.37; 2.11]
British English	1.56 [0.43; 2.68]
Canadian English	1.21 [0.16; 2.21]
Danish	1.29 [-0.08; 2.64]
Dutch	0.97 [-0.12; 2.08]
Enga	1.33 [-0.25; 2.87]
Finnish & Swedish	1.19 [-0.27; 2.68]
French	1.53 [-0.01; 3.13]
German	1.34 [0.34; 2.35]
Hungarian	1.14 [-0.35; 2.62]
Italian	1.3 [0.03; 2.62]
Jamaican English	0.68 [-0.77; 2.18]

Japanese	1.04 [0.15; 1.99]
Kannada	1.46 [-0.17; 3.1]
Kenyan & Fijian	0.72 [-0.78; 2.22]
Korean	1 [-0.51; 2.47]
Mandarin Chinese	1.17 [0.15; 2.17]
Mbendjele	1.07 [-0.42; 2.57]
Mentawai	0.84 [-0.67; 2.35]
New Zealand English	1.65 [0.13; 3.09]
Norwegian	0.83 [-0.33; 1.99]
Nyangatom	1.37 [-0.15; 2.88]
Nu	9.41 [4.22; 21.65]

Supplementary Table 9.1: Model parameter estimates for f_0

Model parameters for f_0 variability	**Estimate**
Standard Deviation of Languages	0.21 [0.01; 0.58]
Standard Deviation of Studies within Languages	0.76 [0.60; 0.95]
Standard Deviation of Measurements	0.10 [0.01; 0.23]
Age (months)	0.00 [-0.01; 0.01]
Australian English	0.49 [0.03; 0.93]

Bislama	0.29 [-0.83; 1.4]
British English	0.57 [-0.45; 1.58]
Canadian English	0.82 [-0.16; 1.77]
Cantonese Chinese	0.51 [-0.93; 1.94]
Danish	1.21 [-0.19; 2.58]
Dutch	0.17 [-1.04; 1.45]
Enga	0.83 [-0.64; 2.3]
Finnish & Swedish	0.4 [-0.87; 1.73]
French	0.13 [-1.22; 1.5]
German	0.14 [-0.72; 1.01]
Hungarian	0.27 [-0.96; 1.51]
Italian	0.4 [-0.7; 1.48]
Japanese	0.55 [-0.16; 1.26]
Kannada	0.54 [-0.93; 2.04]
Kenyan & Fijian	0.55 [-0.81; 1.88]
Korean	0.31 [-0.99; 1.59]
Mandarin Chinese	0.42 [-0.47; 1.3]
Mbendjele	0.81 [-0.53; 2.13]

Mentawai	0.37 [-0.93; 1.67]
New Zealand English	0.55 [-0.7; 1.78]
Norwegian	0.89 [-0.16; 1.93]
Nyangatom	0.53 [-0.83; 1.85]
Polish	0.66 [-0.54; 1.85]
Quechua	-0.03 [-1.35; 1.29]
Quechua & Achuar	0.18 [-1.14; 1.49]
Scottish English	0.45 [-0.86; 1.71]
Spanish	-0.19 [-1.42; 1.06]
Sri Lankan Tamil	0.36 [-1.02; 1.69]
Swedish	0.16 [-1.3; 1.59]
Tagalog	0.26 [-1.02; 1.56]
Toposa	0.57 [-0.79; 1.93]
Tsimane	0.35 [-0.95; 1.67]
US English	0.92 [0.19; 1.56]
Spontaneous Speech	0.39 [0.05; 0.72]
Nu	31.17 [11.06; 67.91]

Supplementary Table 9.2: Model parameter estimates for f_0 variability

Model Parameters for Vowel Space Area	Estimate
Standard Deviation of Languages	0.37 [0.02; 0.87]
Standard Deviation of Studies within Languages	0.61 [0.41; 0.86]
Standard Deviation of Measurements	0.10 [0.00; 0.27]
Age (months)	-0.01 [-0.02; 0.01]
Australian English	0.93 [0.36; 1.42]
Bislama	1.94 [0.5; 3.29]
British English	0.99 [-0.32; 2.28]
Canadian English	1.54 [0.16; 2.83]
Cantonese Chinese	0.13 [-1.07; 1.41]
Danish	0.22 [-0.88; 1.36]
Dutch	-0.33 [-1.52; 0.96]
Enga	0.87 [-0.59; 2.32]
Finnish & Swedish	1.05 [-0.33; 2.37]
French	0.29 [-1.18; 1.79]
German	0.63 [-0.57; 1.86]
Hungarian	0.92 [-0.39; 2.23]

Jamaican English	0.53 [-0.76; 1.9]
Japanese	1.05 [-0.17; 2.23]
Kannada	1.24 [-0.33; 2.82]
Mandarin Chinese	0.84 [-0.14; 1.8]
Mbendjele	1.9 [0.36; 3.28]
Mentawai	0.21 [-1.14; 1.57]
New Zealand English	0.92 [-0.4; 2.2]
Norwegian	0.11 [-0.91; 1.18]
Nyangatom	0.96 [-0.44; 2.32]
Polish	0.41 [-0.82; 1.68]
Quechua	0.43 [-0.9; 1.79]
Quechua Achuar	1.82 [0.34; 3.17]
Russian	0.77 [-0.54; 2.12]
Spanish	0.5 [-0.7; 1.69]
Swedish	1.47 [0.28; 2.6]
Toposa	0.2 [-1.11; 1.56]
Tsimane	1.31 [-0.12; 2.67]
US English	0.66 [-0.24; 1.56]
Nu	18.61 [4.86; 50.08]

Supplementary Table 9.3: Model parameter estimates for vowel space area

Model parameters for articulation rate	Estimate
Standard Deviation of Languages	0.41 [0.02; 1.14]
Standard Deviation of Studies within Languages	0.74 [0.42; 1.19]
Standard Deviation of Measurements	0.23 [0.02; 0.46]
Age (months)	0.02 [0.00; 0.05]
Australian English	-1.19 [-2.14; -0.23]
Bislama	-0.49 [-2.06; 0.97]
British English	-1.53 [-3.32; 0.25]
Canadian English	-0.96 [-2.74; 0.92]
Cantonese Chinese	-1.97 [-3.77; -0.11]
Danish	-2.03 [-3.7; -0.25]
Dutch	-0.69 [-1.97; 0.49]
French	-1.33 [-3.08; 0.45]
German	-1.38 [-2.9; 0.14]
Italian	-1.18 [-2.81; 0.45]
Japanese	-1.65 [-3.42; 0.2]

Kenyan & Fijian	-1.26 [-2.74; 0.25]
Korean	-1.3 [-2.92; 0.32]
Mandarin Chinese	-0.56 [-2.03; 0.82]
Sri Lankan Tamil	-0.79 [-2.41; 0.82]
Tagalog	-1.5 [-3.11; 0.21]
US English	-0.9 [-2.03; 0.07]
Spontaneous Speech	0.95 [-0.08; 1.88]
Nu	23.64 [5.90; 58.56]

Supplementary Table 9.4: Model parameter estimates for articulation rate

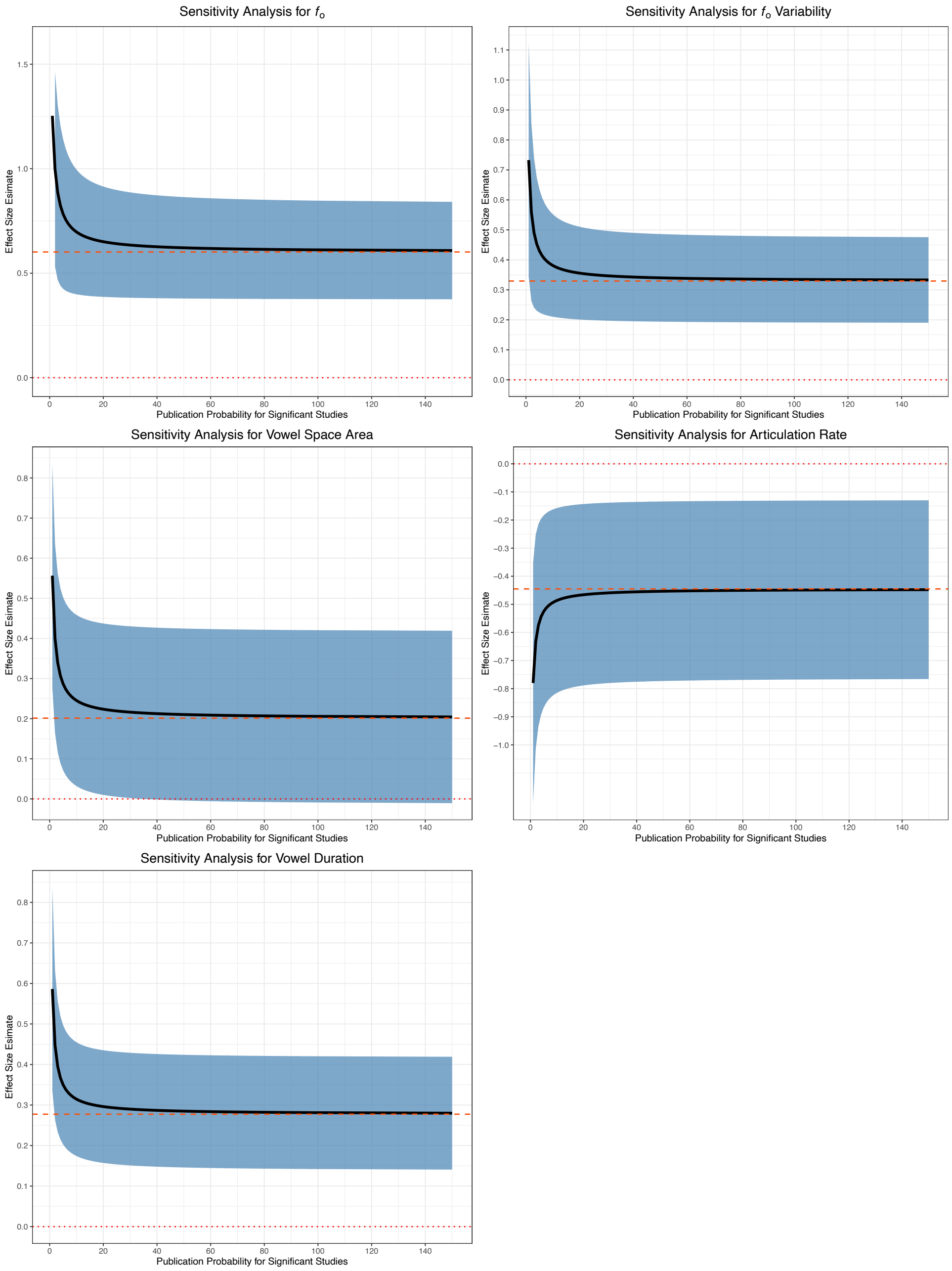
Model parameters for vowel duration	**Estimate**
Standard Deviation of Languages	0.39 [0.02; 1.14]
Standard Deviation of Studies within Languages	0.50 [0.12; 0.92]
Standard Deviation of Measurements	0.16 [0.01; 0.36]
Age (months)	-0.02 [-0.05; 0.01]
Australian English	0.6 [-0.08; 1.21]
British English	1.08 [-0.43; 2.42]
Canadian English	0.88 [-0.49; 2.18]
Danish	0.09 [-1.12; 1.44]

Jamaican English	0.17 [-1.04; 1.5]
Japanese	0.36 [-0.62; 1.36]
Mandarin Chinese	0.92 [-0.23; 1.98]
Norwegian	0.32 [-0.59; 1.35]
Scottish English	0.55 [-0.63; 1.7]
Swedish	-0.29 [-1.34; 1]
US English	0.84 [-0.18; 1.76]
Nu	4.90 [2.19; 11.19]

Supplementary Table 9.5: Model parameter estimates for vowel duration

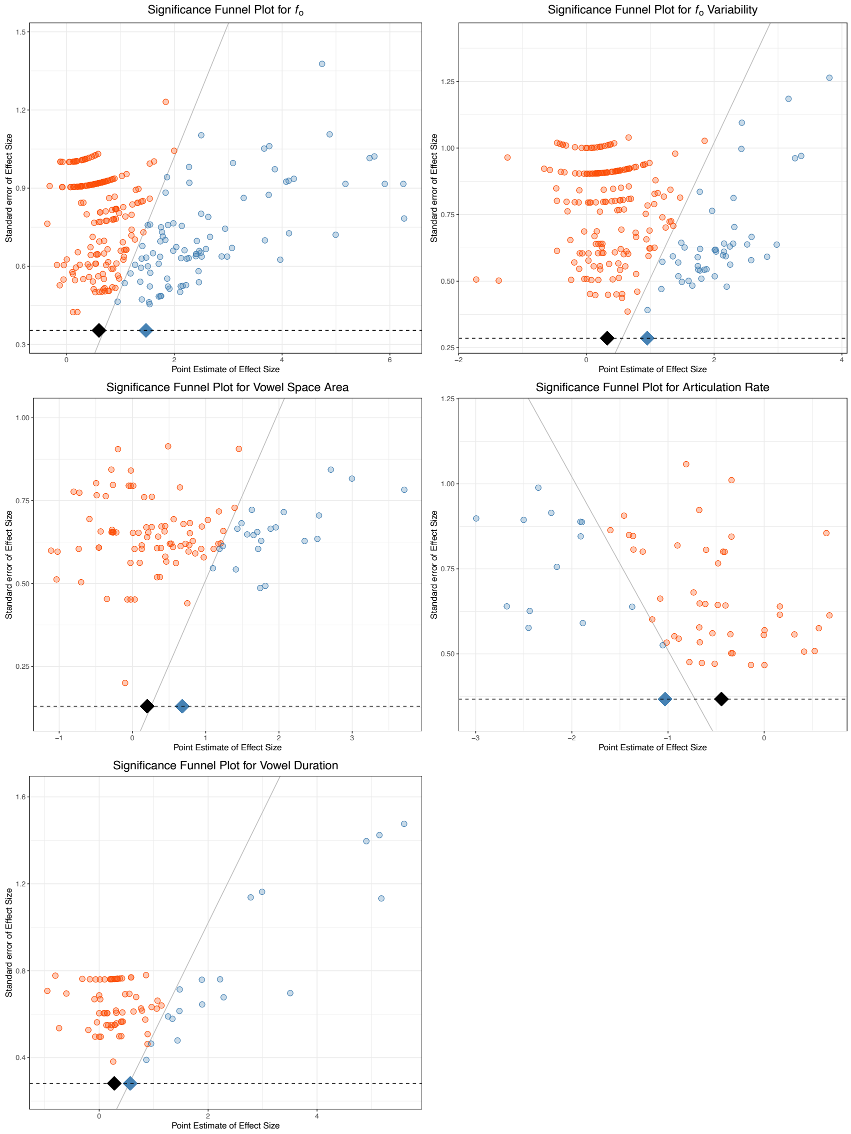
S10: Publication Bias Sensitivity Plots

The plot indicates what happens to the effect size if the publication probability is x times higher for significant studies than for non-significant studies. An effect size estimate of 0.0 is indicated by the orange dotted line, and the worst-case point estimate (see below) is indicated by the dashed red line.



Supplementary Figure 10.1: Sensitivity plots for each acoustic variable, showing the effect size estimate as a function of severity of publication bias.

Studies on the diagonal line have exactly p = 0.05. Black diamond: worst-case estimate of effect size based only on non-significant studies. Blue diamond: estimate of effect size for all studies. These plots help to determine the extent to which the non-affirmative studies' point estimates are systematically smaller than the entire set of point estimates. As a simple heuristic, when the diamonds are close to one another, our quantitative sensitivity analyses will typically indicate that the meta-analysis is fairly robust to publication bias. When the diamonds are distant or if the grey diamond represents a negligible effect size, then our sensitivity analyses may indicate that the meta-analysis is not robust.



Supplementary Figure 10.2: Significance funnel plots for each acoustic variable.

S11: Overview of Languages and Sample Sizes

Language	Total Sample Size
American English	2942
Australian English	1049
Bislama	36
British English	156
Canadian English	96
Cantonese Chinese	80
Danish	170
Dutch	335
French	30
German	710
Hungarian	234
Italian	110
Jamaican English	40
Japanese	1441
Kenyan & Fijian	45
Korean	87
Mandarin Chinese	373
Norwegian	924
Quiche Mayan	3
Russian	10
Scottish English	380
Spanish	17
Sri Lankan Tamil	84
Swedish	86
Tagalog	87

Supplementary Table 11.1 Overview of total sample size (i.e., number of speakers) according to language

S12: Overview of Studies, Measures & Number of Effect Sizes

Study	Measure	Language	Number of Effect Sizes
(Albin_&_Echols_1996)	F0	US English	4
(Amano_et_al_2006)	F0	Japanese	69
(Amano_et_al_2006)	F0V	Japanese	70
(Andruski_&_Kuhl_1996)	F0	US English	2
(Andruski_&_Kuhl_1996)	VD	US English	2
(Ashby_2004)	F0	Australian English	1
(Ashby_2004)	F0V	Australian English	1
(Benders_2013)	F0	Dutch	2
(Benders_2013)	VSA	Dutch	2
(Benders_StGeorge_&_Fletcher_2021)	F0	Dutch	2
(Bergeson_et_al_2006)	F0V	US English	2
(Bergeson_Miller_&_McCune_2006)	F0	US English	2
(Bohn_2013)	F0	Danish	1
(Bohn_2013)	VD	Danish	1
(Bohn_2013)	VSA	Danish	1
(Broen_1972)	AR	US English	2
(Broesch_&_Bryant_2015)	AR	Kenyan & Fijian	1
(Broesch_&_Bryant_2015)	AR	US English	1
(Broesch_&_Bryant_2015)	F0	Kenyan & Fijian	1
(Broesch_&_Bryant_2015)	F0	US English	1
(Broesch_&_Bryant_2015)	F0V	Kenyan & Fijian	1
(Broesch_&_Bryant_2015)	F0V	US English	1
(Broesch_&_Bryant_2018)	AR	Bislama	1
(Broesch_&_Bryant_2018)	AR	US English	1
(Broesch_&_Bryant_2018)	F0	Bislama	1
(Broesch_&_Bryant_2018)	F0	US English	1
(Broesch_&_Bryant_2018)	F0V	Bislama	1
(Broesch_&_Bryant_2018)	F0V	US English	1
(Buckler_Goy_&_Johnson_2018)	F0	Canadian English	1
(Buckler_Goy_&_Johnson_2018)	VD	Canadian English	1
(Burnham_et_al_2002)	F0	Australian English	1

(Burnham_et_al_2002)	VSA	Australian English	1
(Burnham_et_al_2015)	VSA	US English	7
(Church_2002)	AR	Canadian English	1
(Cooper_et_al_1997)	F0	US English	1
(Cooper_et_al_1997)	F0V	US English	1
(Cristia_&_Seidl_2014)	VD	US English	1
(Cristia_&_Seidl_2014)	VSA	US English	4
(Cristia_2010)	AR	US English	2
(Cristia_2010)	F0	US English	2
(Cristia_2010)	F0V	US English	2
(De_Palma_&_VanDam_2017)	F0	US English	2
(Dideriksen_&_Fusaroli_2018)	AR	Danish	1
(Dideriksen_&_Fusaroli_2018)	F0	Danish	1
(Dideriksen_&_Fusaroli_2018)	F0V	Danish	1
(Dideriksen_&_Fusaroli_2018)	VSA	Danish	1
(Dodane_&_Al-Tamimi_2007)	AR	British English	1
(Dodane_&_Al-Tamimi_2007)	AR	French	1
(Dodane_&_Al-Tamimi_2007)	AR	Japanese	1
(Dodane_&_Al-Tamimi_2007)	VSA	British English	1
(Dodane_&_Al-Tamimi_2007)	VSA	French	1
(Dodane_&_Al-Tamimi_2007)	VSA	Japanese	1
(Englund_&_Behne_2006)	VD	Norwegian	6
(Englund_&_Behne_2006)	VSA	Norwegian	6
(Englund_2018)	F0	Norwegian	6
(Englund_2018)	VD	Norwegian	12
(Fernald_&_Simon_1984)	AR	German	1
(Fernald_&_Simon_1984)	F0	German	1
(Fernald_1989)	F0	British English	2
(Fernald_1989)	F0	French	2
(Fernald_1989)	F0	German	2
(Fernald_1989)	F0	Italian	2
(Fernald_1989)	F0	Japanese	2
(Fernald_1989)	F0	US English	2
(Fernald_1989)	F0V	British English	2
(Fernald_1989)	F0V	French	2
(Fernald_1989)	F0V	German	2
(Fernald_1989)	F0V	Italian	2

(Fernald_1989)	F0V	Japanese	2
(Fernald_1989)	F0V	US English	2
(Fisher_&_Tokura_1995)	F0V	US English	1
(Garcia-Sierra_et_al_2021)	VSA	Spanish	1
(Gergely_et_al_2017)	F0	Hungarian	6
(Gergely_et_al_2017)	F0V	Hungarian	6
(Gergely_et_al_2017)	VSA	Hungarian	6
(Grieser_&_Kuhl_1988)	F0	Mandarin Chinese	1
(Grieser_&_Kuhl_1988)	F0V	Mandarin Chinese	1
(Han_De_Jong_&_Kager_2020)	F0	Dutch	2
(Han_De_Jong_&_Kager_2020)	F0	Mandarin Chinese	2
(Han_De_Jong_&_Kager_2020)	F0V	Dutch	2
(Han_De_Jong_&_Kager_2020)	F0V	Mandarin Chinese	2
(Han_et_al_2021)	AR	Dutch	2
(Han_et_al_2021)	AR	Mandarin Chinese	2
(Hartman_2013)	VD	US English	1
(Hartman_2013)	VSA	US English	2
(Hartman_Ratner_&_Newman_2017)	VD	US English	3
(Hartman_Ratner_&_Newman_2017)	VSA	US English	4
(Hilton_et_al_2022)	F0	Bislama	1
(Hilton_et_al_2022)	F0	Canadian English	1
(Hilton_et_al_2022)	F0	Enga	1
(Hilton_et_al_2022)	F0	Finnish & Swedish	1
(Hilton_et_al_2022)	F0	Kannada	1
(Hilton_et_al_2022)	F0	Mandarin Chinese	2
(Hilton_et_al_2022)	F0	Mbendjele	1
(Hilton_et_al_2022)	F0	Mentawai	1
(Hilton_et_al_2022)	F0	New Zealand English	1
(Hilton_et_al_2022)	F0	Nyangatom	1
(Hilton_et_al_2022)	F0	Polish	3
(Hilton_et_al_2022)	F0	Quechua & Achuar	1
(Hilton_et_al_2022)	F0	Spanish	2
(Hilton_et_al_2022)	F0	Toposa	1
(Hilton_et_al_2022)	F0	Tsimane	1
(Hilton_et_al_2022)	F0	US English	1
(Hilton_et_al_2022)	F0V	Bislama	1
(Hilton_et_al_2022)	F0V	Canadian English	1

(Hilton_et_al_2022)	F0V	Enga	1
(Hilton_et_al_2022)	F0V	Finnish & Swedish	1
(Hilton_et_al_2022)	F0V	Kannada	1
(Hilton_et_al_2022)	F0V	Mandarin Chinese	2
(Hilton_et_al_2022)	F0V	Mbendjele	1
(Hilton_et_al_2022)	F0V	Mentawai	1
(Hilton_et_al_2022)	F0V	New Zealand English	1
(Hilton_et_al_2022)	F0V	Nyangatom	1
(Hilton_et_al_2022)	F0V	Polish	3
(Hilton_et_al_2022)	F0V	Quechua	1
(Hilton_et_al_2022)	F0V	Quechua & Achuar	1
(Hilton_et_al_2022)	F0V	Spanish	2
(Hilton_et_al_2022)	F0V	Toposa	1
(Hilton_et_al_2022)	F0V	Tsimane	1
(Hilton_et_al_2022)	F0V	US English	1
(Hilton_et_al_2022)	VSA	Bislama	1
(Hilton_et_al_2022)	VSA	Canadian English	1
(Hilton_et_al_2022)	VSA	Enga	1
(Hilton_et_al_2022)	VSA	Finnish & Swedish	1
(Hilton_et_al_2022)	VSA	Kannada	1
(Hilton_et_al_2022)	VSA	Mandarin Chinese	2
(Hilton_et_al_2022)	VSA	Mbendjele	1
(Hilton_et_al_2022)	VSA	Mentawai	1
(Hilton_et_al_2022)	VSA	New Zealand English	1
(Hilton_et_al_2022)	VSA	Nyangatom	1
(Hilton_et_al_2022)	VSA	Polish	3
(Hilton_et_al_2022)	VSA	Quechua	2
(Hilton_et_al_2022)	VSA	Quechua & Achuar	2
(Hilton_et_al_2022)	VSA	Spanish	2
(Hilton_et_al_2022)	VSA	Toposa	1
(Hilton_et_al_2022)	VSA	Tsimane	1
(Hilton_et_al_2022)	VSA	US English	1
(Igarashi_et_al_2013)	F0	Japanese	3
(Igarashi_et_al_2013)	F0V	Japanese	3
(Ikeda_&_Masataka_1999)	F0	Japanese	1
(Ikeda_&_Masataka_1999)	F0V	Japanese	1
(Inoue_et_al_2011)	F0	Japanese	1

(Inoue_et_al_2011)	F0V	Japanese	1
(Inoue_et_al_2011)	VD	Japanese	1
(Kalashnikova_&_Burnham_2018)	F0	Australian English	5
(Kalashnikova_&_Burnham_2018)	VSA	Australian English	5
(Kalashnikova_et_al_2017)	VSA	Australian English	1
(Kalashnikova_et_al_2020)	F0	Australian English	1
(Kalashnikova_et_al_2020)	VSA	Australian English	1
(Kempe_2009)	F0V	Scottish English	2
(Kempe_2009)	VD	Scottish English	2
(Kondaurova_&_Bergeson_2011)	F0	US English	6
(Kondaurova_&_Bergeson_2011)	VD	US English	6
(Kondaurova_Bergeson_&_Dilley_2012)	VSA	US English	1
(Kondaurova_et_al_2012)	VD	US English	4
(Kondaurova_et_al_2013)	AR	US English	9
(Kondaurova_et_al_2013)	F0	US English	9
(Kondaurova_et_al_2013)	F0V	US English	11
(Kuhl_et_al_1997)	VSA	Russian	1
(Kuhl_et_al_1997)	VSA	Swedish	1
(Kuhl_et_al_1997)	VSA	US English	1
(Lahey_&_Ernestus_2014)	AR	Dutch	1
(Lam_&_Kitamura_2012)	F0	Australian English	1
(Lam_&_Kitamura_2012)	F0V	Australian English	3
(Lam_&_Kitamura_2012)	VD	Australian English	3
(Lam_&_Kitamura_2012)	VSA	Australian English	1
(Lee_et_al_2014)	AR	Australian English	5
(Lee_et_al_2014)	F0	Australian English	5
(Liu_et_al_2009)	F0V	Mandarin Chinese	1
(Liu_et_al_2009)	VD	Mandarin Chinese	1
(Liu_Kuhl_&_Tsao_2003)	VSA	Mandarin Chinese	2
(Liu_Tsao_&_Kuhl_2007)	F0	Mandarin Chinese	1
(Liu_Tsao_&_Kuhl_2007)	VD	Mandarin Chinese	1
(Liu_Tsao_&_Kuhl_2009)	F0	Mandarin Chinese	1
(Liu_Tsao_&_Kuhl_2009)	VSA	Mandarin Chinese	1
(Lovcevic_et_al_2020)	F0V	Australian English	1
(Lovcevic_et_al_2020)	VD	Australian English	6
(Lovcevic_et_al_2020)	VSA	Australian English	2
(Lovcevic_Kalashnikova_&_Burnham_2020)	F0	Australian English	6

(Marklund_&_Gustavsson_2020)	VD	Swedish	1
(Marklund_&_Gustavsson_2020)	VSA	Swedish	1
(Masataka_1992)	F0	Japanese	1
(Masataka_1992)	F0V	Japanese	1
(McMurray_et_al_2013)	F0	US English	1
(McMurray_et_al_2013)	F0V	US English	1
(Miyazawa_et_al_2017)	F0	Japanese	1
(Miyazawa_et_al_2017)	F0V	Japanese	1
(Miyazawa_et_al_2017)	VD	Japanese	1
(Miyazawa_et_al_2017)	VSA	Japanese	1
(Naoi_et_al_2012)	F0	Japanese	1
(Naoi_et_al_2012)	F0V	Japanese	1
(Narayan_&_McDermott_2016)	AR	Korean	6
(Narayan_&_McDermott_2016)	AR	Sri Lankan Tamil	6
(Narayan_&_McDermott_2016)	AR	Tagalog	6
(Narayan_&_McDermott_2016)	F0	Korean	6
(Narayan_&_McDermott_2016)	F0	Sri Lankan Tamil	6
(Narayan_&_McDermott_2016)	F0	Tagalog	6
(Narayan_&_McDermott_2016)	F0V	Korean	6
(Narayan_&_McDermott_2016)	F0V	Sri Lankan Tamil	6
(Narayan_&_McDermott_2016)	F0V	Tagalog	6
(Niwano_&_Sugai_2002)	F0	Japanese	4
(Niwano_&_Sugai_2002)	F0V	Japanese	4
(Ogle_&_Maidment_1993)	F0	British English	1
(Ogle_&_Maidment_1993)	F0V	British English	1
(Outters_et_al_2020)	F0V	German	2
(Outters_Schreiner_Behne_&_Mani_2020)	F0	German	2
(Raneri_2015)	AR	US English	4
(Raneri_2015)	F0	US English	4
(Raneri_2015)	F0V	US English	4
(Räsänen_et_al_2017)	F0	Canadian English	1
(Räsänen_et_al_2017)	F0V	Canadian English	1
(Ratner_&_Pye_1984)	F0	Quiche Mayan	1
(Ratner_&_Pye_1984)	F0	US English	1
(Rattanasone_Burnham_&_Reilly_2013)	VSA	Cantonese Chinese	6
(Rosslund_et_al_2021)	F0	Norwegian	1
(Rosslund_et_al_2021)	F0V	Norwegian	1

(Rosslund_et_al_2021)	VD	Norwegian	1
(Rosslund_et_al_2021)	VSA	Norwegian	1
(Sheehan_2008)	AR	US English	2
(Sheehan_2008)	F0	US English	4
(Sheehan_2008)	F0V	US English	4
(Shute_&_Wheldall_1989)	F0	British English	2
(Shute_&_Wheldall_2001)	F0	British English	2
(Shute_&_Wheldall_2001)	F0V	British English	2
(Smith_&_Strader_2014)	F0	US English	1
(Smith_&_Strader_2014)	F0V	US English	1
(Steen_&_Englund_2021)	F0	Norwegian	6
(Steen_&_Englund_2021)	F0V	Norwegian	6
(Steen_&_Englund_2021)	VD	Norwegian	6
(Steen_&_Englund_2021)	VSA	Norwegian	2
(Stern_et_al_1983)	F0V	US English	4
(Sulpizio_et_al_2018)	AR	German	1
(Sulpizio_et_al_2018)	AR	Italian	1
(Sulpizio_et_al_2018)	F0	German	1
(Sulpizio_et_al_2018)	F0	Italian	1
(Sulpizio_et_al_2018)	F0V	German	1
(Sulpizio_et_al_2018)	F0V	Italian	1
(Sundberg_1998)	F0	Swedish	1
(Sundberg_1998)	F0V	Swedish	1
(Sundberg_1998)	VD	Swedish	2
(Tajima_et_al_2013)	VD	Japanese	6
(Tang_&_Maidment_1996)	AR	Cantonese Chinese	1
(Tang_&_Maidment_1996)	F0V	Cantonese Chinese	1
(Tang_et_al_2017)	VD	Mandarin Chinese	6
(Tang_et_al_2017)	VSA	Mandarin Chinese	1
(Trainor_Austin_&_Desjardins_2000)	F0V	Canadian English	1
(Trainor_et_al_2000)	F0	Canadian English	1
(Uther_et_al_2007)	F0	British English	1
(Uther_et_al_2007)	VD	British English	1
(Uther_et_al_2007)	VSA	British English	1
(VanDam_&_De_Palma_2014)	F0	US English	2
(Vosoughi_&_Roy_2012)	F0	US English	5
(Vosoughi_&_Roy_2012)	F0V	US English	5

(Vosoughi_&_Roy_2012)	VD	US English	5
(Wassink_et_al_2007)	VSA	Jamaican English	1
(Wassink_Wright_&_Franklin_2007)	F0	Jamaican English	1
(Wassink_Wright_&_Franklin_2007)	VD	Jamaican English	2
(Weirich_&_Simpson_2019)	F0	German	12
(Weirich_&_Simpson_2019)	F0V	German	12
(Weirich_&_Simpson_2019)	VSA	German	12
(Wieland_et_al_2015)	VSA	US English	2
(Xu_et_al_2013)	VSA	Australian English	1

Supplementary Table 12.1 Overview of studies, measures, languages and number of effect sizes

Study II

Infant-Directed Speech Does Not Always Involve Exaggerated Vowel Distinctions: Evidence From Danish

Christopher Cox[1, 2, 3], Christina Dideriksen[1, 2], Tamar Keren-Portnoy[3], Andreas Roepstorff [1, 2],
Morten H. Christiansen[1, 2, 4], Riccardo Fusaroli[1, 2]

[1] School of Communication and Culture, Aarhus University,
Jens Chr. Skous Vej 2, Building 1485, 8200 Aarhus, Denmark

[2] Interacting Minds Centre, Aarhus University,
Jens Chr. Skous Vej 4, Building 1483, 8200 Aarhus, Denmark

[3] Department of Language and Linguistic Science, University of York,
Vanbrugh College, Heslington, York YO10 5 DD, United Kingdom

[4] Department of Psychology, Cornell University,
Ithaca, NY 14853, United States of America

Cox, C., 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, 1672–1696. https://doi.org/10.1111/cdev.13950

Acknowledgements

This research was supported in part by the Danish Council for Independent Research (FKK Grant DFF-7013-00074) to MHC. This project also received a seed funding grant from the Interacting Minds Centre. All of the computation done for this project was performed on the UCloud interactive HPC system, which is managed by the eScience Center at the University of Southern Denmark. We would also like to thank Ocke-Schwen Bohn for his insightful comments on an earlier version of this paper. The data necessary to reproduce the analyses presented here are publicly accessible at the following URL: https://osf.io/ywf9m/. The analytic code necessary to reproduce the analyses presented in this paper is publicly accessible at: https://osf.io/ywf9m/. The materials (i.e., videos and audio files) necessary to attempt to replicate the findings presented here are not publicly accessible. The analyses presented here were not preregistered.

Correspondence: chris.mm.gmail.com (Chris Cox)

University of York
York Graduate Research School
Research Degree Thesis Statement of Authorship

Note that where a paper has multiple authors, the statement of authorship can focus on the key contributing/corresponding authors.

Candidate name	Christopher Martin Mikkelsen Cox
Department	Department of Linguistics & Cognitive Science
Thesis title	How Social Contingency Shapes Early Child-Caregiver Interactions

Title of the work (paper/chapter)	Infant-directed speech does not always involve exaggerated vowel distinctions: Evidence from Danish	
Publication status	**Published**	X
	Accepted for publication	
	Submitted for publication	
	Unpublished and unsubmitted	
Citation details (if applicable)	Cox, C., Dideriksen, C., Keren-Portnoy, T., Roepstorff, A., Christiansen, M. H., & Fusaroli, R. (2022). Infant-directed speech does not always involve exaggerated vowel distinctions: Evidence from Danish. *Child Development* 94, 1672-1696.	

Description of the candidate's contribution to the work*	Conceptualization, Methodology, Software, Validation, Formal analysis, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision, Project administration
Approximate percentage contribution of the candidate to the work	70 %
Signature of the candidate	Chris Cox
Date (DD/MM/YY)	24-11-23

<u>Co-author contributions</u>

By signing this Statement of Authorship, each co-author agrees that:
(i) the candidate has accurately represented their contribution to the work;
(ii) if required, permission is granted for the candidate to include the work in their thesis (note that this is separate from copyright considerations).

Name of co-author	Morten H. Christiansen
Contact details of co-author	Department of Psychology, Cornell University, Ithaca, NY 14853, USA
Description of the co-author's contribution to the work*	Commented and edited various drafts of the paper and helped with the review process
Approximate percentage contribution of the co-author to the work	10%
Signature of the co-author	
Date (DD/MM/YY)	24/11/23

Name of co-author	Christina Dideriksen
Contact details of co-author	dideriksenchristina@gmail.com
Description of the co-author's contribution to the work*	Data Collection
Approximate percentage contribution of the co-author to the work	5%
Signature of the co-author	
Date (DD/MM/YY)	28-11-23

Name of co-author	Tamar Keren-Portnoy
Contact details of co-author	tamar.keren-portnoy@york.ac.uk
Description of the co-author's contribution to the work*	Read and commented on several versions of the paper
Approximate percentage contribution of the co-author to the work	5%

Signature of the co-author	
Date (DD/MM/YY)	3-12-23

Name of co-author	Riccardo Fusaroli
Contact details of co-author	fusaroli@cc.au.dk
Description of the co-author's contribution to the work*	Conceptualization, Methodology, Software, Validation, Formal analysis, Resources, Data Curation, Writing - Original Draft, Visualization, Project administration
Approximate percentage contribution of the co-author to the work	10%
Signature of the co-author	
Date (DD/MM/YY)	24-11-23

Infant-Directed Speech Does Not Always Involve Exaggerated Vowel Distinctions: Evidence From Danish

Abstract

This study compared the acoustic properties of 26 (100% female, 100% monolingual) Danish caregivers' spontaneous speech addressed to their 11-24-month-old infants (infant-directed speech, IDS) and an adult experimenter (adult-directed speech, ADS). The data were collected between 2016-2018 in Aarhus, Denmark. Prosodic properties of Danish IDS conformed to cross-linguistic patterns, with a higher pitch, greater pitch variability and slower articulation rate than ADS. However, an acoustic analysis of vocalic properties revealed that Danish IDS had a reduced or similar vowel space, higher within-vowel variability, raised formants, and lower degree of vowel discriminability compared to ADS. None of the measures, except articulation rate, showed age-related differences. These results push for future research to conduct theory-driven comparisons across languages with distinct phonological systems.

Introduction

A mother sits on the floor with her 13-month-old infant. The mother points towards a cuddly toy and says in a high-pitched, animated voice: *"Look! A penguin! Do you want to pet the penguin?"*. The mother picks up a stuffed animal from the floor, holds it in front of the infant and repeats "*Look! A penguin!*" The infant then looks towards the mother and provides a rough attempt at repeating the word. The mother replies in an excited tone: "*Yes, it's a penguin!*".

In the above scenario, we might expect the caregiver to address her child using a spontaneous form of language known as infant-directed speech (IDS). Across a diverse intersection of languages and cultures, the form of speech that adults direct to infants differs from that directed to other adults (i.e., adult-directed speech, ADS) in systematic ways (Cox, Bergmann, et al., 2022; Hilton et al., 2022). The acoustic characteristics of IDS have been studied extensively, and some clear commonalities have emerged across a wide variety of languages and cultures (e.g., Broesch & Bryant, 2015; Fernald et al., 1989; Gergely et al., 2017). For example, caregivers increase their vocal pitch and pitch variability, slow down their speech, and produce acoustically exaggerated vowels (e.g., Broesch & Bryant, 2015; Han et al., 2020; Kalashnikova et al., 2017; Kuhl et al., 1997). Many of the acoustic properties of IDS change as the infant becomes older in ways that allow caregivers to scaffold infants' early social and linguistic development (Cox, Bergmann, et al., 2022; Fusaroli et al., 2019; Fusaroli, Weed, et al., 2021; Warlaumont et al., 2014). This has led to the proposal that this speech style may serve various functions, such as allowing caregivers to express affect (cf., Benders, 2013), to regulate infants' emotional states (cf., Bryant & Barrett, 2007; Kitamura & Lam, 2009; Papoušek et al., 1991) as well as to facilitate language development (cf., Golinkoff et al., 2015; Kuhl, 2000). However, some questions about IDS require further investigation: i) To what extent do common IDS properties generalise across a more diverse set of languages (cf., Blasi et al., 2022; Christiansen et al., 2022; Kidd & Garcia, 2022)? ii) To what extent does the acoustic expression of IDS exhibit dynamic changes across the span of early development? iii) Does IDS provide a clearer input than ADS, and how does this relate to the internal structure of vowel categories? In the remainder of this introduction, we detail each of these points and explain the aims of this comparative acoustic study of Danish IDS and ADS.

The extent to which the acoustic expression of IDS exhibits similar properties across languages with different phonological structures (e.g., Broesch & Bryant, 2015; Fernald et al., 1989; Han et al., 2020) and cultures with different parenting behaviours (e.g., Bergelson et al., 2019; Casillas et al., 2020) remains under-investigated. A recent meta-analysis of the acoustic properties of IDS found a high degree of between-study variability, part of which may arise due to differences in phonological systems across languages as well as methodological differences across studies of IDS (Cox, Bergmann, et al., 2022). Across these studies, there was a strong tendency for caregivers to produce IDS with a slower articulation rate as well as an elevated and more variable pitch across a wide variety of languages and cultures (Broesch & Bryant, 2015; Han et al., 2020; Narayan & McDermott, 2016). However, there were significant cross-linguistic differences in the degree to which caregivers produced acoustically exaggerated vowels (Cox, Bergmann, et al., 2022). Some languages, for example, showed a vowel space reduction in IDS (e.g., Benders, 2013; Englund & Behne, 2005; Rattanasone et al., 2013) and a higher degree of within-vowel variability (Cristia & Seidl, 2014; Martin et al., 2015; Miyazawa et al., 2017; Rosslund et al., 2022). The overrepresentation of North American English in all of these studies (Christiansen et al., 2022), which appears to provide a particularly exaggerated form of IDS (Fernald et al., 1989; Floccia et al., 2016), may inflate the importance of these findings. New data from a more diverse set of human cultures and languages (cf., Kidd & Garcia, 2022) can thus likely offer new insights into the possible forms of IDS in early infant development.

The current study focuses on Danish, a language characterised by a high degree of phonetic reduction, schwa assimilation, and a wide variety of vocalic sounds (cf., Basbøll, 2005; Højen & Nazzi, 2016; Trecca et al., 2021). This peculiar sound structure means that speakers of Danish often produce a speech stream with few or no clear spectral discontinuities to allow infants to engage in word segmentation (cf., Trecca et al., 2019, 2021). This in turn has led researchers to posit that Danish may be particularly difficult for infants to learn (Bleses et al., 2008a, 2008b, 2011), and even result in a vowel bias in word learning rather than a consonant bias (Højen & Nazzi, 2016). Looking-while-listening studies with Danish two-and-half-year-olds have shown that the vowel-rich nature of Danish has a negative impact on the processing of both known and novel words (Trecca et al., 2018, 2020). These findings provide potential explanations for the tendency for Danish infants to fall behind on some early linguistic milestones compared to children learning other languages (Bleses et al., 2008a, 2008b, 2011).

The peculiar sound structure of Danish raises the question as to how caregivers modify the acoustic properties of IDS. Two earlier investigations of IDS in Danish have painted a puzzling picture (Bohn, 2013; Dideriksen & Fusaroli, 2018). Both studies reported that IDS exhibited a slower articulation rate as well as an elevated f_o variability; however, the studies also reported either similar degree of vowel separability (Bohn, 2013) or even a slight reduction in vowel separability in IDS compared to ADS (Dideriksen & Fusaroli, 2018). The measure of between-vowel separability in these studies was derived by calculating the total area enclosed by steady-state formant frequencies of the three peripheral /i/-/a/-/u/ vowels, which has been posited to be a measure of speech clarity (J. Lam et al., 2012; Whitfield & Mehta, 2019), as discussed further below. The vowel inventory of Danish consists of ten monophthongal vowel phonemes, seven of which are positioned in the front region of the vowel space (Grønnum, 1998). Because there are both short and long versions of each of these vowel phonemes – and only the long versions can be combined with the suprasegmental, creaky-voice feature of *stød* – the vowel inventory is estimated as comprising 30 phonologically distinct vowels, although estimates of this number differ depending on how the vowels are counted (Trecca et al., 2021). Given the high number of vowels that Danish-learning infants have to acquire, we might have expected caregivers (albeit unconsciously) to increase the separability between IDS vowel phonemes to provide a clearer speech input for their infants. However, these previous studies of Danish IDS provided a limited picture of its acoustic expression. Firstly, Bohn (2013) presented results from a laboratory experiment with a high degree of experimental control. Parents were asked to talk about explicitly labelled toy animals to elicit specific words containing the /i/, /u/ and /a/ vowel tokens. Because the type of experimental task used to study IDS can influence the strength of the vowel space area effect (Cox, Bergmann, et al., 2022; cf., Marklund & Gustavsson, 2020), it is unclear whether this finding extends to spontaneous speech in Danish. Secondly, Dideriksen and Fusaroli (2018) used a coarse-grained approach to formant estimation by basing their data on automatic extractions of any type of voiced segment from the speech signal (cf., Degottex et al., 2014), which might unwittingly introduce biases in the estimates. To build upon the groundwork laid by these studies, we i) relied on data collected in naturalistic settings, with caregivers and infants engaging in the types of interaction that take place in day-to-day activities, and ii) conducted a detailed acoustic analysis of the internal distributions of each of the vowel categories in Danish IDS and ADS.

One of the benefits of IDS may derive from caregivers' dynamic adaptation to infants' developmental states (Fusaroli et al., 2019; Smith & Trainor, 2008) and responsiveness to infant vocalisations (Fusaroli, Weed, et al., 2021; Goldstein & Schwade, 2008; Ko et al., 2016; Nguyen et al., 2022; Warlaumont et al., 2014). Cross-linguistic studies of changes in the acoustic properties of IDS as a function of infant age, for example, indicated that caregivers increase their rate of articulation (Kondaurova et al., 2013; Lee et al., 2014; Narayan & McDermott, 2016; Raneri, 2015) and reduce the median f_o in IDS as their infants become older (Gergely et al., 2017; Han et al., 2020; Kondaurova et al., 2013; Niwano & Sugai, 2002; Vosoughi & Roy, 2012). Most studies on the properties of f_0 variability and vowel space area, on the other hand, did not find evidence of any developmental shifts across a wide variety of age ranges (Benders, 2013; Burnham et al., 2015; Cristia & Seidl, 2014; Gergely et al., 2017; Hartman, 2013; Kalashnikova & Burnham, 2018; Lovcevic et al., 2020; Weirich & Simpson, 2019; Wieland et al., 2015). Two studies of vowel space area did show a shift over time, but contradicted each other in terms of the direction of the shift; whereas Japanese-speaking caregivers were reported to exhibit a gradual vowel space expansion as the infant became older (Dodane & Al-Tamimi, 2007), Cantonese Chinese-speaking caregivers were reported to reduce their vowel space as a function of infant age (Rattanasone et al., 2013). These acoustic modifications in IDS according to infants' preferences (Kitamura & Burnham, 1998; Singh et al., 2002) and processing limitations (Christiansen & Chater, 2016; Saffran & Kirkham, 2018) suggest that IDS involves interactive reciprocity, where active participation and reciprocity both play a crucial role (Goldstein & Schwade, 2008; Ko et al., 2016; Nguyen et al., 2022; Warlaumont et al., 2014). Investigating which aspects of the acoustic expression of IDS changes in the span of early infant development allows crucial insight into the different functions served by IDS (Bryant, 2022; Bryant & Barrett, 2007; Cox, Bergmann, et al., 2022). The question of how each of the acoustic features of Danish IDS undergo change with infant age will also be investigated in the current cross-sectional sample of participants.

A large body of research shows that the speech directed to infants has properties that facilitate speech processing (García-Sierra et al., 2016, 2021). For example, IDS stimuli can facilitate neural processing (Peter et al., 2016), induce faster word recognition (Song et al., 2010) and produce better word segmentation of an artificial language composed of nonsense syllables (Thiessen et al., 2005). The extent to which parents produce acoustically exaggerated vowels in IDS has been shown to have an effect on concurrent speech discrimination skills (García-Sierra et al., 2016, 2021; H. Liu et al., 2003), later expressive vocabulary (Hartman et al., 2017; Kalashnikova & Burnham, 2018), and the complexity of vocalizations at a later point in time (Marklund et al., 2021). Other descriptive studies showed positive correlations between pitch modulations in IDS and expressive vocabulary size (Porritt et al., 2014; Rosslund et al., 2022; Spinelli et al., 2017).

The facilitatory effects of IDS are generally attributed to its tendency to increase the clarity of the speech addressed to children (Hartman et al., 2017; Kuhl et al., 1997; H. Liu et al., 2003). Studies of adult speech perception have indicated a clear relation between vowel space expansion and speech intelligibility, the majority of which are conducted on American English (Bradlow et al., 1996, 2003; Ferguson & Kewley-Port, 2002, 2007; J. Lam et al., 2012; H.-M. Liu et al., 2005; Whitfield & Goberman, 2017; Whitfield & Mehta, 2019). Computational models also learned vowel categories better if the location of category centroids were more distant from each another (De Boer & Kuhl, 2003; Eaves et al., 2016; McMurray et al., 2009; Vallabha et al., 2007). Other experiments showed that measures of speech clarity correlated with articulation rate and speech segment durations, suggesting that temporal aspects of speech may also play an important role (Ferguson & Kewley-Port, 2007; J. Lam & Tjaden, 2013; Searl & Evitts, 2013).

There is substantial evidence that vowel space areas tend to be larger in IDS than in ADS (e.g., Cristia & Seidl, 2014; Gergely et al., 2017; Kalashnikova & Burnham, 2018; C. Lam & Kitamura, 2012; Marklund & Gustavsson, 2020; Miyazawa et al., 2017; Weirich & Simpson, 2019); however, a number of languages, such as Dutch (Benders, 2013), Norwegian (Englund, 2018; Englund & Behne, 2005; Steen & Englund, 2021) and Cantonese (Rattanasone et al., 2013) show vowel space reduction in IDS. Although a lot of the literature on speech clarity is based on vowel space area, a number of authors have raised concerns about potential limitations to traditional vowel space measures (cf., Whitfield & Goberman, 2014, 2017). One of these limitations is that studies relying only on the acoustic measure of vowel space area disregard the crucial assumption that vowel space expansion in and of itself does not entail a greater degree of speech clarity. A number of studies have reported an increased degree of within-vowel variability in IDS (Cristia & Seidl, 2014; Martin et al., 2015; McMurray et al., 2013; Miyazawa et al., 2017; Rosslund et al., 2022). For example, Miyazawa et al. (2017) showed that Japanese mothers expanded the vowel space area of the IDS to their 18-20-month-old infants; however, due to an increase in within-vowel variability, this expansion did not lead to more distinct categories compared to those in ADS. It is thus only by computing the extent of within-vowel variability and between-vowel discriminability that claims about the clarity and overlap of phonetic categories can be made. The limitations of considering vowel space area alone thus makes it unclear whether the facilitative properties of IDS should be attributed to vowel space expansion, or whether this effect arises as an unintended side effect of another acoustic variable, such as vocal tract shortening through raising of the larynx (Kalashnikova et al., 2017) or smiling (Englund, 2018), or whether it simply occurs as a side effect of a slower articulation rate and therefore being able to reach articulatory targets (Whitfield & Goberman, 2017). To provide a fuller picture of the extent of within-vowel variability and between-vowel discriminability in Danish IDS and ADS, we manually annotated 9267 individual vowel tokens and analysed the extent of compactness of individual vowel categories.

The present study was designed to provide a comprehensive analysis of the input from which Danish infants learn language by comparing the prosodic and vocalic properties of Danish caregivers' spontaneous IDS and ADS. To build on the insights from previous studies of Danish IDS (Bohn, 2013; Dideriksen & Fusaroli, 2018) and IDS across distinct languages (Cox, Bergmann, et al., 2022), we engaged in cumulative science practices and incorporated statistical results from these studies into our models to compare the findings. We turned our focus to the following three questions concerning the acoustic expression of Danish IDS and ADS:

i) To what extent are five acoustic properties that have been reported extensively in the literature on IDS (Cox, Bergmann, et al., 2022) expressed in Danish ADS and IDS – fundamental frequency (f_o), f_o variability, articulation rate, vowel duration, and vowel space area?

ii) Do any of the acoustic properties in Danish IDS exhibit age-related changes over the course of early infant development?

iii) How do Danish parents negotiate the balance between within-vowel variability and between-vowel separability? Do vowel categories in IDS exhibit a higher degree of discriminability than in ADS?

Methods

Participants & Recording Context

The speech recordings for this study consisted of spontaneous speech from 26 Danish-speaking mothers of infants between the ages of 11 and 24 months. The data overlapped somewhat with the dataset of a previous study of Danish IDS (Dideriksen & Fusaroli, 2018), which provided a coarse-grained analysis with a subset of the participants (N = 10) in this dataset. Based on a precision analysis of a sample size of 26 participants (cf. Figure S5.1 and S5.2 in the supplementary materials), repeated measures from this size of sample provided sufficient data to obtain credible intervals that allowed us to draw meaningful conclusions for the moderate effect sizes (Hedges' *g*

$\simeq 0.5$) we expected from previous meta-analyses of the field (Cox, Bergmann, et al., 2022). The precision analysis indicated that – assuming a moderate effect size of 0.5 – a within-subjects design with 26 participants allowed estimates with a standard deviation of 0.05. This was quite sufficient to draw robust conclusions about potential differences between the speech styles. The mothers spoke Danish as their first language, did not report any health problems, and their infants were not at risk for any linguistic or cognitive disabilities. All mothers provided informed consent. Each mother participated in two recording sessions: i) to elicit spontaneous IDS, the mother was recorded for 60 minutes in a free play session with her infant, and ii) to elicit spontaneous ADS, the mother participated in a casual conversation about summer holiday plans with an experimenter for 15 minutes. To increase the level of ecological validity, we took care to make the recording settings as natural as possible in order to capture the types of interaction that take place in day-to-day activities. To ensure that infants and caregivers felt comfortable, the recording sessions took place in participants' homes and used two video cameras (a Panasonic HDC-HS700 and GoPro Hero 5) and a high-quality wearable microphone. The experimenter interacted with the mother to elicit ADS, but remained absent during IDS recordings. Each mother was instructed to interact with her child as she would normally do in an everyday situation. The 26 infants (10 female) were aged between 11 and 24 months of age (mean = 16 months, SD = 4.4), as shown in Table S0.1 in the supplementary materials.

Acoustic Analysis

Diarisation of the Speech Samples

In order to partition the IDS speech recordings into homogenous segments and identify utterance boundaries according to speaker identity, we used an open-source diarisation tool known as ALICE (i.e., Automatic LInguistic unit Count Estimator) (Räsänen, Seshadri, Lavechin, et al., 2021). This software provides automatic diarisation using a voice type classifier trained on over 200 hours of day-long audio recordings of infants learning a number of typologically distinct languages (Räsänen, Seshadri, Lavechin, et al., 2021). This voice type classifier uses the pyannote-audio library (Lavechin et al., 2020) and a neural network architecture called SincNet (Ravanelli & Bengio, 2018). Using this neural network architecture, the diarisation tool extracts talker information from an input audio waveform and classifies utterances into talker categories. For the IDS recordings, we thus relied on this automatic diarisation process and retained only segments spoken by the caregiver.

Because the ADS recordings involved two female interactants (i.e., the caregiver and experimenter), there would be a high risk of confusing the two voices in the automatic diarisation process. We therefore transcribed all the audio files for the ADS speech samples and subsequently discarded utterances from the experimenter. An utterance was defined as a vocal production by one speaker that i) was not interrupted by another speaker (e.g., backchannels were not considered) and ii) did not contain a pause longer than a second. In most cases, the infant was in the room during the ADS recordings; however, interruptions from the infant were rare and all utterances from the infant as well as those directed toward the infant were discarded from the speech samples analysed.

To ensure compatibility of the distinct diarisation methods for the IDS and ADS recordings, we checked the accuracy of the ALICE output by comparing it to manually determined timecodes for 13 of the 26 recordings (cf. S10 in Supplementary Materials). Our analysis showed a substantial degree of agreement between the two methods for the timecodes (Kappa = 0.72 [0.65; 0.84]) as well as substantial agreement between the extracted measures across the two types of diarization (cf., Figure S10.2-S10.3).

Altogether, this diarisation and data extraction process resulted in measures from 22,033 individual utterances: 3544 utterances in ADS and 18,489 utterances in IDS. The average length of utterance in ADS is more varied and on average includes longer utterances, as shown in Table 1. Although the number of utterances is smaller in ADS, the difference in mean length of utterance ensures that we have enough data on which to compare the speech styles, as shown by the comparable vocalic measures in Table 2.

Speech Style	**No. of Utterances**	**Mean Utterance Length (s)**	**SD of Utterance Length (s)**
ADS	3544	3.95	5.70
IDS	18,489	1.53	1.05

Table 1: Information about length of utterances in the two speech styles

Extraction of Prosodic Measures

In order to extract the utterance-level acoustic data for this project, we wrote a custom Praat script based on the principles of De Jong & Wempe's (2009) script for syllable detection. We tested the script on a speech sample to obtain appropriate parameter values and ran the script with the same parameter values on all speech samples. The script detected potential syllables by extracting peaks of intensity above a threshold of 2 dB above the median intensity of the utterance, with a time window of 64ms and a time step of 16ms. This captured the general tendency for a syllable nucleus to exhibit a higher degree of intensity than most surrounding sounds. The script then compared the intensity values of these potential syllables to the preceding dips in intensity and disregarded intensity peaks that do not show a preceding dip of at least 2 dB with respect to the current peak. By dividing the total number of syllables with the duration of the individual utterances, we calculated the articulation rate for each utterance. At the same time, the script extracted the pitch contour of the utterance using a window size of 100ms and time step of 20ms. The script then summarised the data in terms of median f_o (centralised measure of f_o in Hz), interquartile range of f_o (in Hz), as well as articulation rate (syllables per second). We thus had 22,033 individual utterance-level data points for each of these prosodic measures. We should note that these acoustic properties exhibit interdependence, as shown by the correlation network plots for each of the speech styles in Figure S4.1 of the supplementary materials. Moreover, the process of automatically extracting these acoustic measurements from the audio may have generated some errors. To combat the influence of these potential measurement errors, we chose to use robust regression over outlier detection and deletion for the following two reasons: i) it is difficult if not impossible to establish an objective definition of what counts as a measurement error versus what is an inherent property of the distribution, and ii) outlier deletion wastes data and can lead to underestimation of the variance (Yellowlees et al., 2016). We ran an additional check to determine whether this choice influenced our estimates by comparing the results of models with and without outlier deletion (i.e., values above and below two standard deviations from the mean). As shown in Table S1.5 in the supplementary materials, the control analysis of each of the utterance-level acoustic variables indicated similar model estimates and credible intervals for the data with and without outlier deletion.

Extraction of Vocalic Measures

To analyse the duration and formants of individual vowel categories in Danish across the two speech styles, we manually segmented the onsets and offsets of 9267 vowels using Praat (Boersma & Weenink, 2022). The vowel tokens were not annotated if the acoustic signal was disrupted by ambient noise or overlapping speech, if the vowel was too short to identify a stable midpoint, or if clear formants were not present due to whispered speech or creaky voice (including *stød*). For each speaker, we annotated any vowel that conformed to the above criteria until we had sampled a minimum of 150 vowel tokens in each speech style. The total number of vowels ended up being slightly higher than 7800 due to our exploration of how to extract the vowels in the initial phases of the analysis process; that is, for subjects AF, AN and CL we have 488, 792 and 486 vowel tokens, respectively, while the number of vowel tokens for the rest of the subjects are between 307-348. This sampling approach to the spontaneous speech data allows us to gain insight into the vowel types produced by each speaker while limiting degrees of freedom in the analysis process. To perform control analyses of the vowel data, moreover, we annotated each vowel according to the following four binary properties: phonological length (long vs. short), stress (stressed vs. unstressed), focus (whether the vowel appears in a focused constituent in an utterance) and word type (content word vs. function word). We wrote a custom Praat script to extract the first three formants at the temporal midpoint of each vocalic interval, vowel duration as well as the above four binary properties. We should note that the short versions of four Danish vowels, notably

/ɛ/, /œ/, /ɔ/ and /u/, have been shown to be slightly more centralised than the long versions in the Aarhus dialect (Steinlen, 2005). We assumed that the ratio of vowels with short and long durations remained similar in the two speech styles, thereby not influencing the between-style comparison of vowel quality. To check this assumption, we computed the number of vowels in each of the speech styles across different quantiles of vowel duration. As shown in Table 2, we see similar distributions of the number of vowels across quantiles, phonological length and in the total number of vowels analysed.

	No. of vowels < 20%	**No. of vowels 20-40%**	**No. of vowels 40-60%**	**No. of vowels 60-80%**	**No. of vowels 80-100%**	**No. of Long Vowels**	**No. of Short Vowels**	**Total no. of vowels**
ADS	877	934	908	889	917	390	4133	4523
IDS	983	913	947	965	937	505	4239	4744

Table 2: Overview of the number of vowels in each quantile of vowel duration as well as across phonological length for each speech style. The quantiles of duration are based on the entire dataset of vowels.

A random sample of three spectrograms for each of the seven border vowels in Danish IDS from one of the speakers is shown in Figure 1. The script considered the maximum value for the formant search to be 5,500 Hz and applied pre-emphasis to frequencies above 50 Hz. The raw vowel formant values for each vowel category are provided in Table 3.

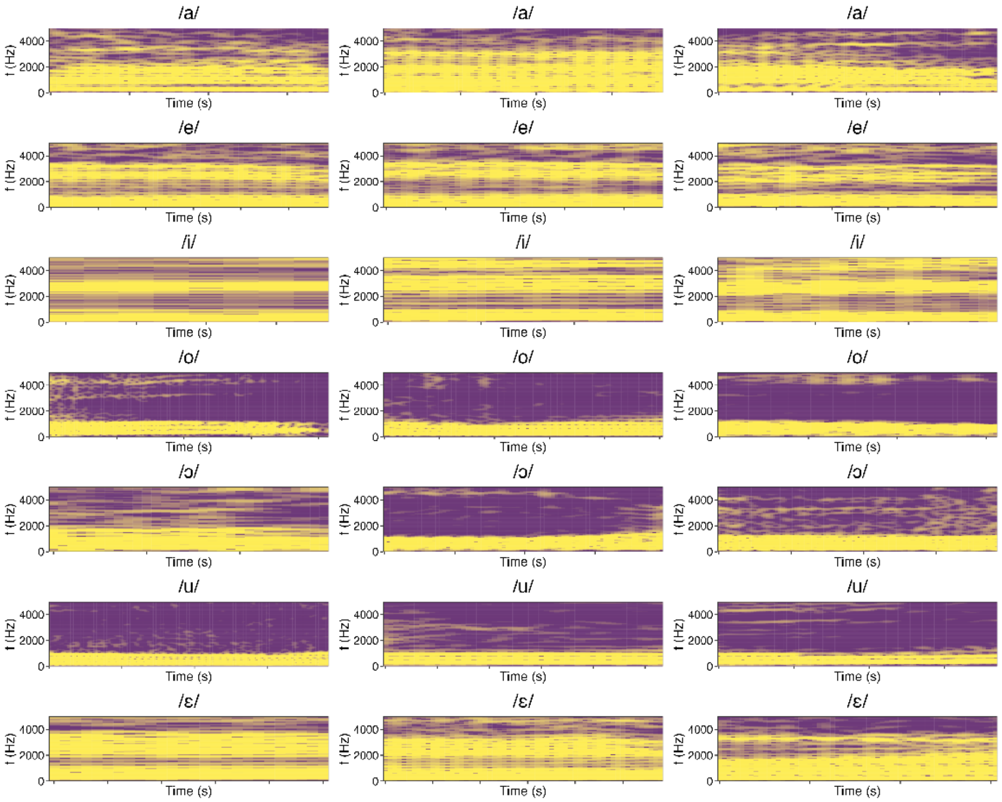


Figure 1: Three random samples of vowel tokens for seven of the peripheral vowels analysed. The more yellow the spectrogram, the higher the amplitude of the frequency present in the audio signal. The limits of the x-axis show the total time for each of the vowel tokens displayed.

These vowel formant data were then imported and further analysed in R Studio (RStudio Team, 2020). To minimise the acoustic variation that arises due to physiological and anatomical differences across speakers, the vowel formant data were normalised using the vowel-extrinsic methods of Lobanov's (1971) z-score transformation. This normalisation procedure has been shown to reduce inter-speaker anatomical/physiological variation while preserving phonemic variation, without requiring logarithmic scale transformation (Adank et al., 2004; Nearey, 1978). The procedure involved subtracting the mean speaker-specific formant value from all vowel tokens for a given speaker, and then dividing by the speaker-specific standard deviation for that formant. This procedure expresses formant values according to the hypothetical centre of each speaker's vowel space and was developed for automatic speech recognition purposes (Lobanov, 1971). The rest of the results in this paper are reported using the normalised vowel formant data. The raw formant data and code has been made available on OSF: https://osf.io/ywf9m/?view_only=d99fc6dbc61546febff619b8674a7943.

ADS					IDS				
Vowel	**n**	**Dur (SD)**	**F1 (SD)**	**F2 (SD)**	**Vowel**	**n**	**Dur (SD)**	**F1 (SD)**	**F2 (SD)**
æ	158	0.097 (0.06)	651.9 (114.7)	1879.3 (240.8)	æ	148	0.087 (0.04)	696.3 (111.4)	1888.6 (325.9)
ɑ	1040	0.083 (0.04)	792.4 (139.4)	1598 (204.4)	ɑ	968	0.08 (0.05)	792.7 (157.5)	1711.1 (264.2)
e	674	0.072 (0.04)	483.5 (85.8)	2052.9 (245.7)	e	724	0.071 (0.04)	534.7 (114.2)	2067.6 (384.3)
i	391	0.066 (0.03)	424.4 (79.7)	2201.4 (350)	i	305	0.076 (0.05)	455.7 (100.3)	2217.5 (473)
o	92	0.079 (0.04)	502.5 (105)	1069.1 (207.8)	o	179	0.088 (0.04)	591.2 (108.8)	1145.5 (240.8)
ø	87	0.071 (0.03)	472.8 (77.6)	1708 (160.9)	ø	105	0.076 (0.06)	580.1 (129)	1875.2 (223.5)
œ	6	0.075 (0.02)	529.6 (31.7)	1619.2 (175.4)	œ	31	0.077 (0.04)	617.5 (115)	1769.1 (338.3)
ɶ	103	0.093 (0.03)	688.9 (106.4)	1593.9 (136.9)	ɶ	54	0.088 (0.05)	688.5 (97.6)	1634.9 (157)
ɔ	181	0.08 (0.04)	533.2 (87.8)	1220.6 (205.8)	ɔ	336	0.089 (0.06)	641.2 (115.1)	1318.3 (202.1)
u	101	0.063 (0.05)	448.4 (91.4)	1152 (232.1)	u	157	0.068 (0.07)	487.8 (115.6)	1236.4 (311)
ʌ	863	0.08 (0.06)	635.7 (99.4)	1427.5 (175.7)	ʌ	691	0.079 (0.05)	726.2 (128.4)	1491.7 (210)
y	56	0.066 (0.03)	423.3 (83.6)	1893.8 (290.8)	y	65	0.071 (0.03)	467.6 (94.1)	1998.1 (221.5)
ɛ	576	0.072 (0.04)	581.3 (111.7)	1877.3 (208.5)	ɛ	781	0.073 (0.04)	655.1 (135.5)	1896 (297)

Table 3: Number of tokens (n), median duration (Dur) in seconds, median first formant (F1) and second formant (F2) in Hz as well as their standard deviations (SD) for each target vowel across IDS and ADS speech styles. It should be noted that the numbers in this table refer to the non-normalized formant values for each of the vowel categories. See Figures S7.1-7.4 and Tables S8.1-8.4 to see the raw formant values for each speech style across different contexts of word type, phonological length, focus and stress.

Quantification of Vowel Space Area, Vowel Variability & Distinctiveness

We used the phonR package (McCloy, 2016) to calculate two measures of vowel space area for each subject and speech style: i) the vowel space enclosed by plotting the median z-score-transformed first and second formant values of /i/, /a/ and /u/ in a Cartesian coordinate system, as in traditional studies of vowel space area (Kuhl et al., 1997; H. Liu et al., 2003), and ii) the vowel space encompassed by the median z-score-transformed formant values of all of the border vowels Danish (i.e., /i/, /e/, /ɛ/, /ɑ/, /ʌ/, /ɔ/, /o/, /u/). This latter measure was computed in order to be able to assess the difference between the two speech styles in terms of the total area encompassed by all of the vowels. To avoid confusion between these two measures in the paper, we henceforth use vowel space area to refer to the traditional /i/-/a/-/u/ measure (cf., Kuhl et al., 1997; H. Liu et al., 2003) and refer the reader to section S3 of the supplementary materials whenever we discuss the vowel space area results based on all eight border vowels. The calculation of each of the vowel space area measures is conducted for each subject in each speech style. The measure thus consists of 52 data points (i.e., 26 for each speech style). To facilitate interpretability of this measure, moreover, we z-transformed the vowel space areas by subtracting the mean vowel space area and dividing by the standard deviation. This rescales the vowel space area to be on the same scale as the hedges' *g* standardized effect size; a positive score on this scale indicates vowel space expansion while a negative score indicates vowel space reduction. We also conducted control analyses for vowel space areas based on the border vowels according to the binary properties noted for each of the vowels: stress, focus, length and word type, as shown in Figure S6.2.

Furthermore, we use the extracted formant data to estimate the extent of variability in each of the two formant dimensions for each vowel category across the two speech styles. To compare the extent of variability within each vowel category across the speech styles, we quantified the evidence in favour of a greater degree of variability in formant values within each speaker, as explained further in our choice of statistical models in the next section. Similarly, we adapted Rosslund et al.'s (2022) approach to calculating the distinctiveness of vowel categories in the two speech styles. While vowel variability indicates the compactness of vowel tokens within each category, vowel discriminability quantifies the degree of overlap between categories, which has also been argued to be a measure of speech intelligibility (cf., H. Kim et al., 2011; Whitfield & Goberman, 2014, 2017). We computed a measure of vowel discriminability for each participant within each speech style as the squared distances of category centroids from the overall vowel space centroid, divided by the squared distances of individual vowel tokens from the overall vowel space centroid. This continuous measure thus indexed the proportion of variance in F1 and F2 explained by the vowel category identity: a value of 1 would indicate that vowel membership explained all variance (i.e., a low degree of overlap between categories) and 0 would indicate that vowel membership explained no variance (i.e., a high degree of overlap between categories). Due to its reliance on the squared distances of vowel tokens, this measure of distinctiveness can thus be construed as the proportion of variance explained by category membership; the more distinctive the vowel categories, the higher the explained variance due to category membership (cf., Rosslund et al., 2022).

Statistical Modelling

To assess the extent to which the acoustic properties of Danish ADS and IDS differ, we ran Bayesian multi-level robust regression models of the data. For all of the measures appearing only on a positive range of values (i.e., all measures except vowel space area), we used a logarithmic link function to model the potential long tails of high values in the data (cf., Gabry et al., 2019; McElreath, 2018), as explained further in section S1 of the supplementary materials. For each acoustic measure, we built three models: i) an intercepts-only model with varying effects by participant, ii) a model with speech style (i.e., IDS vs. ADS) as a predictor with varying subjects nested within speech style, iii) a model with an interaction term between speech style and age as a predictor, as well as varying intercepts and slopes for subjects nested within speech style. In the models with speech style as a predictor (i.e., model (ii) and (iii)), we allowed the model to estimate a separate sigma across the two speech styles; that is, a different expected error when predicting data in the two speech styles (i.e., heteroskedasticity). We made this choice to explore the potential effects of a slight imbalance in the number of utterance-level measures for each of the speech styles (cf., Table 1) and because we might expect a greater amount of heterogeneity in caregivers' interactions with children (Englund, 2018).

To model the location (centroid) and scale (variability) of each of the vowel categories under investigation, moreover, we built hierarchical mixed-effects location-scale models for the z-score-transformed F1 and F2 measures. This type of location-scale model represents an extension to multi-level regression models in that they allow estimation of covariances among the random effects, both within and across the location and scale of the predictor (Hedeker et al., 2008; Williams et al., 2019). This form of model offers new insights into the structure of individual vowel categories, estimation of intra-vowel variability, extent of formant raising and the influence of speech style upon these parameters (Rast & Ferrer, 2018). We should note that in our location-scale models of vowel categories, we restrict our analysis to vowel categories for which have above 50 tokens (cf. Table 3) to avoid biasing the estimates.

All computations were performed in R 4.2.0 (R Core Team, 2020) using *brms* 2.16 (Bürkner, 2017) and *cmdstanr* 2.28.2 (Carpenter et al., 2017) in RStudio 1.4 (RStudio Team, 2020). We report the model formulae, model specifications and our choice of priors in section S1.1 of the supplementary materials. We chose weakly informative priors in order to discount extreme values as unlikely and to ensure minimal influence on the posterior estimates of the model (Gelman et al., 2017; Lemoine, 2019). We provide explicit description and visualisation of our choice of priors, as well as prior sensitivity checks in section S1.2-S1.4 of the supplementary materials.

Throughout the rest of this paper, we provide estimates and report 95% credible intervals, evidence ratios, and credibility scores. The credible interval refers to the range of values within which there is a 95% probability that the true parameter value lies given the assumptions of the model. We report these intervals in square brackets. The evidence ratio denotes the ratio of likelihood in favour of a hypothesis. An evidence ratio of 10, then, implies that the hypothesis is 10 times more likely than the alternative. An evidence ratio of 'Inf' (infinite) occurs when all the posterior samples conform to the direction of the hypothesis and not to alternative directions. The credibility score refers to the proportion of posterior samples in the direction of the hypothesis under investigation. We perform leave-one-out information criterion-based model comparison (Vehtari et al., 2017) between the three models and report the differences in expected log pointwise predictive density difference (elpd). This measure refers to the difference between the models in terms of their expected out-of-sample predictive accuracy; that is, it quantifies the extent to which model predictions generalise to an independent dataset. The lower the out-of-sample predictive accuracy, the lower the elpd (Vehtari et al., 2017; Yao et al., 2018).

Results

The results are structured as follows. We start by reporting results that concern the first aim of the study, namely to quantify five acoustic properties of Danish IDS: i) median f_o, ii) f_o variability, iii) articulation rate, iv) vowel duration, and v) vowel space area. We compare these results to those of a previous study on Danish IDS (Bohn, 2013) as well as a recent large-scale meta-analysis of these same acoustic features (Cox, Bergmann, et al., 2022) by incorporating prior statistical information into our models. We then present results pertaining to our second aim, namely to investigate the extent to which the acoustic measures exhibit age-related change. We restrict ourselves to depicting the acoustic measure as a function of age only if age exhibits robust effects (i.e., only for articulation rate). We refer the reader to the age plots in section S2 of the supplementary materials. We then present results that concern the third aim, namely to quantify the properties of within-vowel variability and between-vowel discriminability across the two speech styles.

Median f_o

The data for median f_o indicate that caregivers' utterances in IDS (235.8 Hz with 95% CI [228.5; 243.2]) exhibit a higher median f_o than in ADS (202.9 Hz with 95% CI [196.0; 209.6]), as shown in Figure 2. We obtain strong evidence in favour of the hypothesis that IDS exhibits a higher

f_o (33.0 Hz [27.9; 38.3], evidence ratio: Inf, credibility score: 1, with 26/26 participants displaying a positive effect). The model indicates a greater amount of variability in the distribution of overall median f_o values in IDS (sigma = 45.2 Hz [43.1; 47.5]) than in ADS (sigma = 23.7 Hz [21.0; 26.7]). We accordingly also obtain strong evidence in favour of the hypothesis that IDS exhibits a greater amount of residual variability within speech style (21.5 Hz [19.0; 24.01], evidence ratio: Inf, credibility score: 1, with 26/26 participants displaying a positive effect). The model indicates a robust correlation between IDS and ADS values (*r* = 0.66 [0.38; 0.84]). This correlation implies that there is a clear tendency for the height of subjects' f_o in IDS to depend on their corresponding f_o value in ADS. The model with age as a predictor exhibits less out-of-sample predictive accuracy (elpd: -795.3, se: 44.8) than the model without age, and we accordingly see no robust effect of age (estimate = 0.002 [-0.003; 0.008], evidence ratio = 3.36, credibility score = 0.77), as shown by Figure S2.1 in section S2 of the supplementary materials.

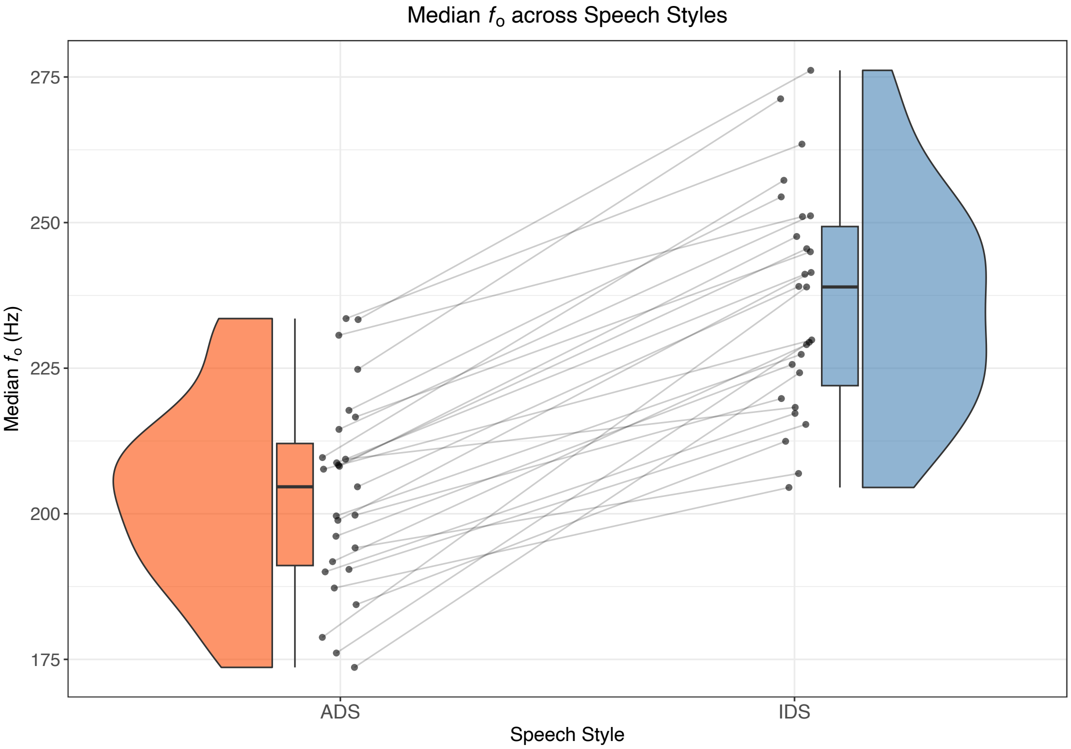


Figure 2: Plot of model estimates for individual subjects' median f_0 across the two speech styles. Each point in each speech style indicates one subject. The points for each subject are connected across the two speech styles with a line. Because model estimates are based on a pooling of repeated measures for each participant, we plot these estimates to provide a more robust picture of differences across speech styles.

f_0 variability

The data for the interquartile range of f_0 indicate that caregivers produce a higher degree of f_0 variability in IDS (74.9 Hz [71.1; 79.0] than in ADS (48.9 Hz [44.3, 53.8]), as shown in Figure 3. Similar to median f_0, we obtain strong evidence in favour of the hypothesis that f_0 variability is higher in IDS than in ADS (26 Hz [21.8; 30.2], evidence ratio: Inf, credibility score: 1, with 26/26 participants displaying a positive effect). The model indicates a greater amount of variability in the distribution of overall f_0 variability values for IDS (sigma = 44.3 Hz [42.7; 46.0] than for ADS (27.1 Hz [23.9; 30.7]). We also obtain strong evidence in favour of the hypothesis that IDS exhibits a greater amount of residual variability in overall f_0 variability values within speech style (17.1 Hz [14.4; 19.7], evidence ratio: Inf, credibility score: 1, with 26/26 participants displaying a positive effect). The model finds weak evidence for a correlation between the speech style estimates for each subject (0.35 [-0.02; 0.65]). The model with age as a predictor exhibits less out-of-sample predictive accuracy (elpd: -530.5, se: 38.2) compared to the model with only speech style as a predictor, and we accordingly find no robust effect of age (estimate = 0.000 [-0.011; 0.010], evidence ratio = 0.85, credibility score = 0.46), as shown by Figure S2.1 in section S2 of the supplementary materials.

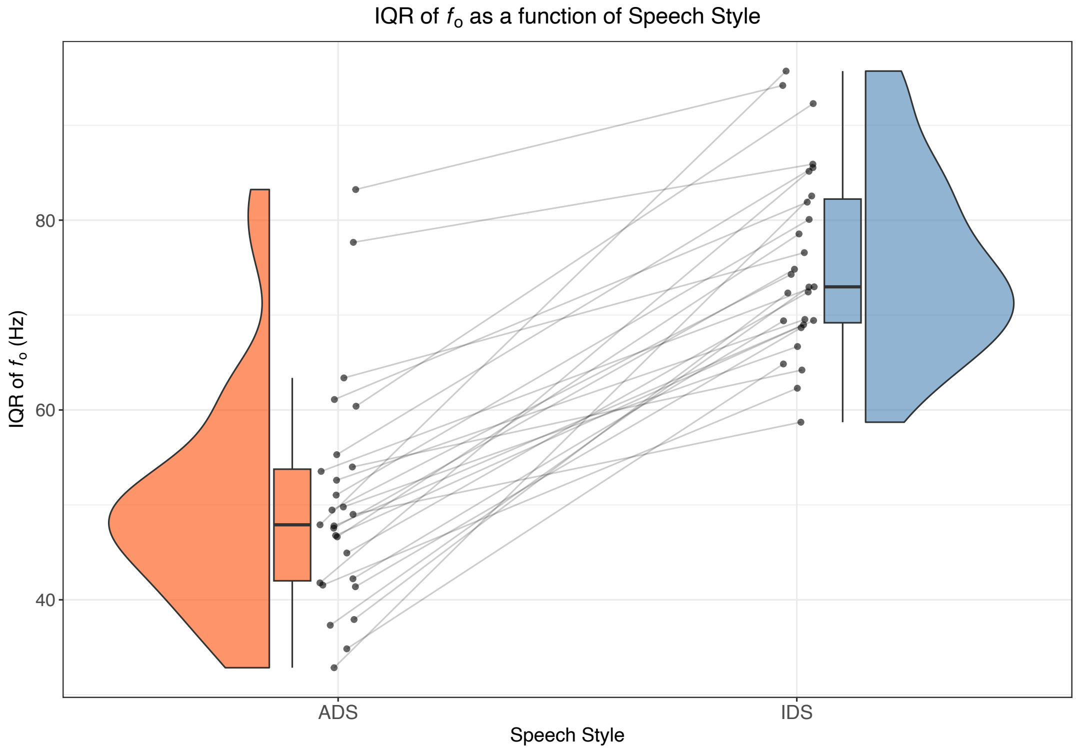


Figure 3: Plot of model estimates for subjects' pitch variability across the two speech styles. Each point in each speech style indicates one subject. The points for each subject across the two speech styles are connected with a line. Because model estimates are based on a pooling of repeated measures for each participant, we plot these estimates to provide a more robust picture of differences across speech styles.

Articulation Rate

The data for articulation rate indicate that caregivers produce fewer syllables per second in IDS (2.96 syll/s with 95% CI [2.89; 3.02]) compared to in ADS (3.72 syll/s with 95% CI [3.61; 3.84]), as shown in Figure 4. We obtain strong evidence in favour of the hypothesis of a slower articulation rate in IDS (-0.77 syll/s [-0.87; -0.67], evidence ratio: Inf, credibility score: 1, with 26/26 participants displaying a negative effect). The model indicates a higher degree of variability in IDS (sigma = 1.1 syll/s [0.958; 1.06]) than in ADS (sigma = 1.07 syll/s [1.06; 1.09]). We accordingly obtain evidence that articulation rate in IDS exhibits a greater amount of residual variability within speech style (0.06 syll/s [0.02; 0.11], evidence ratio = 139.6, credibility score = 0.99, with 26/26 participants displaying a positive effect). The model indicates no clear correlation between the speech style estimates for each subject (*r* = 0.11 [-0.29; 0.49]). The model with age as a predictor exhibits similar out-of-sample predictive accuracy (elpd: -23.4, se: 6.9) compared to the model with only speech style as a predictor, and we observe a robust small effect of age (estimate = 0.006 [0.002; 0.009], evidence ratio = 172.1, credibility score = 0.994), as shown in Figure 5 as well as Figure S2.1 in S2 in the supplementary materials.

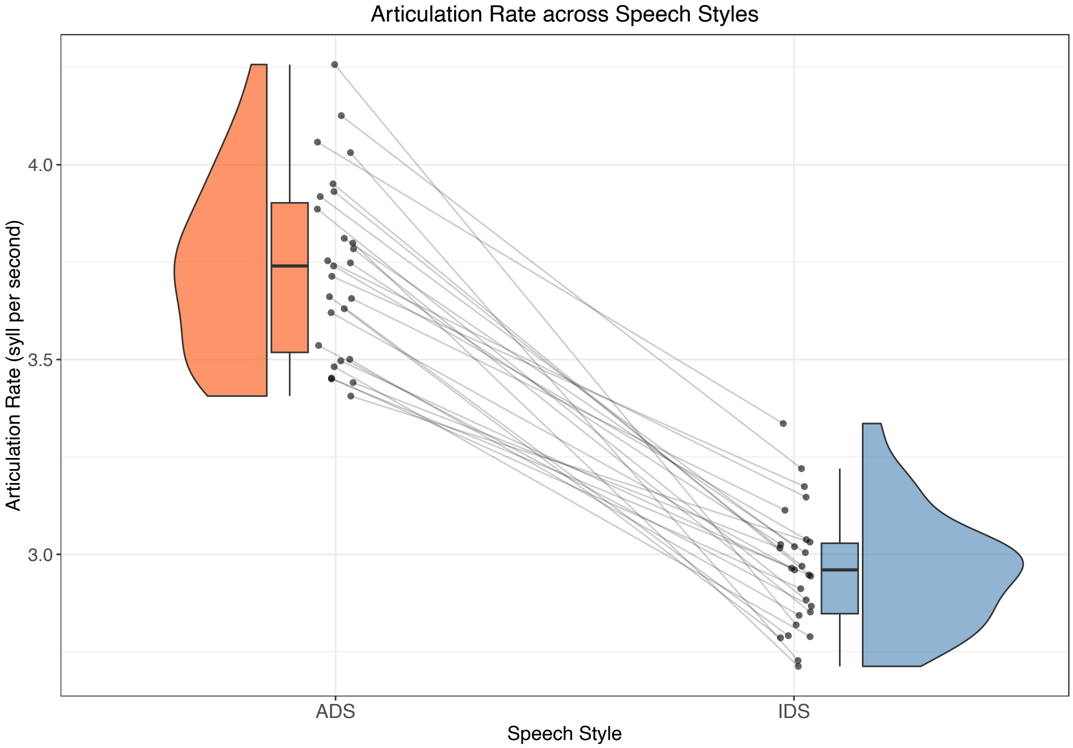


Figure 4: Plot of model estimates for subjects' articulation rate across the two speech styles. Each point in each speech style indicates one subject. The points for each subject are connected across the two speech styles with a line. Because model estimates are based on a pooling of repeated measures for each participant, we plot these estimates to provide a more robust picture of differences across speech styles.

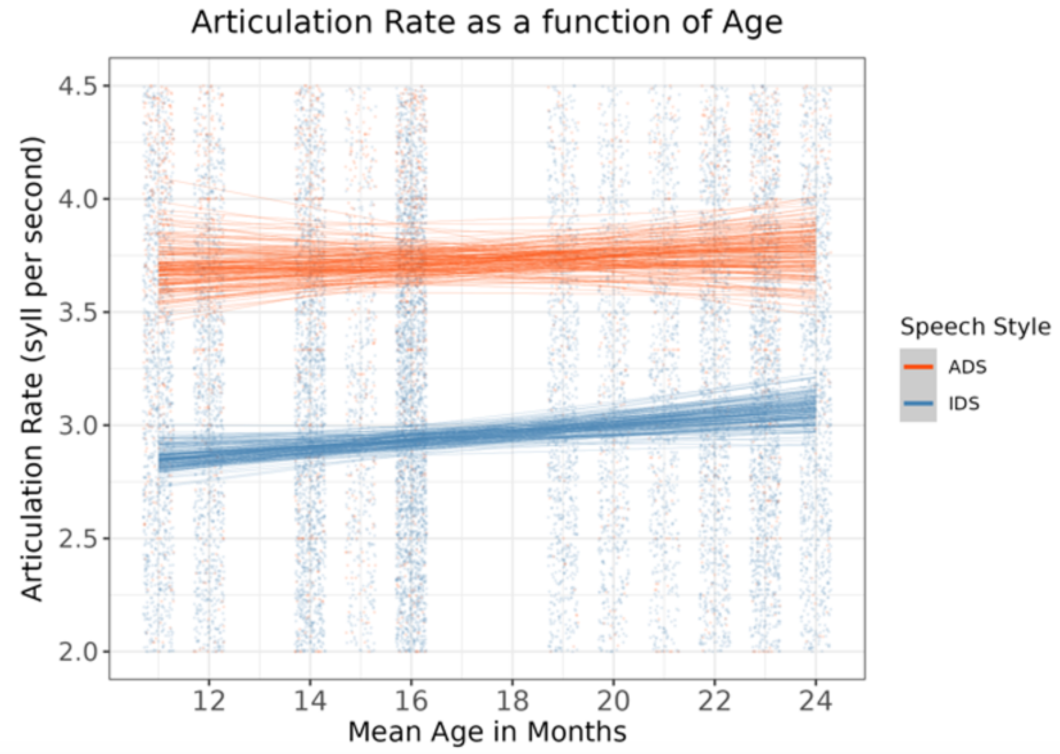


Figure 5: Plot showing how caregivers' articulation rate changes as a function of infant age in the two speech styles. The plot shows 150 posterior predictions from the model including age as a predictor. The plot indicates relative stability for ADS and a relative increase in articulation rate in IDS.

Vowel Duration

The data for vowel duration indicate that caregivers overall produce similar vowel durations in ADS (0.078 s [0.075; 0.081]) and in IDS (0.077 s [0.074;0.081]), as shown in Figure 6. We obtain no evidence in favour of the hypothesis that caregivers' vowels in IDS exhibit longer duration (-0.000 s [-0.004; 0.003], evidence ratio = 0.61, credibility score = 0.38, with only 6/26 participants displaying a reliable positive effect). The model also indicates a similar degree of variability in the distribution of values in IDS (sigma = 0.024 [0.022; 0.025]) and ADS (sigma =

0.024 [0.022; 0.026]). We accordingly obtain no evidence in favour of the hypothesis that IDS exhibits a greater amount of residual variability within speech style (0.000 s [-0.001; 0.002], evidence ratio: 1.33, credibility score: 0.57, with only 5/26 participants showing a reliable positive effect). The model with age as a predictor exhibits less out-of-sample predictive accuracy (elpd: -77.7, se: 14.5) compared to the model without age as a predictor, and we likewise see no robust effect of age (estimate = 0.002 [-0.007; 0.010], evidence ratio = 1.80, credibility score = 0.64), as shown by Figure S2.1 in section S2 of the supplementary materials. We conducted a control analysis of vowel duration to check whether the speech styles exhibited different values across phonologically short and long vowels (cf., Figure S6.1). The control analysis indicated that short vowels are slightly longer in ADS than in IDS, however, not robustly so (estimate = 0.02 [-0.02; 0.06], evidence ratio = 3.1, credibility score = 0.77), whereas long vowels in ADS were shown to be longer than those in IDS (estimate = 0.07 [0.01; 0.13], evidence ratio = 30.5, credibility score = 0.97). We conducted a second control analysis by normalizing vowel duration according to the inverse of the median articulation rate for each speaker (cf. Figure S9.1 and S9.2 in Supplementary Materials). This control analysis indicated that rate-normalised vowel length in IDS (0.23 [0.22; 0.24]) was robustly smaller than in ADS (0.30 [0.29; 0.31]).

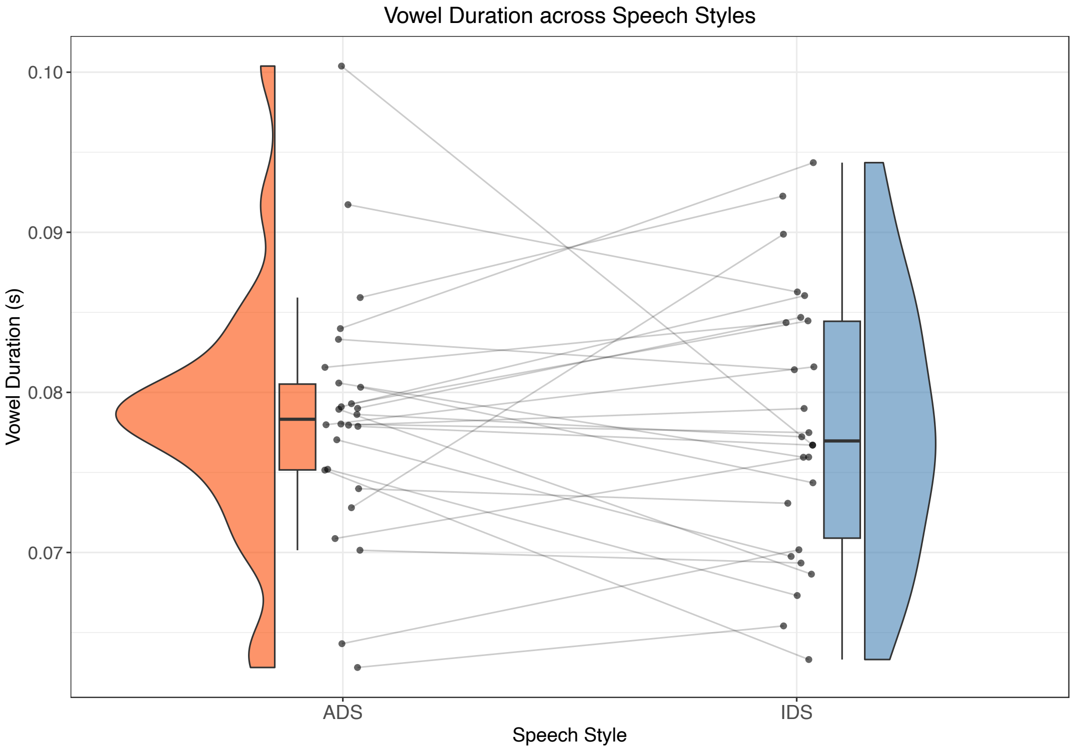


Figure 6: Plot of model estimates for subjects' vowel duration across the two speech styles. Each point in each speech style indicates one subject. The points for each subject are connected across the two speech styles with a line. Because model estimates are based on a pooling of repeated measures for each participant, we plot these estimates to provide a more robust picture of differences across speech styles.

Vowel Space Area

The data for vowel space area indicated that caregivers' vowel space area in ADS (Cohen's *d* = -0.01 [-0.37; 0.34] was greater than that in IDS (Cohen's *d* = -0.44 [-0.81; 0.05]. As shown in Figure 7, we obtained strong evidence in favour of the hypothesis that the vowel space area in IDS was smaller than in ADS (*d* = -0.43 [-0.85; -0.001], evidence ratio = 19.18, credibility score = 0.95, however, the posterior estimates for all 26 participants included the null within their credible interval). The model indicated similar degree of substantial variability in ADS (sigma = 0.88 [0.52; 1.38]) and IDS (sigma = 0.94 [0.55; 1.49]). We obtained no evidence that IDS exhibited a greater degree of residual variability in vowel space area estimates within speech style (*d* = 0.09 [-0.43; 0.62], evidence ratio = 1.58, credibility score = 0.61, with 0/26 participants displaying an effect without the null in their credible interval). The model with age as a predictor exhibited similar out-of-sample predictive accuracy (elpd: -1.3, se: 2.2) compared to the model without age as a predictor; however, as shown in Figure S2.1 in the supplementary materials, the measure exhibited a high degree of uncertainty (estimate = -0.04 [-0.07; -0.01], evidence ratio = 31.64, credibility score = 0.969). These results should be interpreted with caution, especially as the model of vowel space area based on all of the border vowels showed no robust differences between the speech styles (estimate = -0.14 [-0.54; 0.26], evidence ratio = 2.59; credibility score: 0.72, with 0/26 participants displaying a robust positive effect), nor any effect of infant age (estimate = -0.012 [-0.046; 0.022], evidence ratio = 2.59, credibility score: 0.72), as shown in Figures S3.1 and S3.2 in the supplementary materials. We discuss the implications of these results further below. We conducted a control analysis of vowel space area with all border vowels to check whether the speech styles exhibited different values across different contexts of stress, focus, word type and length (cf., Figure S6.2 for all estimates). Across the majority of contexts, there were no differences in vowel space with all border vowels, with the exception of long vowels where the ADS vowel space area tends to be smaller (estimate = -0.44 [-0.87; -0.01], evidence ratio = 20.2, credibility score = 0.95) as well as unstressed vowels where ADS vowel space area appears to more expanded than in IDS (estimate = 0.46 [0.13; 0.8], evidence ratio = 75.77, credibility score = 0.99).

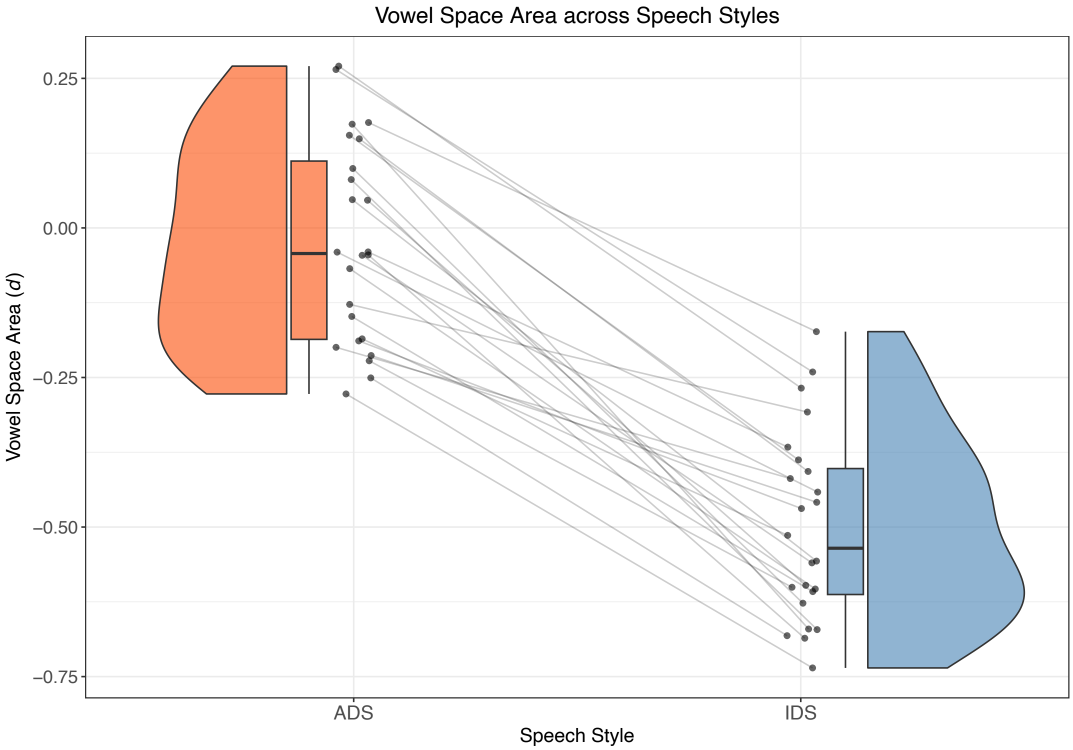


Figure 7: Plot of model estimates for subjects' vowel space area across the two speech styles. Each point in each speech style indicates one subject. The points for each subject are connected across the two speech styles with a line. Because model estimates are based on a pooling of repeated measures for each participant, we plot these estimates to provide a more robust picture of differences across speech styles.

Meta-analytic Priors

In this section, we consider whether our findings change if we statistically integrate information from a recent meta-analysis of IDS (Cox, Bergmann, et al., 2022) and a prior empirical study of Danish IDS (Bohn, 2013) and how our posterior estimates relate to those different priors. In Figure 8, we depict how each of these priors update into posteriors after seeing the data; that is, we incorporated information from other studies and assessed the extent to which our estimates changed (cf., Fusaroli, Grossman, et al., 2021; Parola et al., 2022 for similar approaches). We chose the parameters of the skeptical prior in this analysis with a view to regularising the effect of data on the posterior estimates; that is, this skeptical prior encodes the very low likelihood of large effect sizes (cf. section S1 in supplementary materials). Figure 8 and Table 4 indicate that prosodic properties of Danish IDS (i.e., f_o, f_o variability and articulation rate) conform to the expectations generated by meta-analytic estimates (in purple), although the estimates for Danish show smaller effect sizes for each of the three acoustic measures (cf., Kvarven et al., 2020). Note, however, that our results for f_o contradict those of Bohn (2013) (in green) who found evidence for a null effect. Interestingly, the vowel duration results for Danish contradict the cross-linguistic tendency for vowels to be longer in IDS, as shown by the meta-analytic estimate, but conform to evidence obtained in Bohn's (2013) experimental study of Danish. Lastly, our negative posterior estimate for vowel space area in Danish similarly contradicts cross-linguistic patterns of vowel space expansion, but conforms to the results obtained in Bohn's (2013) experimental study. We should point out that due to the high degree of uncertainty associated with the current sample and the low degree of uncertainty in the meta-analytic estimate based on 32 studies (Cox, Bergmann, et al., 2022), the Danish data do not substantially sway the existing evidence of a general vowel space expansion in IDS across languages. The lack of overlap between the skeptical and meta-analytic posterior distributions supports the idea of Danish behaving differently from most other languages in regard to vowel space area in IDS.

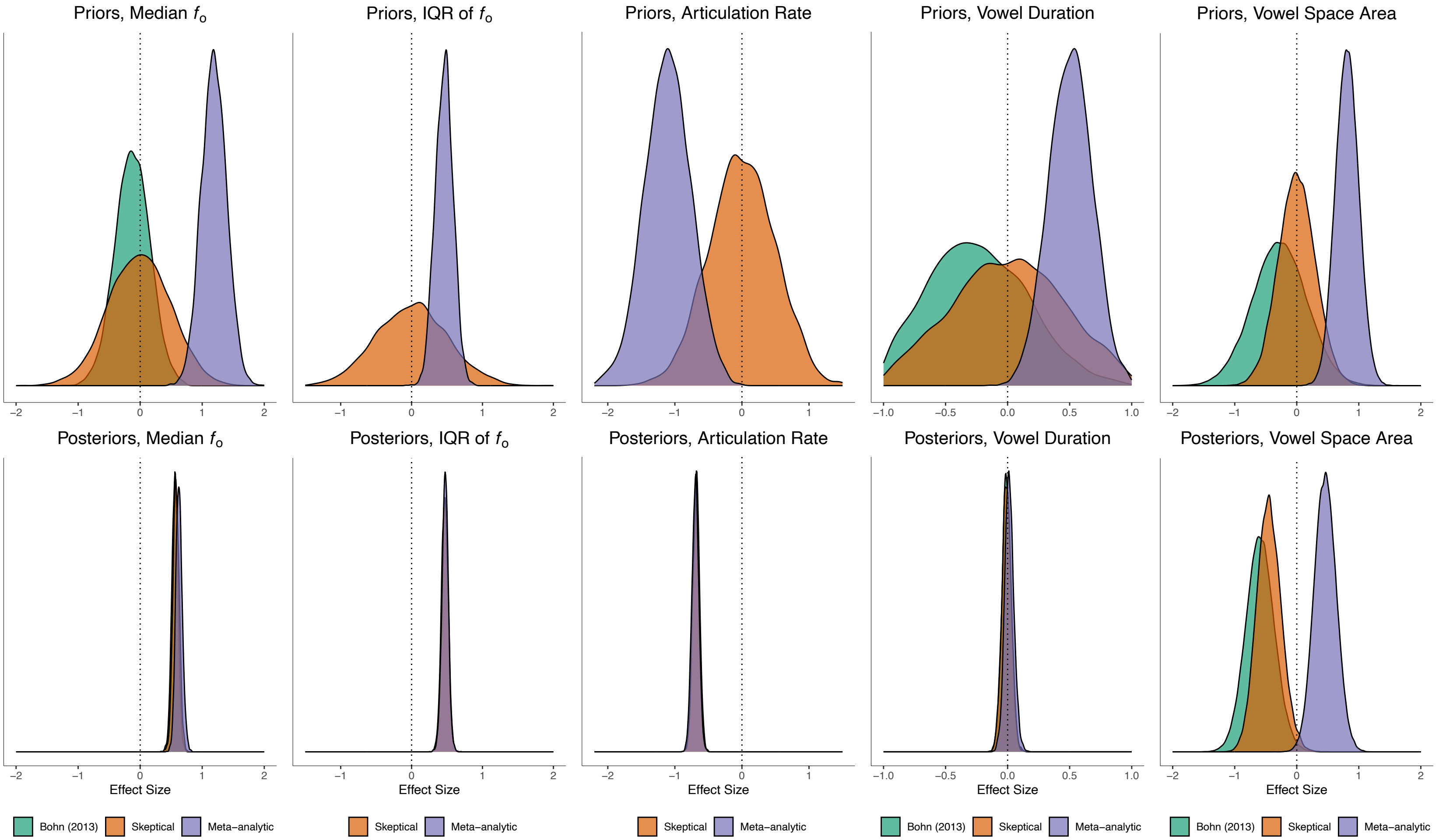


Figure 8: A panel of prior-posterior update plots for each of the acoustic measures, showing how the skeptical and meta-analytic priors change as a result of learning from the data. The meta-analytic estimates consist of a synthesis of data from studies on IDS (Cox, Bergmann, et al., 2022) as well as a recent large-scale cross-linguistic study on the features of IDS (Hilton et al., 2021).

Acoustic Measure	**Estimates from Bohn (2013)**	**Meta-analytic estimates from Cox et al. (2022)**	**Posterior Estimate (w. meta-analytic prior)**	**Posterior Estimate (w. Bohn (2013) prior)**	**Posterior Estimate (w. skeptical prior)**
Median f_0	0.13 [.0.41; 0.67]	1.20 [0.78; 1.62]	0.63 [0.53; 0.74]	0.56 [0.45; 0.66]	0.58 [0.47; 0.68]
IQR of f_0	NA	0.49 [0.15; 0.82]	0.48 [0.38; 0.57]	NA	0.47 [0.37; 0.56]
Articulation Rate	NA	-1.12 [-1.76; -0.42]	-0.72 [-0.82; -0.61]	NA	-0.68 [-0.78; -0.57]
Vowel Duration	0.20 [0.34; 0.74]	0.52 [0.22; 0.83]	-0.02 [-0.05; 0.10]	-0.01 [-0.08, 0.07]	-0.01 [-0.08; 0.06]
Vowel Space Area	-0.28 [-1.14; 0.58]	0.83 [0.53; 1.13]	0.59 [0.30; 0.87]	-0.58 [-1.02; -0.12]	-0.39 [-0.77; 0.03]

Table 4: Effect size estimates for each of the acoustic features in Danish IDS. The columns compare the posterior estimates for the models with meta-analytic priors (Cox, Bergmann, et al., 2022), priors from Bohn's (2013) study, and the models with sceptical priors (cf. S1.1 in supplementary materials for more information about the sceptical priors used here). The meta-analytic estimates consist of a synthesis of data from studies on IDS (Cox, Bergmann, et al., 2022) as well as a recent large-scale cross-linguistic study on the features of IDS (Hilton et al., 2021).

Within-Vowel Variability

This section concerns the third aim of the study, namely to investigate within-vowel variability and between-vowel discriminability in Danish IDS and ADS. The results of the location-scale model pertaining to within-vowel variability indicated that the vowel categories of Danish IDS exhibited a greater degree of within-category variability to those in Danish ADS. This is reflected in Table 5 below, which shows the ratio of evidence in favour of ADS vowel categories showing less variability than IDS vowel categories. The model indicated that all of the vowel clusters in IDS (except /ɑ/) exhibited more variability than in ADS in the dimension of vowel height (i.e., F1), and all of the vowel clusters except /y/ and /ɔ/ exhibit more variability in IDS in the dimension of vowel front-backness (i.e., F2). The model also showed evidence that the majority of the vowels investigated exhibited raising of IDS formants in both height and front-backness dimensions (cf. Table 5). Figure 9 shows posterior samples drawn from the location-scale model of the vowel data and likewise reflects these patterns.

Vowel Cluster	$\mathbf{F1_{ADS} < F1_{IDS}}$	$\mathbf{F2_{IDS} > F2_{ADS}}$	$\mathbf{F1, \sigma_{ADS} < \sigma_{IDS}}$	$\mathbf{F2, \sigma_{ADS} < \sigma_{IDS}}$
i	189	0.38	Inf	399
y	31.5	9.84	7.97	0.77
e	Inf	1.76	Inf	132.00
ø	3999	443.00	3999	9.05
ɛ	Inf	3.19	Inf	25.0
æ	20.86	0.906	3.23	1.53
ɶ	5.68	Inf	0.922	38.2
ɑ	1.56	Inf	56.1	147.00
ʌ	Inf	Inf	Inf	69.2
ɔ	Inf	799	332	0.269
o	49.6	147.00	1.05	4.47
u	60.5	29.3	18.0	147.0

Table 5: Evidence ratios for individual vowel categories. An evidence ratio of > 1 indicates evidence in favour of the hypothesis either i) that the value of F1 or F2 in ADS is lower than that in IDS (columns 2 and 3) or ii) that the F1 or F2 of the vowels in ADS exhibits less variability than that in IDS (columns 4 and 5).

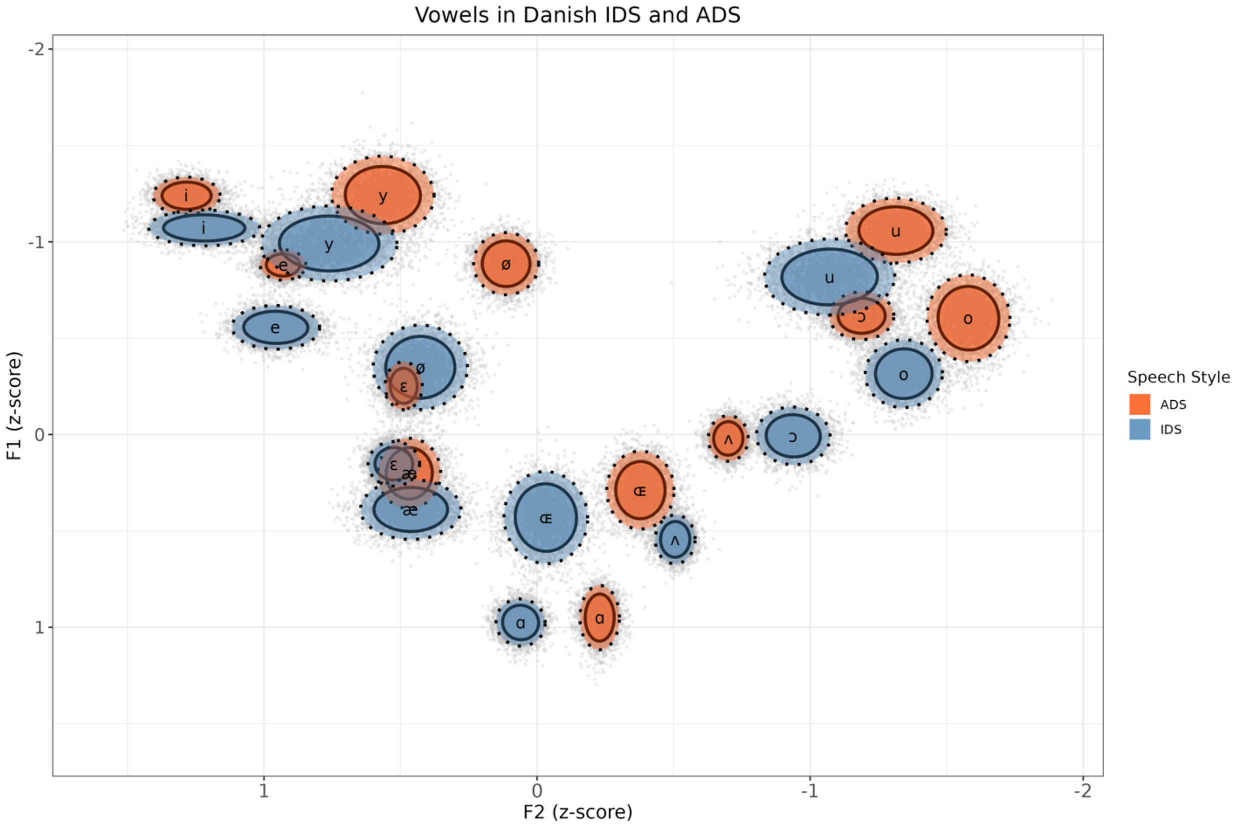


Figure 9: Plot of posterior draws from the location-scale model of vowel categories in Danish IDS (blue) and ADS (orange). The dotted ellipsis encompasses 95% of the vowels while the innermost ellipsis surrounds 80% of the vowel distribution. Note that we only model vowel categories for which we have above 50 tokens (cf. Table 3).

Vowel Discriminability Measure

Similarly, the model of vowel discriminability indicates that caregivers produce less discriminable vowels in IDS (0.54 [0.50; 0.59]) than in ADS (0.67 [0.64; 0.70]), as shown in Figure 10. We obtain strong evidence in favour of the hypothesis that caregivers exhibit a lower proportion of explained variance in IDS (-0.13 [-0.16; -0.09], evidence ratio = Inf, credibility score = 1, with 24/26 participants displaying a reliable negative effect). The model also indicates a similar degree of variability in the distribution of values in IDS (sigma = 0.023 [0.002; 0.088]) and in ADS (sigma = 0.017 [0.002; 0.062]). We accordingly obtain weak evidence in favour of the hypothesis that IDS exhibits a greater amount of variability (0.01 [-0.04; 0.06], evidence ratio = 0.7, credibility score: 0.7, with 0/26 participants showing a reliable positive effect). The model with age as a predictor exhibits similar out-of-sample predictive accuracy (elpd: -1.2, se: 4.7) compared to the model without age as a predictor; however, as shown in Figure S2.1 in the supplementary materials, we see no robust effect of age (estimate = 0.001 [-0.007; 0.0.009], evidence ratio = 0.63, credibility score = 0.38).

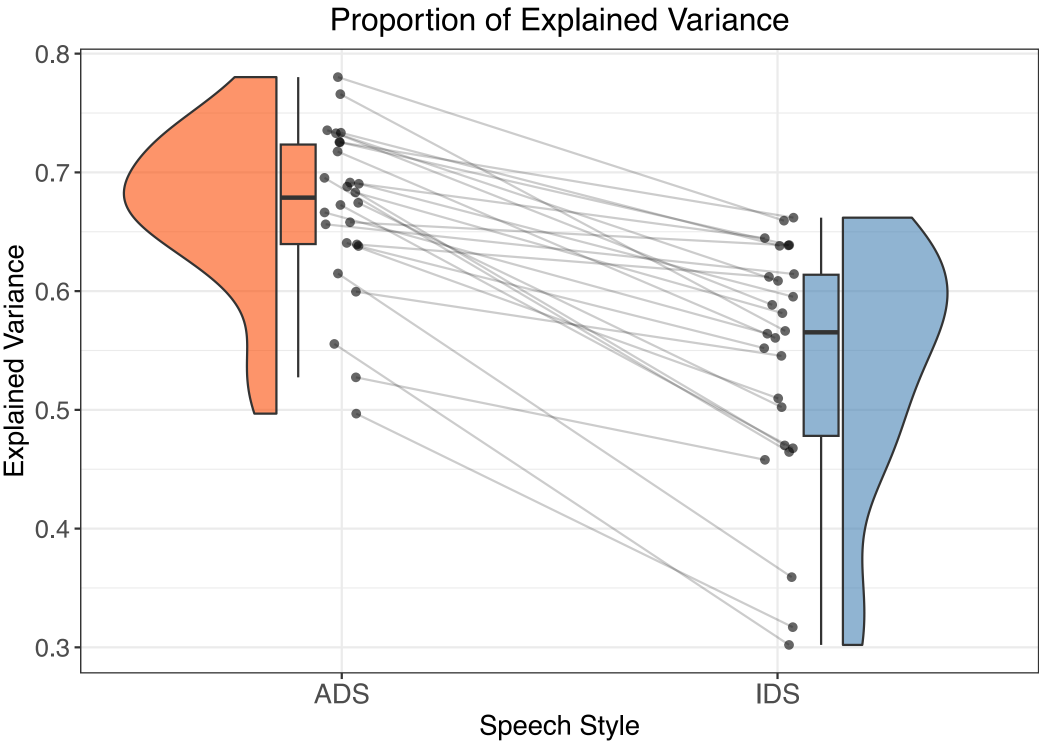


Figure 10: Plot of model estimates for the proportion of explained variance from category membership across the two speech styles. Each point in each speech style indicates one subject. The points for each subject are connected across the two speech styles with a coloured line.

Discussion

This paper set out to investigate caregivers' spontaneous IDS and ADS in Danish and quantify the extent to which caregivers modify the prosodic and vocalic features across the two speech styles. We turned our focus to three different aspects of the acoustic expression of Danish IDS.

Firstly, we looked at the extent to which five acoustic properties that have been reported extensively in the literature on IDS (Cox, Bergmann, et al., 2022) were expressed in Danish. The results indicated that Danish caregivers conformed to cross-linguistic patterns of prosodic properties of IDS, producing IDS with a higher median f_o, a greater degree of f_o variability, and a slower rate of articulation. The vocalic measures of vowel duration and vowel space area in IDS, on the other hand, contradicted cross-linguistic tendencies, with caregivers producing either a reduced or similar vowel space area in IDS (cf., Figure 7 and S3.1), as well as similar vowel durations or slightly longer vowels in ADS (cf., Figure 6 and Figure S6.1). Secondly, we asked whether any of the acoustic properties of Danish caregivers' IDS exhibited change according to infant age. The results indicated no clear age-related changes in the majority of the prosodic and vocalic measures (cf. S2 in supplementary materials), with the exception of articulation rate, which became gradually more similar to Danish ADS in IDS addressed to older infants (cf., Figure 5). Lastly, we asked how parents negotiated the balance between within-vowel variability and between-vowel separability, and whether vowel categories in Danish IDS exhibited a higher degree of discriminability. In our sample of participants, Danish caregivers produced formant raising, a greater degree of within-vowel variability and a lower degree of between-vowel discriminability in IDS when compared with ADS. While our results thus partially support prior cross-linguistic findings (e.g., Englund & Behne, 2005; Rattanasone et al., 2013; Rosslund et al., 2022), they also provide further evidence of the unusual nature of Danish speech and conversational practices (cf., Dideriksen et al., 2022; Trecca et al., 2021 for a review). In the following discussion, we discuss each of these results in turn and argue that they call for more hypothesis-driven comparisons between similar languages (Christiansen et al., 2022) as well as the widespread adoption of cumulative science practices (Brand et al., 2019; Fusaroli, Grossman, et al., 2021).

Danish IDS and Cross-Linguistic Patterns

The finding that Danish caregivers produce speech with an elevated pitch, more varied melody and fewer syllables per second in IDS than in ADS conforms to cross-linguistic patterns of prosodic features of IDS (Cox, Bergmann, et al., 2022; Hilton et al., 2022). The current results provided further evidence showing that the speech style we use when interacting with young children has similar prosodic features across a wide intersection of languages (Broesch & Bryant, 2015; Bryant & Barrett, 2007; Hilton et al., 2022). The above results for Danish and most cross-linguistic studies highlight pitch properties as being the most salient cues in IDS during early development (Broesch & Bryant, 2015; Bryant & Barrett, 2007; Fernald, 1989). These acoustic properties serve the functions of communicating intentions, grabbing an infants' attention, expressing emotions, and encouraging behaviour (Bryant & Barrett, 2007; Fernald & Mazzie, 1991; Fernald & Simon, 1984). This result also reverberates with studies showing that the tendency for IDS to grab infants' attention seems to be driven primarily by pitch elevation (Segal & Newman, 2015). The finding of a higher median f_0 in this study actually contradicts a previous study of Danish IDS (Bohn, 2013); however, the cause of this discrepancy may be due to methodological differences, as we know that recordings of spontaneous speech produces bigger effect sizes compared to those of more controlled speech (Cox, Bergmann, et al., 2022). The explicit integration of Bohn's (2013) results into the prior of our statistical model allowed us to quantify the extent to which this prior updates into a positive posterior estimate after seeing our data (cf. Figure 8). Given the small differences between the posterior estimates with priors from Cox, Bergmann et al.'s (2022) meta-analysis and Bohn's (2013) experimental study, the results suggest that we now have strong evidence that f_0 in Danish IDS is higher than in ADS. It should be noted, however, that the size of the effect is more moderate than the effect size suggested by the meta-analysis of IDS (Cox, Bergmann, et al., 2022).

The greatest difference between the two speech styles was a slower articulation rate in IDS (cf. Table 4). This acoustic property of Danish IDS again conforms to cross-linguistic patterns of acoustic features of IDS (Cox, Bergmann, et al., 2022) and may serve the purpose of easing the cognitive demand involved in young infants' processing of speech (Christiansen & Chater, 2016; Peter et al., 2016; Saffran & Kirkham, 2018; Thiessen et al., 2005). A slowed articulation rate has also been shown to increase the intelligibility of speech (Ferguson & Kewley-Port, 2007; J. Lam & Tjaden, 2013; Searl & Evitts, 2013), to facilitate word recognition and learning (Raneri, 2015; Song et al., 2010) and to be used when introducing unfamiliar words (Han et al., 2018). In terms of prosodic properties of IDS, then, Danish caregivers made acoustic modifications in a way that suggests flexible adaptation to infants' communicative immaturity and developmental needs (Fusaroli, Weed, et al., 2021; Goldstein & Schwade, 2008; Ko et al., 2016; Nguyen et al., 2022; Warlaumont et al., 2014).

Our analysis of vowel space area in Danish IDS and ADS contradicted the general cross-linguistic tendency for vowel space expansion in IDS (e.g., Hartman et al., 2017; H. Liu et al., 2003), but was consistent with earlier acoustic investigations of IDS in Danish (Bohn, 2013; Dideriksen & Fusaroli, 2018) and the growing number of studies showing language specificity in some of the acoustic properties of IDS (e.g., Englund, 2018; Rattanasone et al., 2013; Rosslund et al., 2022). Vowel space reduction, for example, has been found in a number of different languages, such as Dutch (Benders, 2013), Norwegian (Englund, 2018; Englund & Behne, 2005; Steen & Englund, 2021) and Cantonese (Rattanasone et al., 2013). Considering the general connection between speech clarity and vowel space expansion (J. Lam et al., 2012; Whitfield & Goberman, 2017; Whitfield & Mehta, 2019) – as well as the phonetic opacity of Danish sound structure – we may have expected Danish caregivers to increase the separability between the centroids of vowel categories in IDS. Our control analysis of the vowel space area measures (cf., Figure S6.2) shows that even in contexts of emphasis (e.g., focused constituents and content words), we see no clear evidence of vowel space expansion, the only potential exception being long vowel contexts where the vowel space in IDS is slightly expanded. These results highlight the need for more studies to test whether clear speech in Danish as well as other languages can be described with vowel space expansion, or to a greater extent can be measured by a slowed articulation rate or some other mediating acoustic variable (Ferguson & Kewley-Port, 2007; J. Lam & Tjaden, 2013; Searl & Evitts, 2013).

Another feature that has been proposed to aid speech intelligibility involves an exaggeration of differences in vowel duration, thus making relevant phonological differences more salient to children (Seidl & Cristià, 2008; Soderstrom et al., 2003). In our first control analysis of the vowels across different contexts of phonological length (cf. Figure S6.1), we found that phonologically long vowels in ADS exhibit longer duration compared to those in IDS, whereas vowels of short phonological length exhibited similar durations across the two speech styles. These findings of no clear differences between vowel durations across the two speech styles should be interpreted in light of the use of quantity distinctions to distinguish word meaning in Danish. For example, distinctions between the singular *tov* [tɒʊ̯] ‘rope’ and the plural *tove* [ˈtɒʊ̯ː] ‘ropes’ or between *hus* [huːˀs] ‘house’ and *huse* [huːːs] ‘houses’ often relies on subtle quantity differences in vowel length. The finding of a lack of exaggeration of IDS vowel lengths within this complex quantity system may suggest that caregivers rely on more transparent cues to clarify these word meanings (Kjærbæk & Basbøll, 2016), such as clearer consonantal cues or [ɐ]- and [ə]-suffixation suffixation (e.g., *tove* [ˈtʌwə] ’ropes’ and *huse* [ˈhuːsə] ‘houses’). Moreover, because longer segmental duration tends to co-occur with a slowed articulation rate (Panneton et al., 2006; Song et al., 2010), we conducted a second control analysis where we normalised vowel duration by the inverse of the median articulation rate for each speaker. Rate-normalised vowel length was shorter in IDS than in ADS (cf., Figure S9.1), implying that vowels constitute a smaller proportion of the speech stream in IDS.

Because rate-normalised vowel length exhibited similar slopes for phonologically long and short vowels across speech styles (cf., Figure S9.2), the results still suggest a lack of exaggerated quantity distinctions between long and short vowels in IDS. These findings hold interesting implications in the context of Danish phonetic structure, where the frequent reduction of obstruents to vocalic sounds (i.e., consonant reduction) has been argued to reduce the salience of cues that allow infants to extract information from the speech stream (Bleses et al., 2008a, 2008a; Højen & Nazzi, 2016; Trecca et al., 2019). The finding that rate-normalised vocalic material comprises a smaller part of the speech stream in IDS may tentatively suggest that Danish caregivers to a greater extent exaggerate consonantal cues in IDS, presumably to aid Danish-learning infants in segmenting a highly vocalic speech stream (Bleses et al., 2008a; Trecca et al., 2019, 2020). These intriguing findings in Danish IDS and ADS vowels require further experimental investigation and demonstrate the importance of considering subtle distinctions across phonological systems in discovering differences in the acoustic expression of IDS and ADS.

Stability & Dynamic Changes in Features

The second aim of this paper was to investigate the extent to which the acoustic expression of Danish IDS changes with infant age. The results indicated no clear age-related changes in the majority of the prosodic and vocalic measures (cf. S2 in supplementary materials), with the exception of articulation rate, which became gradually more similar to Danish ADS the older the infants were (cf., Figure 5). This finding of an age-related change conformed to other cross-linguistic studies of age-related changes in the acoustic properties of IDS (Kondaurova et al., 2013; Lee et al., 2014; Narayan & McDermott, 2016; Raneri, 2015) and may reflect caregivers' adaptation to infants' gradual improvement in their processing of the speech stream over the span of early development (Christiansen & Chater, 2016; Peter et al., 2016; Saffran & Kirkham, 2018; Werker & Tees, 1999). The lack of change in vowel duration, however, which often goes hand in hand with articulation rate, contradicted longitudinal studies in several languages, indicating that caregivers often decreased the relative vowel duration differences in IDS and ADS as infants became older (Englund & Behne, 2005; Hartman et al., 2017; Vosoughi & Roy, 2012). The lack of age-related changes in pitch properties likewise contradicted the majority of studies that suggest that caregivers reduce the median f_0 and f_0 variability in IDS as their infants become older (Amano et al., 2006; Gergely et al., 2017; Han et al., 2020; Kondaurova et al., 2013; Niwano & Sugai, 2002; Stern et al., 1983; Vosoughi & Roy, 2012). Similarly, we saw no changes in either vowel space expansion or vowel discriminability. The stability in these measures may indicate their continued function of increasing infant attention and social motivation as well as expressing affect (Fernald, 1989; Fernald & Kuhl, 1987; Kitamura & Lam, 2009; although see Ma et al., 2011); however, any conclusions that we can make about age-related changes may be limited by our reliance on a cross-sectional sample, as discussed further in the *Limitations & Future Directions* section below.

Within-Vowel Variability & Between-Vowel Discriminability

One of the main objectives of this study was to broaden our knowledge of the internal distributions of vowel categories in Danish IDS and ADS by investigating each of the members in the large vowel inventory. The results indicated that caregivers produce more variable vowel categories in IDS compared to ADS, providing another example to add to the growing number of studies showing less compact vowel categories in IDS (cf., Cristia & Seidl, 2014; Martin et al., 2015; McMurray et al., 2013; Miyazawa et al., 2017; Rosslund et al., 2022). This larger degree of variability in IDS vowel categories also influenced the between-vowel discriminability in IDS, with vowels being more overlapping in IDS compared to ADS. This lower degree of vowel discriminability did not appear to change across infant ages (cf. Figure S2).

The combination of less between-vowel separability (i.e., a similar or reduced vowel space area) and a higher degree of within-vowel variability in IDS at first glance contradicted the prominent hypothesis that IDS serves to clarify the speech signal and help infants learn phonetic categories (Hartman et al., 2017; H. Liu et al., 2003). However, the presence of variability may benefit infants in a number of different ways. One way that variability may benefit infants is by leading the infant to a greater degree of abstraction from individual categories and a more robust system of categorisation (Perry et al., 2010; Raviv et al., 2022; Rost & McMurray, 2009, 2010). For example, Rost and McMurray (2009) found that infants trained on labels spoken by a single speaker failed to distinguish between labels for visual objects that had a minimal phonological difference, whereas infants who were exposed to labels uttered by multiple speakers succeeded. Houston and Jusczyk (2000) similarly showed that increasing the number of speakers during familiarisation facilitated 5-7-month-old infants' generalization of sound patterns to novel speakers. Infant word learning has likewise been shown to rely on variability; Perry et al.'s (2010) longitudinal study showed that infants trained on a more variable set of stimuli exhibited greater generalisation to novel stimuli of these categories. These studies indicate that a greater degree of variability may allow infants to abstract away from instances that are not good exemplars of category (Eaves et al., 2016).

The above notions resonate with theories suggesting that infants identify phonological distinctions by observing and processing statistical regularities in the speech stream (e.g., Pierrehumbert, 2003). This construal of language development posits that infants form a phonological system through a gradual process of matching acoustic input with memories of similar events in an active constructive process. Infants' perceptual development may therefore admit a crucial role for variability, as within-category variability can allow infants to compute statistics over multiple dimensions concurrently (Pierrehumbert, 2003).

Vowel variability may benefit infant learning in a second way: Lower levels of variability and complexity can lead to habituation, which can counteract learning by causing low attention (Colombo & Mitchell, 2009; Hunter et al., 1983; Paulus, 2022). In this sense, a greater degree of variability may be necessary for attention and learning in phonetic category development (Christiansen & Chater, 2016; Raviv et al., 2022). For example, an experimental study has shown that 14-month-old infants learned faster from a single speaker who spoke with a greater degree of variability in duration, pitch, and pitch variability (Galle et al., 2015). The notion that variability in the speech stream can attract infants' attention goes hand in hand with the results obtained in this study that f_0, f_0 variability and articulation rate all exhibited a greater degree of variability within Danish IDS than within ADS. The beneficial role of variability may thus consist primarily in its ability to grab and maintain infant attention (Englund, 2018).

A third possibility is that the greater degree of vowel variability could potentially be a side effect of other articulatory features specific to IDS. This may include an elevated pitch and pitch variability, which has been shown to impact both F1 and F2 measures (McMurray et al., 2013). Relatedly, the tendency for caregivers to raise their larynx in IDS – either to convey non-threatening behavior (Kalashnikova et al., 2017), to mimic infant production (Cristia, 2013; Polka et al., 2022), to grab infant attention (Masapollo et al., 2016), or to convey positive affect (Benders, 2013; Saint-Georges et al., 2013; Singh et al., 2002), or a combination thereof – would produce a shorter vocal tract and result in an increase in both the first and second formants for all of the vowels. The results indeed indicated a leftwards and downwards transformation (i.e., raised F1 and F2 values) in the IDS formant centroids for the majority of the vowel categories investigated (cf. Table 5).

Another articulatory factor that might shorten the vocal tract and produce the shifts in formant frequencies would be through smiling. Smiling involves a retraction of the lips and widening of the mouth, with a resultant increase in the first and second vowel formants (Barthel & Quené, 2015; Tartter, 1980). The acoustic origin of smiling in the animal kingdom has been posited to derive from the desire to raise the resonant frequencies of the vocal tract to sound smaller and convey appeasement towards others (Ohala, 1980; Xu & Chuenwattanapranithi, 2007). This last explanation would admit a crucial role for the audio-visual component of language development and would conform to evidence indicating that infants can integrate audio-visual speech stimuli at an early point in development (Cox, Keren-Portnoy, et al., 2022) and flexibly take advantage of intersensory redundancy when processing complex audio-visual speech stimuli (Bastianello et al., 2022; Hillairet de Boisferon et al., 2017; Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015). Infants also preferentially attend to infant-directed faces (H. I. Kim & Johnson, 2014), and smiling during interaction produces a greater frequency of speechlike syllabic infant vocalisations (Hsu et al., 2001). The observed formant raising in IDS could thus be in part motivated by these visual and emotional accompaniments to verbal communication (Englund, 2018). It is important to recognize that raising the formants in IDS does not necessarily mean that caregivers cannot simultaneously expand the vowel space. These two features of IDS can work together, and exploring how they interact and contribute to the proposed functions of IDS will be a valuable topic for future research.

Limitations & Future Directions

The results of this study highlight the need for detailed cross-linguistic analysis of the acoustic properties of IDS and demonstrate the value of comparatively and critically incorporating statistical information from prior studies (Brand et al., 2019; Devezer et al., 2019; Fusaroli, Grossman, et al., 2021). There are several limitations of this study, which are important to keep in mind for planning future studies. The first limitation concerns the characteristics of our participant sample. The lack of age-related effects for most of the acoustic features of IDS examined here must be interpreted in light of relying on a cross-sectional sample, as some of the ages were represented by only one or two participants. Because the infants of the caregivers under investigation were between 11 and 24 months of age, moreover, they would already have a certain level of exposure and knowledge of the phonetic categories of their first language (cf., Kuhl, 2000). The current results, for example, cannot rule out the possibility that caregivers initially produce a greater degree of vowel separability to younger infants, as indicated in studies of other languages (cf., Hartman et al., 2017; Kuhl et al., 1997; H. Liu et al., 2003). If IDS is construed as a form of speaker adaptation, we might also expect infants' developmental status as well as the kinship status and familiarity of the interlocutor to affect the generalisability of the results. Future research exploring the effects of diverse speaker characteristics would provide important insights into factors affecting the acoustic properties of IDS (e.g., Kaplan et al., 2001; Lam-Cassettari & Kohlhoff, 2020; Steen & Englund, 2021; Woolard et al., 2022). The recent expansion in the availability of cross-linguistic data (e.g., MacWhinney, 2014) and technological improvements to perform automatic transcription (e.g., Cychosz et al., 2021; Räsänen, Seshadri, Lavechin, M, et al., 2021) allow more detailed analyses of how IDS differs across individuals, genders, languages and infant ages. To take full advantage of cumulative science practices, we would encourage researchers to share utterance- and vowel-level data in open repositories. This approach to future investigations of IDS would allow for a higher resolution of how the acoustic properties of IDS differ between individuals, languages, and infant ages. In line with this recommendation, the data and code used in this manuscript are available in the following open repository: https://osf.io/ywf9m/?view_only=d99fc6dbc61546febff619b8674a7943.

A second limitation of this study concerns the focus on Danish without any comparison to other languages or cultures that differ in a key moderator of interest. For example, a comparison between Danish and a similar vowel-rich language without quantity distinctions would provide important insights into how phonological structure can influence the acoustic expression of IDS. By conducting theory-driven comparisons of the acoustic properties of IDS across a diverse intersection of languages and cultures (Christiansen et al., 2022; Deffner et al., 2021), we can obtain a fuller picture of the cross-cultural and cross-linguistic variables that moderate the acoustic expression of the speech style. One approach would be to investigate culturally similar societies that differ in the key phonological variables of interest (e.g., number of front vowels, consonantal lenition or schwa assimilation). By keeping unobserved cultural variables roughly comparable, such as socio-economic status and child-rearing practices, we can isolate the influence of language structure and facilitate causal inference about specific acoustic modifications in caregivers' IDS. Another approach would include computational models with IDS and ADS speech data from a diverse intersection of languages to formulate testable predictions on whether the benefits of the speech style derive from the improved information structure of the speech signal (e.g., Eaves et al., 2016; Ludusan et al., 2021; McMurray et al., 2009). This in turn would allow us to examine and answer questions as to whether caregivers' acoustic modifications in IDS relate mainly to phonological structure or cultural practice.

Conclusion

The present study was designed as a comparative analysis of the acoustic properties of Danish ADS and IDS. Firstly, the results indicated that pitch, melody and articulation rate in IDS were modified by Danish caregivers in similar ways to other languages. However, Danish vowels in IDS were articulated with no acoustic exaggeration and with similar durations compared to ADS. The findings for the vocalic properties of Danish IDS here thus add to a small subset of studies finding no vocalic exaggeration in IDS and provide further evidence of the peculiar nature of Danish sound structure. Secondly, articulation rate was the only property of Danish IDS to exhibit dynamic change and became more similar to Danish ADS when directed to older infants. Thirdly, Danish caregivers produced a greater degree of within-vowel variability and a lower degree of between-vowel discriminability in IDS when compared with ADS. These findings highlight the need for future studies to conduct theory-driven comparisons of the acoustic expression of IDS across a wide intersection of languages with distinct phonological systems.

List of References

Adank, P., Smits, R., & Van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America*, *116*(5), 3099–3107.

Amano, S., Nakatani, T., & Kondo, T. (2006). Fundamental frequency of infants' and parents' utterances in longitudinal recordings. *The Journal of the Acoustical Society of America*, *119*(3), 1636–1647. https://doi.org/10.1121/1.2161443

Barthel, H., & Quené, H. (2015). *Acoustic-phonetic properties of smiling revised: Measurements on a natural video corpus*. Proceedings of the 18th international congress of phonetic sciences.

Basbøll, H. (2005). *The phonology of Danish*. Oxford University Press.

Bastianello, T., Keren-Portnoy, T., Majorano, M., & Vihman, M. (2022). Infant looking preferences towards dynamic faces: A systematic review. *Infant Behavior and Development*, *67*, 101709.

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

Bergelson, E., Casillas, M., Soderstrom, M., Seidl, A., Warlaumont, A. S., & Amatuni, A. (2019). What do North American babies hear? A large-scale cross-corpus analysis. *Developmental Science*, *22*(1), e12724.

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

Bleses, D., Basbøll, H., & Vach, W. (2011). Is Danish difficult to acquire? Evidence from Nordic past-tense studies. *Language and Cognitive Processes*, *26*(8), 1193–1231.

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O., & Basbøll, H. (2008a). Early vocabulary development in Danish and other languages: A CDI-based comparison. *Journal of Child Language*, *35*(3), 619–650.

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O., & Basbøll, H. (2008b). The Danish Communicative Developmental Inventories: Validity and main developmental trends. *Journal of Child Language*, *35*(3), 651–669.

Boersma, P., & Weenink, D. (2022). *Praat: Doing phonetics by computer [Computer program]. Version 6.2.05, retrieved 5 January 2022 from http://www.praat.org/.*

Bohn, O.-S. (2013). *Acoustic characteristics of Danish infant directed speech*. *19*(1), 060055.

Bradlow, A. R., Kraus, N., & Hayes, E. (2003). Speaking clearly for children with learning disabilities. *Journal of Speech, Language, and Hearing Research*, *46*(1), 80–97.

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.

Brand, C. O., Ounsley, J. P., van der Post, D. J., & Morgan, T. J. H. (2019). Cumulative Science via Bayesian Posterior Passing: An introduction. *Meta-Psychology*, *3*. https://doi.org/10.15626/MP.2017.840

Broesch, T. L., & Bryant, G. A. (2015). Prosody in infant-directed speech is similar across western and traditional cultures. *Journal of Cognition and Development*, *16*(1), 31–43.

Bryant, G. A. (2022). Vocal communication across cultures: Theoretical and methodological issues. *Philosophical Transactions of the Royal Society B*, *377*(1841), 20200387. https://doi.org/10.1098/rstb.2020.0387

Bryant, G. A., & Barrett, H. C. (2007). Recognizing intentions in infant-directed speech: Evidence for universals. *Psychological Science*, *18*(8), 746–751. https://doi.org/10.1111/j.1467-9280.2007.01970.x

Bürkner, P.-C. (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

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., 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

Casillas, M., Brown, P., & Levinson, S. C. (2020). Early language experience in a Tseltal Mayan village. *Child Development*, *91*(5), 1819–1835. https://doi.org/10.1111/cdev.13349

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*, E62. https://doi.org/10.1017/S0140525X1500031X

Christiansen, M. H., Kallens, P. C., & Trecca, F. (2022). Towards A Comparative Approach to Language Acquisition. *Current Directions in Psychological Science*, *31*, 131–138. https://doi.org/10.1177/09637214211049229

Colombo, J., & Mitchell, D. W. (2009). Infant visual habituation. *Neurobiology of Learning and Memory*, *92*(2), 225–234.

Cox, C. M. M., Bergmann, C., Fowler, E., Keren-Portnoy, T., Roepstorff, A., Bryant, G. A., & Fusaroli, R. (2022). A Systematic Review and Bayesian Meta-Analysis of the Acoustic Features of Infant-Directed Speech. *Nature Human Behaviour*, 1–20. https://doi.org/10.1038/s41562-022-01452-1

Cox, C. M. M., Keren-Portnoy, T., Roepstorff, A., & Fusaroli, R. (2022). A Bayesian meta-analysis of infants' ability to perceive audio–visual congruence for speech. *Infancy*, *27*(1), 67–96.

Cristia, A. (2013). Input to language: The phonetics and perception of infant-directed speech. *Language and Linguistics Compass*, *7*(3), 157–170.

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

Cychosz, M., Cristia, A., Bergelson, E., Casillas, M., Baudet, G., Warlaumont, A. S., Scaff, C., Yankowitz, L., & Seidl, A. (2021). Vocal development in a large-scale crosslinguistic corpus. *Developmental Science*, e13090. https://doi.org/10.1111/desc.13090

De Boer, B., & Kuhl, P. K. (2003). Investigating the role of infant-directed speech with a computer model. *Acoustics Research Letters Online*, *4*(4), 129–134. https://doi.org/10.1121/1.1613311

De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, *41*(2), 385–390.

Deffner, D., Rohrer, J. M., & McElreath, R. (2021). A Causal Framework for Cross-Cultural Generalizability. *Advances in Methods and Practices in Psychological Science*, *5*(3). https://doi.org/10.1177/25152459221106366

Degottex, G., Kane, J., Drugman, T., Raitio, T., & Scherer, S. (2014). *COVAREP—A collaborative voice analysis repository for speech technologies*. 960–964. https://doi.org/10.1109/ICASSP.2014.6853739

Devezer, B., Nardin, L. G., Baumgaertner, B., & Buzbas, E. O. (2019). Scientific discovery in a model-centric framework: Reproducibility, innovation, and epistemic diversity. *PloS One*, *14*(5), e0216125.

Dideriksen, C., Christiansen, M. H., Dingemanse, M., Højmark-Bertelsen, M., Johansson, C., Tylén, K., & Fusaroli, R. (2022). Language specific constraints on conversation: Evidence from Danish and Norwegian. *PsyArxiv*.

Dideriksen, C., & Fusaroli, R. (2018). *CDS is not what you think-Hypoarticulation in Danish Child Directed Speech*.

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).

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

Englund, K. T. (2018). Hypoarticulation in infant-directed speech. *Applied Psycholinguistics*, *39*(1), 67–87.

Englund, K. T., & Behne, D. M. (2005). Infant directed speech in natural interaction—Norwegian vowel quantity and quality. *Journal of Psycholinguistic Research*, *34*(3), 259–280. https://doi.org/10.1007/s10936-005-3640-7

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.

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.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, 1497–1510. https://doi.org/10.2307/1130938

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, *10*(3), 279–293.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology*, *27*(2), 209.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, *20*(1), 104–113. https://doi.org/10.1037/0012-1649.20.1.104

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

Floccia, C., Keren-Portnoy, T., DePaolis, R., Duffy, H., Delle Luche, C., Durrant, S., White, L., Goslin, J., & Vihman, M. (2016). British English infants segment words only with exaggerated infant-directed speech stimuli. *Cognition*, *148*, 1–9. https://doi.org/10.1016/j.cognition.2015.12.004

Fusaroli, R., Grossman, R., Bilenberg, N., Cantio, C., Jepsen, J. R. M., & Weed, E. (2021). Toward a cumulative science of vocal markers of autism: A cross-linguistic meta-analysis-based investigation of acoustic markers in American and Danish autistic children. *Autism Research*, 1–12. https://doi.org/10.1002/aur.2661

Fusaroli, R., Weed, E., Fein, D., & Naigles, L. (2019). Hearing me hearing you: Reciprocal effects between child and parent language in autism and typical development. *Cognition*, *183*, 1–18. https://doi.org/10.1016/j.cognition.2018.10.022

Fusaroli, R., Weed, E., Fein, D., & Naigles, L. (2021). Caregiver linguistic alignment to autistic and typically developing children. *PsyArxiv*.

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization in Bayesian workflow. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *182*(2), 389–402.

Galle, M. E., Apfelbaum, K. S., & McMurray, B. (2015). The role of single talker acoustic variation in early word learning. *Language Learning and Development*, *11*(1), 66–79.

García-Sierra, A., Ramírez-Esparza, N., & Kuhl, P. K. (2016). Relationships between quantity of language input and brain responses in bilingual and monolingual infants. *International Journal of Psychophysiology*, *110*, 1–17.

García-Sierra, A., Ramírez-Esparza, N., Wig, N., & Robertson, D. (2021). Language learning as a function of infant directed speech (IDS) in Spanish: Testing neural commitment using the positive-MMR. *Brain and Language*, *212*, 104890.

Gelman, A., Simpson, D., & Betancourt, M. (2017). The prior can often only be understood in the context of the likelihood. *Entropy*, *19*(10), 555.

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), 1–10. https://doi.org/10.1038/s41598-017-13883-2

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19*(5), 515–523. https://doi.org/10.1111/j.1467-9280.2008.02117.x

Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby) talk to me: The social context of infant-directed speech and its effects on early language acquisition. *Current Directions in Psychological Science*, *24*(5), 339–344. https://doi.org/10.1177/0963721415595345

Grønnum, N. (1998). Danish. *Journal of the International Phonetic Association*, *28*(1–2), 99–105.

Han, M., de Jong, N. H., Kager, R., & Bertolini AB, K. M. (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*, *2*(1), 331–344.

Han, M., De Jong, N., & Kager, R. (2020). Pitch properties of infant-directed speech specific to word-learning contexts: A cross-linguistic investigation of Mandarin Chinese and Dutch. *Journal of Child Language*, *47*(1), 85–111. https://doi.org/10.1017/S0305000919000813

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

Hedeker, D., Mermelstein, R. J., & Demirtas, H. (2008). An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. *Biometrics*, *64*(2), 627–634.

Hillairet de Boisferon, A., Tift, A. H., Minar, N. J., & Lewkowicz, D. J. (2017). Selective attention to a talker's mouth in infancy: Role of audiovisual temporal synchrony and linguistic experience. *Developmental Science*, *20*(3), e12381.

Hilton, C. B., Moser, C. J., Bertolo, M., Lee-Rubin, H., Amir, D., Bainbridge, C. M., Simson, J., Knox, D., Glowacki, L., & Galbarczyk, A. (2021). Acoustic regularities in infant-directed speech and song across cultures. *BioRxiv*, 2020–04.

Hilton, C. B., Moser, C. J., Bertolo, M., Lee-Rubin, H., Amir, D., Bainbridge, C. M., Simson, J., Knox, D., Glowacki, L., Galbarczyk, A., Jasienska, G., Ross, C. T., Neff, M. B., Martin, A., Cirelli, L. K., Trehub, S. E., Song, J., Kim, M., Schachner, A., … Mehr, S. A. (2022). Acoustic regularities in infant-directed speech and song across cultures. *Nature Human Behaviour*, 1–12. https://doi.org/10.1038/s41562-022-01410-x

Højen, A., & Nazzi, T. (2016). Vowel bias in Danish word-learning: Processing biases are language-specific. *Developmental Science*, *19*(1), 41–49.

Hsu, H.-C., Fogel, A., & Messinger, D. S. (2001). Infant non-distress vocalization during mother-infant face-to-face interaction: Factors associated with quantitative and qualitative differences☆. *Infant Behavior and Development*, *24*(1), 107–128.

Hunter, M. A., Ames, E. W., & Koopman, R. (1983). Effects of stimulus complexity and familiarization time on infant preferences for novel and familiar stimuli. *Developmental Psychology*, *19*(3), 338.

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*, 170306. https://doi.org/10.1098/rsos.170306

Kaplan, P. S., Bachorowski, J., Smoski, M. J., & Zinser, M. (2001). Role of clinical diagnosis and medication use in effects of maternal depression on infant-directed speech. *Infancy*, *2*(4), 537–548. https://doi.org/10.1207/S15327078IN0204_08

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

Kim, H., Hasegawa-Johnson, M., & Perlman, A. (2011). Vowel contrast and speech intelligibility in dysarthria. *Folia Phoniatrica et Logopaedica*, *63*(4), 187–194.

Kim, H. I., & Johnson, S. P. (2014). Detecting ‘infant-directedness’ in face and voice. *Developmental Science*, *17*(4), 621–627.

Kitamura, C., & Burnham, D. (1998). *Acoustic and affective qualities of IDS in English*. Fifth international conference on spoken language processing. http://www.isca-speech.org/archive.

Kitamura, C., & Lam, C. (2009). Age-specific preferences for infant-directed affective intent. *Infancy*, *14*(1), 77–100. https://doi.org/10.1080/15250000802569777

Kjærbæk, L., & Basbøll, H. (2016). Interaction between input frequency, transparency and productivity in acquisition of noun plural inflection in Danish. *Poznan Studies in Contemporary Linguistics*, *52*(4), 663–686.

Ko, E.-S., Seidl, A., Cristia, A., Reimchen, M., & Soderstrom, M. (2016). Entrainment of prosody in the interaction of mothers with their young children. *Journal of Child Language*, *43*(2), 284–309. https://doi.org/10.1017/S0305000915000410

Kondaurova, M. V., Bergeson, T. R., & Xu, H. (2013). Age-Related Changes in Prosodic Features of Maternal Speech to Prelingually Deaf Infants with Cochlear Implants. *Infancy*, *18*(5), 825–848. https://doi.org/10.1111/infa.12010

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences*, *97*(22), 11850–11857.

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

Kvarven, A., Strømland, E., & Johannesson, M. (2020). Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nature Human Behaviour*, *4*(4), 423–434.

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

Lam, J., & Tjaden, K. (2013). Acoustic-perceptual relationships in variants of clear speech. *Folia Phoniatrica et Logopaedica*, *65*(3), 148–153.

Lam, J., Tjaden, K., & Wilding, G. (2012). Acoustics of clear speech: Effect of instruction. *Journal of Speech, Language, and Hearing Research*, *55*(6), 1807–1821.

Lam-Cassettari, C., & Kohlhoff, J. (2020). Effect of maternal depression on infant-directed speech to prelinguistic infants: Implications for language development. *PLoS One*, *15*(7), e0236787.

Lavechin, M., Bousbib, R., Bredin, H., Dupoux, E., & Cristia, A. (2020). An open-source voice type classifier for child-centered daylong recordings. *ArXiv Preprint ArXiv:2005.12656*.

Lee, C. S., Kitamura, C., Burnham, D., & McAngus Todd, N. P. (2014). On the rhythm of infant-versus adult-directed speech in Australian English. *The Journal of the Acoustical Society of America*, *136*(1), 357–365. https://doi.org/10.1121/1.4883479

Lemoine, N. P. (2019). Moving beyond noninformative priors: Why and how to choose weakly informative priors in Bayesian analyses. *Oikos*, *128*(7), 912–928.

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences*, *109*(5), 1431–1436.

Liu, H., Kuhl, P. K., & Tsao, F. (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. (2005). The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *The Journal of the Acoustical Society of America*, *117*(6), 3879–3889.

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America*, *49*(2B), 606–608.

Ludusan, B., Mazuka, R., & Dupoux, E. (2021). Does Infant-Directed Speech Help Phonetic Learning? A Machine Learning Investigation. *Cognitive Science*, *45*(5), e12946.

Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word learning in infant-and adult-directed speech. *Language Learning and Development*, *7*(3), 185–201.

MacWhinney, B. (2014). *The CHILDES project: Tools for analyzing talk, Volume II: The database*. Psychology Press. https://doi.org/10.4324/9781315805641

Marklund, E., & Gustavsson, L. (2020). The dynamics of vowel hypo-and hyperarticulation in Swedish infant-directed speech to 12-month-olds. *Frontiers in Communication*, 79.

Marklund, E., Marklund, U., & Gustavsson, L. (2021). An Association Between Phonetic Complexity of Infant Vocalizations and Parent Vowel Hyperarticulation. *Frontiers in Psychology*, 2873.

Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., & Cristia, A. (2015). Mothers speak less clearly to infants than to adults: A comprehensive test of the hyperarticulation hypothesis. *Psychological Science*, *26*(3), 341–347. https://doi.org/10.1177/0956797614562453

Masapollo, M., Polka, L., & Ménard, L. (2016). When infants talk, infants listen: Pre-babbling infants prefer listening to speech with infant vocal properties. *Developmental Science*, *19*(2), 318–328.

McCloy, D. (2016). *phonR: tools for phoneticians and phonologists. R package version 1.0-7.*

McElreath, R. (2018). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC.

McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: Insights from a computational approach. *Developmental Science*, *12*(3), 369–378. https://doi.org/10.1111/j.1467-7687.2009.00822.x

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

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. (1978). Vowel space normalization in synthetic stimuli. *The Journal of the Acoustical Society of America*, *63*(S1), S5–S5.

Nguyen, V., Versyp, O., Cox, C. M. M., & Fusaroli, R. (2022). A systematic review and Bayesian meta-analysis of the development of turn taking in adult-child vocal interactions. *Child Development*. https://doi.org/10.1111/cdev.13754

Niwano, K., & Sugai, K. (2002). Intonation contour of Japanese maternal infant-directed speech and infant vocal response. *The Japanese Journal of Special Education*, *39*(6), 59–68.

Ohala, J. J. (1980). The acoustic origin of the smile. *The Journal of the Acoustical Society of America*, *68*(S1), S33–S33.

Panneton, R., Kitamura, C., Mattock, K., & Burnham, D. (2006). Slow speech enhances younger but not older infants' perception of vocal emotion. *Research in Human Development*, *3*(1), 7–19. https://doi.org/10.1207/s15427617rhd0301_2

Papoušek, M., Papoušek, H., & Symmes, D. (1991). The meanings of melodies in motherese in tone and stress languages. *Infant Behavior and Development*, *14*(4), 415–440. https://doi.org/10.1016/0163-6383(91)90031-M

Parola, A., Simonsen, A., Lin, J. M., Zhou, Y., Huiling, W., Ubukata, S., Koelkebeck, K., Bliksted, V., & Fusaroli, R. (2022). Voice patterns as markers of schizophrenia: Building a cumulative generalizable approach via cross-linguistic and meta-analysis based investigation. *MedRxiv*.

Paulus, M. (2022). Should infant psychology rely on the violation-of-expectation method? Not anymore. *Infant and Child Development*, *31*(1), e2306.

Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn locally, think globally: Exemplar variability supports higher-order generalization and word learning. *Psychological Science*, *21*(12), 1894–1902. https://doi.org/10.1177/0956797610389189

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), 1–14. https://doi.org/10.1038/srep34273

Polka, L., Masapollo, M., & Ménard, L. (2022). Setting the stage for speech production: Infants prefer listening to speech sounds with infant vocal resonances. *Journal of Speech, Language, and Hearing Research*, *65*(1), 109–120.

Pons, F., Bosch, L., & Lewkowicz, D. J. (2015). Bilingualism modulates infants' selective attention to the mouth of a talking face. *Psychological Science*, *26*(4), 490–498.

Porritt, L. L., Zinser, M. C., Bachorowski, J.-A., & Kaplan, P. S. (2014). Depression diagnoses and fundamental frequency-based acoustic cues in maternal infant-directed speech. *Language Learning and Development*, *10*(1), 51–67.

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

Raneri, D. P. (2015). Infant-directed speech: Maternal pitch variability, rate of speech, and child language outcomes. *ProQuest Dissertations Publishing*, 1–56.

Räsänen, O., Seshadri, S., Lavechin, M., Cristia, A., & Casillas, M. (2021). ALICE: An open-source tool for automatic measurement of phoneme, syllable, and word counts from child-centered daylong recordings. *Behavior Research Methods*, *53*(2), 818–835. https://doi.org/10.3758/s13428-020-01460-x

Räsänen, O., Seshadri, S., Lavechin, M, Cristia, A., & Casillas, M. (2021). An open-source tool for automatic measurement of phoneme, syllable, and word counts from child-centered daylong recordings. *Behavior Research Methods*, *53*(2), 818–835. https://doi.org/10.3758/s13428-020-01460-x

Rast, P., & Ferrer, E. (2018). A mixed-effects location scale model for dyadic interactions. *Multivariate Behavioral Research*, *53*(5), 756–775.

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

Ravanelli, M., & Bengio, Y. (2018). *Speaker recognition from raw waveform with sincnet*. 1021–1028.

Raviv, L., Lupyan, G., & Green, S. C. (2022). How variability shapes learning and generalization. *Trends in Cognitive Sciences*, *26*(6), 462–483.

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.

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*(2), 339–349. https://doi.org/10.1111/j.1467-7687.2008.00786.x

Rost, G. C., & McMurray, B. (2010). Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning. *Infancy*, *15*(6), 608–635. https://doi.org/10.1111/j.1532-7078.2010.00033.x

RStudio Team. (2020). *RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.*

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology*, *69*, 181–203. https://doi.org/10.1146/annurev-psych-122216-011805

Saint-Georges, C., Chetouani, M., Cassel, R., Apicella, F., Mahdhaoui, A., Muratori, F., Laznik, M.-C., & Cohen, D. (2013). Motherese in interaction: At the cross-road of emotion and cognition?(A systematic review). *PloS One*, *8*(10), e78103.

Searl, J., & Evitts, P. M. (2013). Tongue–palate contact pressure, oral air pressure, and acoustics of clear speech. *Journal of Speech, Language, and Hearing Research*, *56*(3), 826–839.

Seidl, A., & Cristià, A. (2008). Developmental changes in the weighting of prosodic cues. *Developmental Science*, *11*(4), 596–606. https://doi.org/10.1111/j.1467-7687.2008.00704.x

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

Smith, N. A., & Trainor, L. J. (2008). Infant-directed speech is modulated by infant feedback. *Infancy*, *13*(4), 410–420. https://doi.org/10.1080/15250000802188719

Soderstrom, M., Seidl, A., Nelson, D. G. K., & Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language*, *49*(2), 249–267. https://doi.org/10.1016/S0749-596X(03)00024-X

Song, 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

Spinelli, M., Fasolo, M., & Mesman, J. (2017). Does prosody make the difference? A meta-analysis on relations between prosodic aspects of infant-directed speech and infant outcomes. *Developmental Review*, *44*, 1–18. https://doi.org/10.1016/j.dr.2016.12.001

Steen, V. B., & Englund, N. (2021). Child-directed Speech in a Norwegian Kindergarten Setting. *Scandinavian Journal of Educational Research*, 1–14. https://doi.org/10.1080/00313831.2021.1897873

Steinlen, A. K. (2005). *The influence of consonants on native and non-native vowel production: A cross-linguistic study* (Vol. 30). Gunter Narr Verlag.

Stern, D. N., Spieker, S., Barnett, R., & 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

Tartter, V. C. (1980). Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception & Psychophysics*, *27*(1), 24–27.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*(1), 53–71.

Trecca, F., Bleses, D., Højen, A., Madsen, T. O., & Christiansen, M. H. (2020). When too many vowels impede language processing: An eye-tracking study of Danish-learning children. *Language and Speech*, *63*(4), 898–918.

Trecca, F., Bleses, D., Madsen, T. O., & Christiansen, M. H. (2018). Does sound structure affect word learning? An eye-tracking study of Danish learning toddlers. *Journal of Experimental Child Psychology*, *167*, 180–203.

Trecca, F., McCauley, S. M., Andersen, S. R., Bleses, D., Basbøll, H., Højen, A., Madsen, T. O., Ribu, I. S. B., & Christiansen, M. H. (2019). Segmentation of highly vocalic speech via statistical learning: Initial results from Danish, Norwegian, and English. *Language Learning*, *69*(1), 143–176.

Trecca, F., Tylén, K., Højen, A., & Christiansen, M. H. (2021). Danish as a Window Onto Language Processing and Learning. *Language Learning*, *71*, 799–833. https://doi.org/10.1111/lang.12450

Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences*, *104*(33), 13273–13278. https://doi.org/10.1073/pnas.0705369104

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Vosoughi, S., & Roy, D. K. (2012). A longitudinal study of prosodic exaggeration in child-directed speech. *Proceedings of the Speech Prosody, 6th International Conference*.

Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2014). A social feedback loop for speech development and its reduction in autism. *Psychological Science*, *25*(7), 1314–1324. https://doi.org/10.1177/0956797614531023

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

Werker, J. F., & Tees, R. C. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology*, *50*(1), 509–535. https://doi.org/10.1146/annurev.psych.50.1.509

Whitfield, J. A., & Goberman, A. M. (2014). Articulatory–acoustic vowel space: Application to clear speech in individuals with Parkinson's disease. *Journal of Communication Disorders*, *51*, 19–28.

Whitfield, J. A., & Goberman, A. M. (2017). Articulatory-acoustic vowel space: Associations between acoustic and perceptual measures of clear speech. *International Journal of Speech-Language Pathology*, *19*(2), 184–194.

Whitfield, J. A., & Mehta, D. D. (2019). Examination of clear speech in Parkinson disease using measures of working vowel space. *Journal of Speech, Language, and Hearing Research*, *62*(7), 2082–2098.

Williams, D. R., Zimprich, D. R., & Rast, P. (2019). A Bayesian nonlinear mixed-effects location scale model for learning. *Behavior Research Methods*, *51*(5), 1968–1986.

Woolard, A., Lane, A. E., Campbell, L. E., Whalen, O. M., Swaab, L., Karayanidis, F., Barker, D., Murphy, V., & Benders, T. (2022). Infant and Child-Directed Speech Used with Infants and Children at Risk or Diagnosed with Autism Spectrum Disorder: A Scoping Review. *Review Journal of Autism and Developmental Disorders*, *9*, 290–306. https://doi.org/10.1007/s40489-021-00253-y

Xu, Y., & Chuenwattanapranithi, S. (2007). Perceiving anger and joy in speech through the size code. *Proceedings of the International Conference on Phonetic Sciences*, *16*, 2105–2108.

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

Yellowlees, A., Bursa, F., Fleetwood, K. J., Charlton, S., Hirst, K. J., Sun, R., & Fusco, P. C. (2016). The appropriateness of robust regression in addressing outliers in an anthrax vaccine potency test. *Bioscience*, *66*(1), 63–72.

Supplementary Materials

S0: Information about Participants

Subject	Sex	Age	Subject	Sex	Age
PV	m	11	**LK**	m	16
RV	f	11	**LV**	m	16
TL	m	11	**MV**	m	16
DM	m	12	**SAF**	m	19
LM	m	12	**SC**	f	19
ME	m	12	**AF**	m	20
KA	m	14	**JE**	f	21
MPS	f	14	**AN**	m	22
SF	m	14	**IA**	f	23
SS	m	14	**MA**	f	23
SA	f	15	**MS**	f	23
CL	f	16	**JA**	m	24
KV	m	16	**TM**	f	24

Table S0.1: Details about the 26 participants under investigation, ordered according to age. Sex refers to the biological sex of the infant. The column Age refers to infants' age in months.

S1: Model Structures & Priors

S1.1: Our Choices of Priors

For the models with speech style as a predictor, we used the following formula and priors, with a logarithmic link function for all of the measures on the positive scale:

Model Structure 1: acous.variable ~ 0 + SpeechStyle + (0 + SpeechStyle|Subject),
sigma ~ 0 + SpeechStyle + (0 + SpeechStyle|Subject)

Measures	Intercepts	Sigmas	SD	DoF	Cor
Median f_0	N(5.5,0.5)	N(2, 1)	N(0,1)	G(2,.1)	lkj(2)
IQR of f_0	N(4,0.5)	N(2, 1)	N(0,1)	G(2,.1)	lkj(2)
Articulation rate	N(1,0.5)	N(0,.5)	N(0,1)	G(2,.1)	lkj(2)
Vowel Duration	N(-2.5,0.5)	N(1.1)	N(1,1)	G(2,.1)	lkj(2)
Vowel Space Area	N(0,0.3)	N(1, 0.5)	N(1,1)	G(2,.1)	lkj(2)
Effect Size Models	N(0,0.5)	N(0.1)	N(1,1)	G(2,.1)	lkj(2)
Location-scale	N(0.1.5)	N(0,1)	N(1,1)	G(2,.1)	lkj(2)
Discriminability	N(0.5,0.2)	N(0,2)	N(0,1)	G(2,.1)	lkj(2)

Table S1.1: Priors for the parameters in the varying-interceps, varying-slopes model for each acoustic parameter. N() refers to a normal distribution, G() indicates a gamma distribution, lkj() refers to the Lewandowski-Kurowicka-Joe distribution.

For the model with an interaction term between speech style and age, we used the same priors as above, with the additional Gaussian prior of N(0.02) on age. We ran the models with the following formula structure:

Model Structure 2: acous.variable ~ 0 + SpeechStyle + SpeechStyle:ChildAge + (0 + SpeechStyle|Subject)

For the mixed-effects location-scale model, we modelled the data with the following formula and the above priors:

Model Structure 3: F1/F2 ~ 0 + SpeechStyle:vowel + (0 + SpeechStyle:vowel|Subject),
sigma ~ 0 + SpeechStyle:vowel + (0 + SpeechStyle:vowel|Subject)

We ran these models with Hamiltonian Monte Carlo samplers with 2 parallel chains of 5000 iterations each, an adapt delta of 0.99 and a maximum tree depth of 20 to allow for a robust estimation process without divergent transitions. The influence of priors was assessed by i) performing prior and posterior predictive checks (cf. S1.2), prior-posterior update checks (cf. S1.3) as well as prior sensitivity checks (cf. S1.4). For all of the models, we ensured that i) the Rhat statistics were no higher than 1.0, ii) there were no divergent transitions in the process of estimation, and iii) the number of effective bulk and tail samples was above 200.

S1.2: Prior & Posterior Predictive Checks

As noted above, we performed quality checks of the models by carrying out prior and posterior predictive checks. As shown in Figures S1.2.1-S1.2.2, the prior predictive checks indicate that our priors predict values within the order of magnitude of the distribution. The posterior predictive checks indicate that the models have captured the distributions of data for each of the acoustic measures. These plots provide reassurance that our models capture relevant aspects of the overall distributions of dependent variables (cf., Gabry et al., 2019).

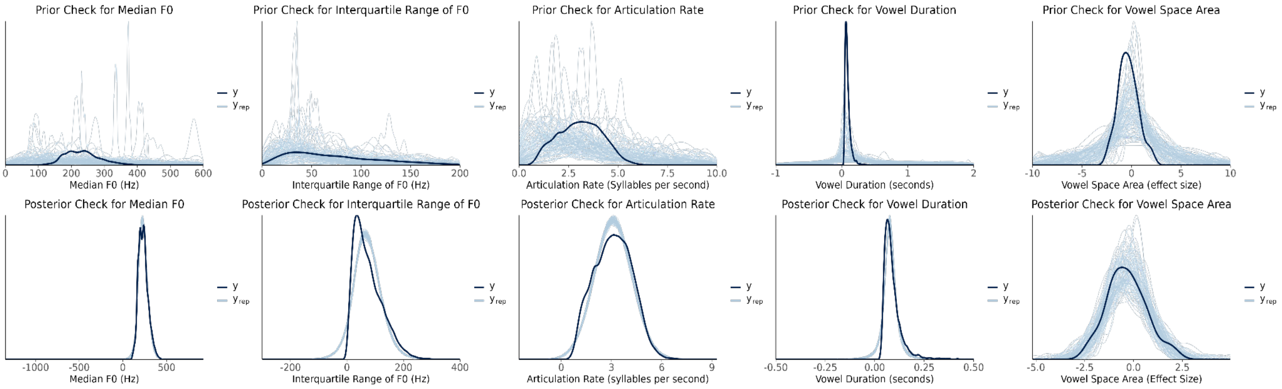


Figure S1.2.1: Panel of plots showing the prior and posterior predictive checks (in lightblue) and the data (in black) for each of the acoustic measures under investigation. The predictive checks are shown on the outcome scale.

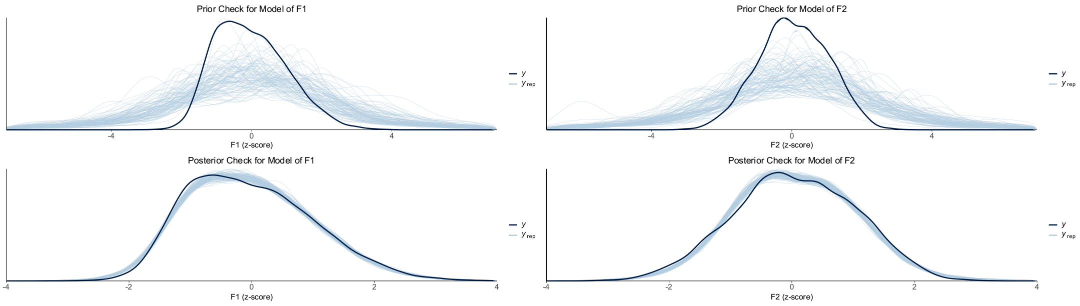
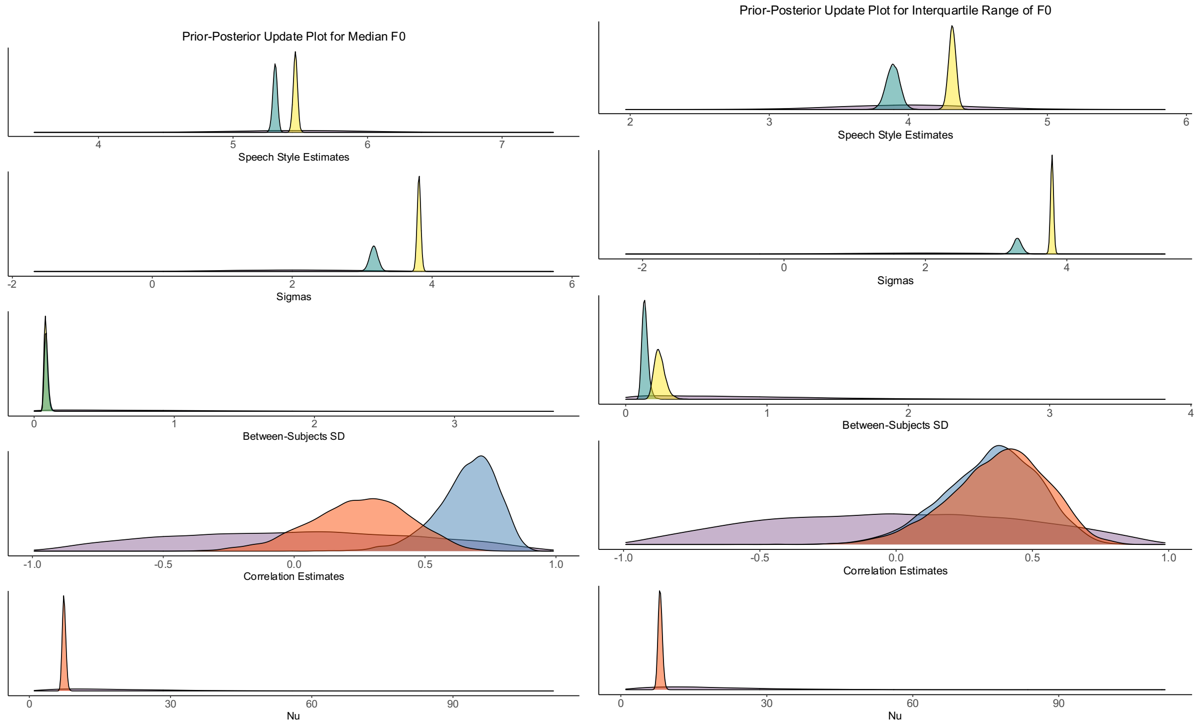
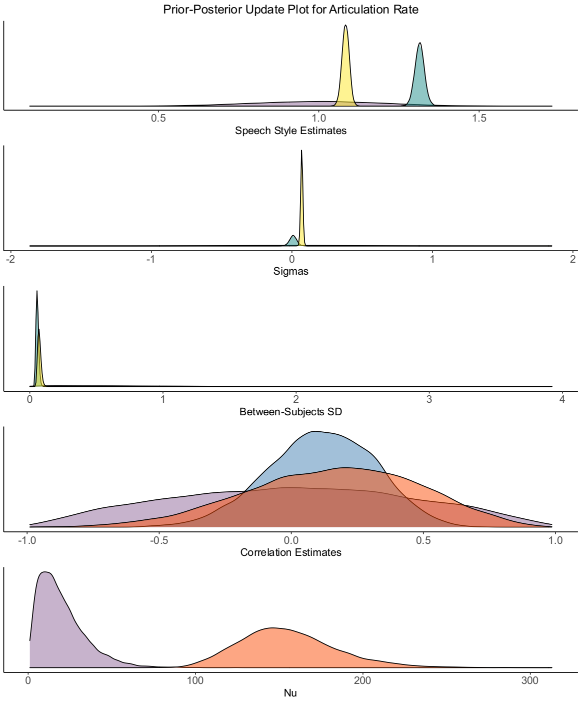


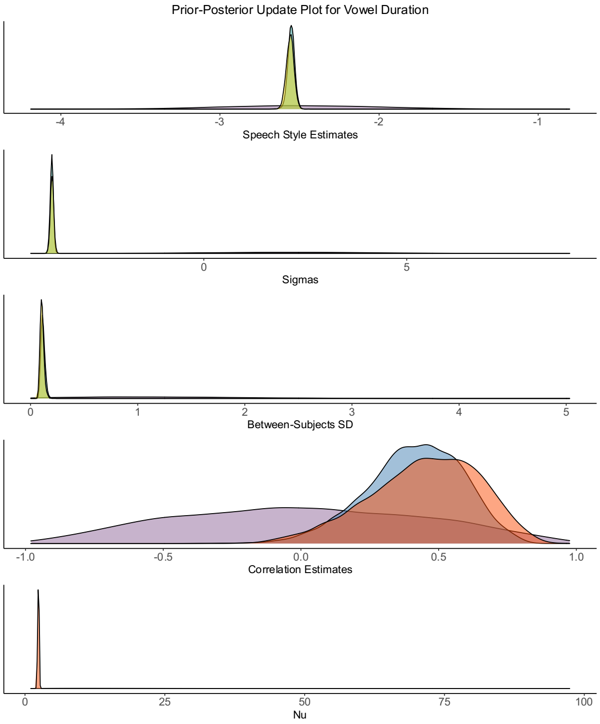
Figure S1.2.2: Panel of plots showing the prior and posterior predictive checks (in lightblue) and the data (in black) for the location-scale model of vowel categories. These plots indicate that our models have captured relevant aspects of the data.

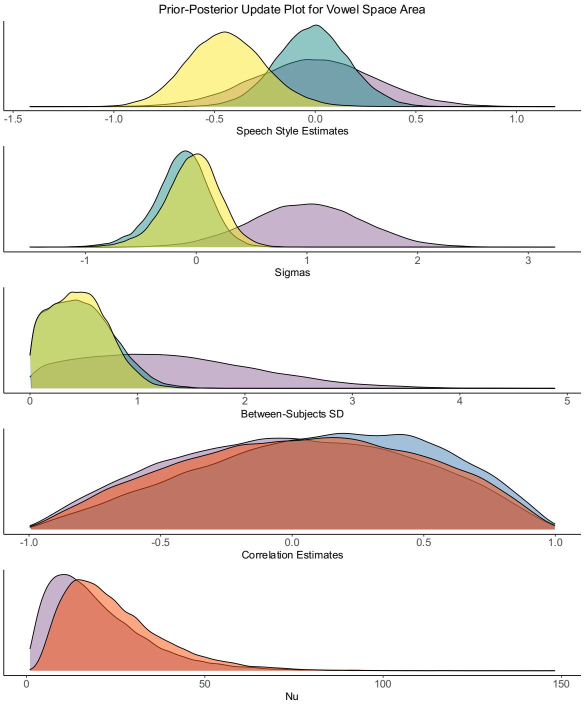
S1.3: Prior-Posterior Update Plots

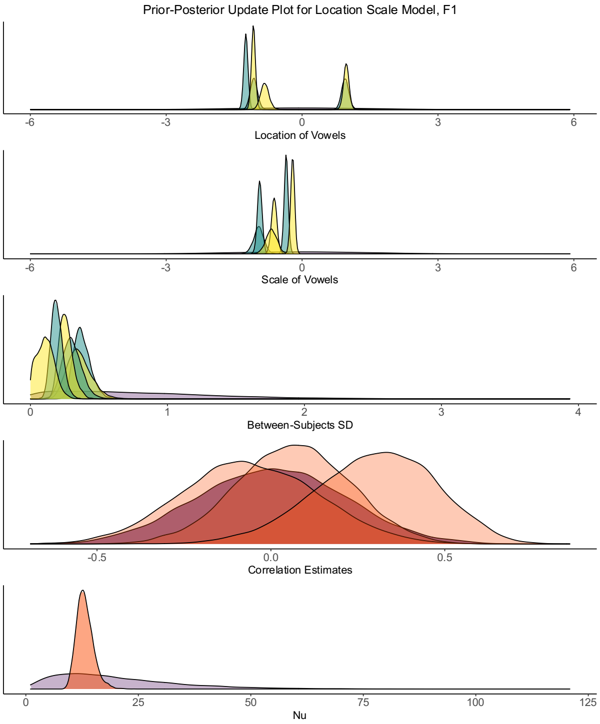
A second quality check of the models was carried out by plotting the prior distributions against the posterior estimates of the model. As shown in the below plots in Figure S1.3, the posteriors exhibit lower variance than the priors, suggesting that the models have learned from the data and provide additional reassurance that our models have captured relevant information.




Prior-Posterior Update Plot for Articulation Rate


Prior-Posterior Update Plot for Vowel Duration


Prior-Posterior Update Plot for Vowel Space Area


Prior-Posterior Update Plot for Location Scale Model, F1

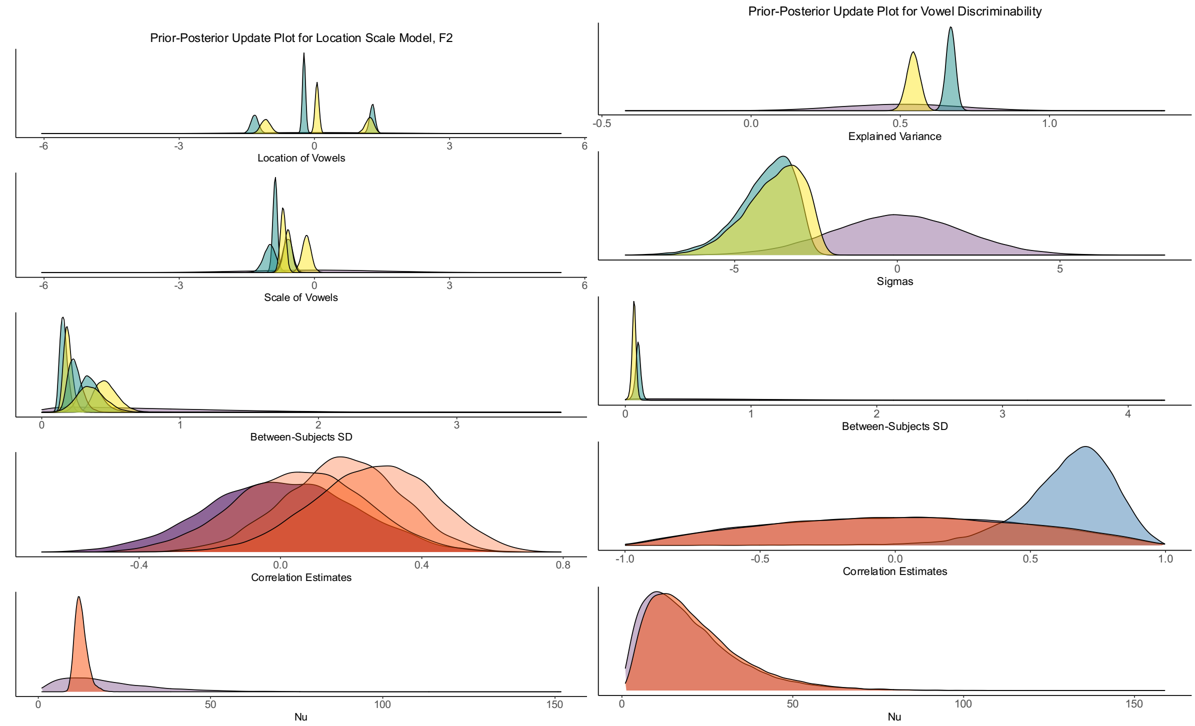


Figure S1.3: Panel of plots showing the prior-posterior update plots for each of the acoustic measures under investigation. For the intercepts, sigma parameters and between-subjects standard deviation (SD), yellow represents IDS and blue represents ADS.

S1.4: Prior Sensitivity Checks

A third quality check of the models was performed by assessing the extent to which the uncertainty of our priors affected posterior estimates. As shown in Figure S1.4, the posterior estimates (on the y-axis) exhibit stability at our choices of priors (i.e., the dashed vertical line). This pattern provides evidence that our choice of priors did not unduly affect model estimations. The only possible exception is the standard deviation of the prior for vowel space area, which as a consequence might regularise the posterior estimate for IDS to a slightly more conservative value. Given that these participant-level data necessarily exhibit a high degree of uncertainty, and the common cross-linguistic tendency for IDS to exhibit vowel space expansion (Cox et al., 2022), we consider the conservative estimate to provide a better representation of the data.

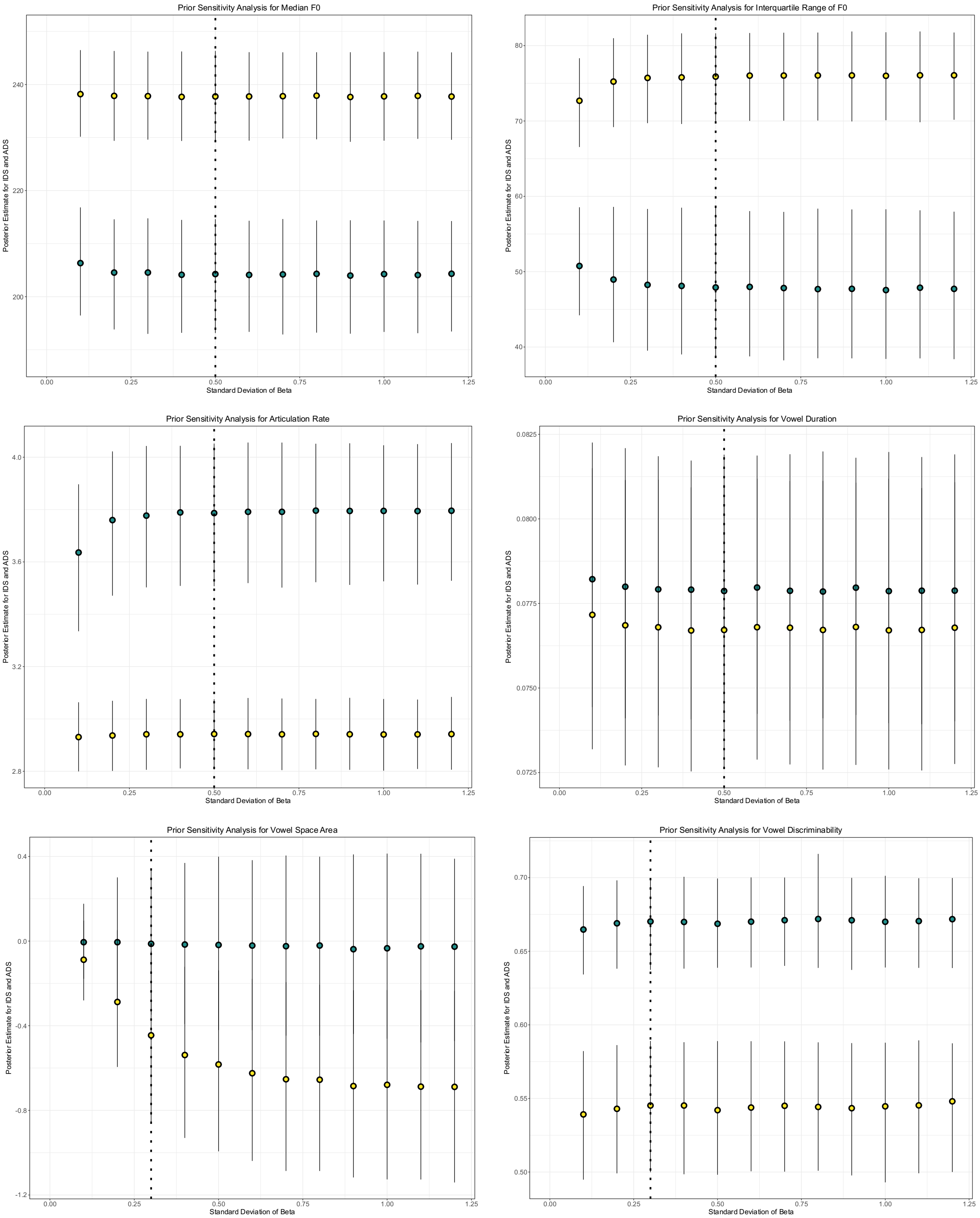


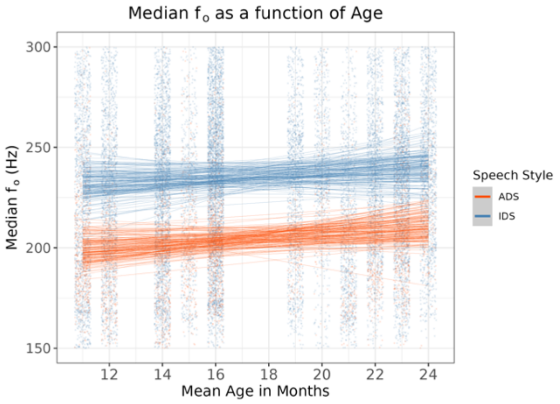
Figure S1.4: Panel of plots showing how the ADS and IDS intercept estimates change with different standard deviations for the prior. These prior sensitivity checks were carried out on a subset of the data (1/4 of the full datasets) in order to decrease the computation time (which would otherwise take a few days). The vertical dashed line indicates the standard deation of the prior chosen for the model.

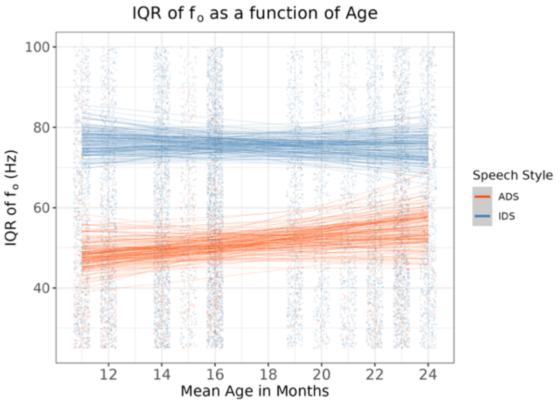
S1.5: Control Analysis of Outlier Detection and Exclusion

Acoustic Variable	Speech Style	With outlier removal (i.e. +- 2 SD)	Without outlier removal	Outlier Percentage
Median f_0	ADS	200 Hz [188; 214]	202.9 Hz [196.0; 209.6]	1.58
	IDS	227 Hz [217; 238]	235.8 Hz [228.5; 243.2]	5.30
f_0 variability	ADS	50.6 [46.3; 55.1]	48.9 Hz [44.3, 53.8]	1.78
	IDS	72.5 [69.6; 75.5]	74.9 Hz [71.1; 79.0]	4.93
Articulation Rate	ADS	3.66 [3.56; 3.75]	3.72 syll/s [3.61; 3.84]	4.68
	IDS	2.94 [2.88; 3.00]	2.96 syll/s [2.89; 3.02]	2.23

Table S1.5: Control analysis of whether outlier detection and deletion of outliers above and below two standard deviations from the mean would influence the results for the automatically extracted utterance-level acoustic measures. The estimates were obtained from running models with *Model Structure 1* as described in section S1.1 above.

S2: Age Models





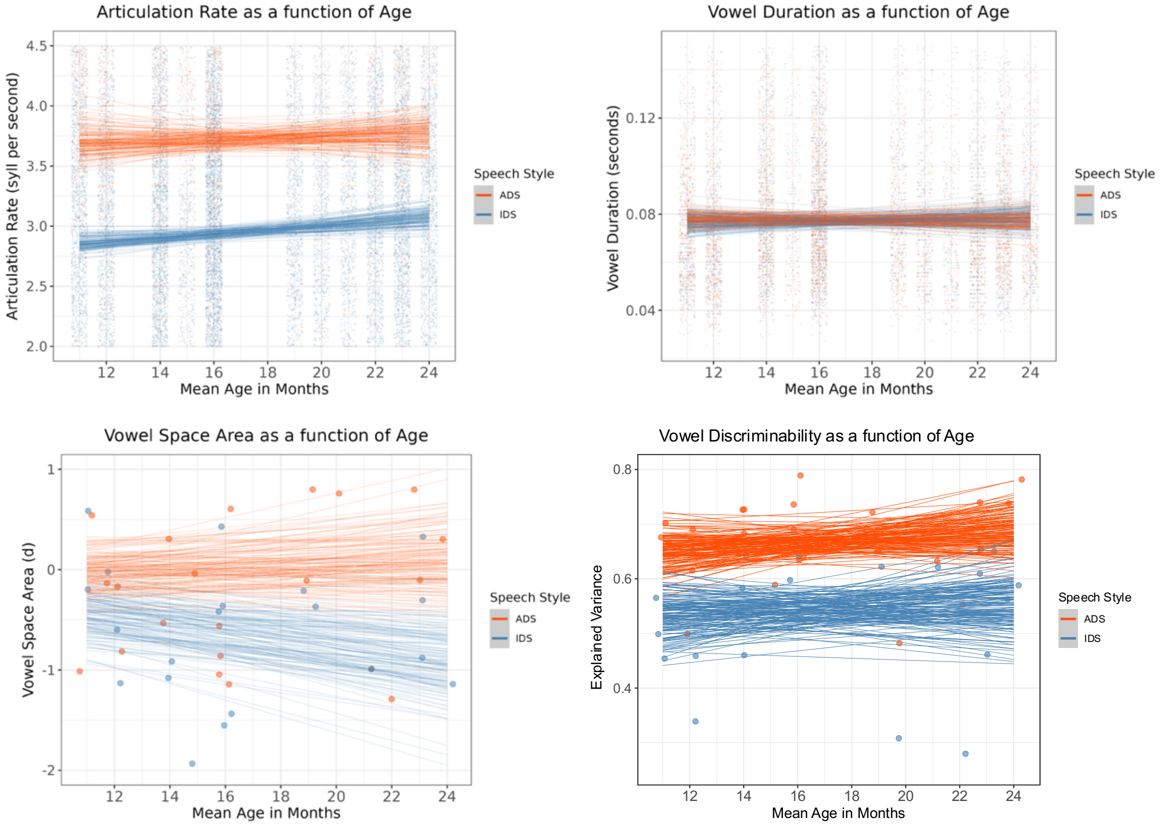


Figure S2: Panel of spaghetti plots showing how the respective acoustic measures change as a function of infant age. The plots show 150 posterior predictions from the model including an interaction between speech style and age. All of the acoustic measures except articulation rate indicate no reliable change over the course of early infancy.

S3: Vowel space area model with all ten border vowels

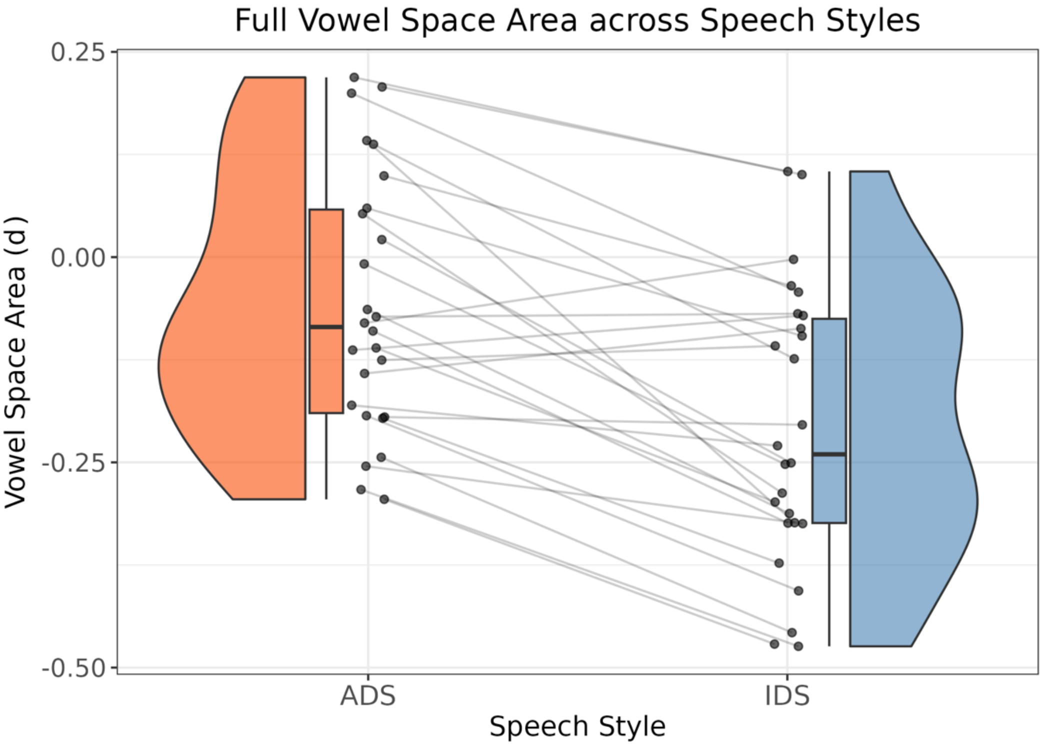


Figure S3.1: Plot of model estimates for subjects' vowel space area based on all 10 of the border vowels across the two speech styles. Each point indicates one subject. The points of each subject are connected across the two speech styles with a line.

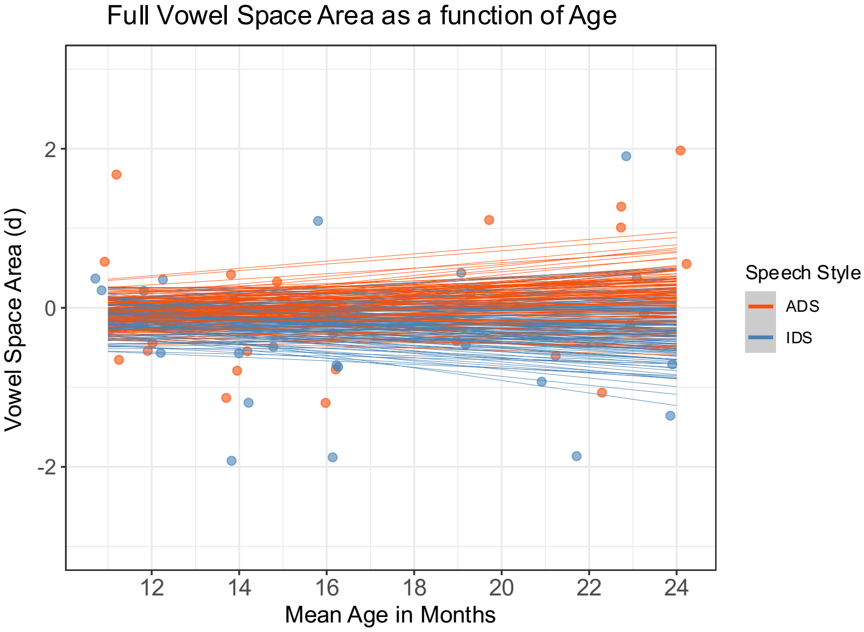


Figure S3.2: A spaghetti plots showing how vowel space area based on all ten border vowels changes as a

function of infant age. The plots show 150 posterior predictions from the model including an interaction between speech style and age.

S4: Correlation Network Plot

Figure S4.1 shows the results of a correlation network plot for the participant-level estimates for each subject. The networks indicate a strong interdependence among the acoustic measures under investigation in both caregivers' IDS and ADS. This result highlights the need to consider that each of the acoustic modifications to speech in IDS occur within a complex system of relationships among multiple levels of linguistic structure (e.g., Hawkins, 2012).

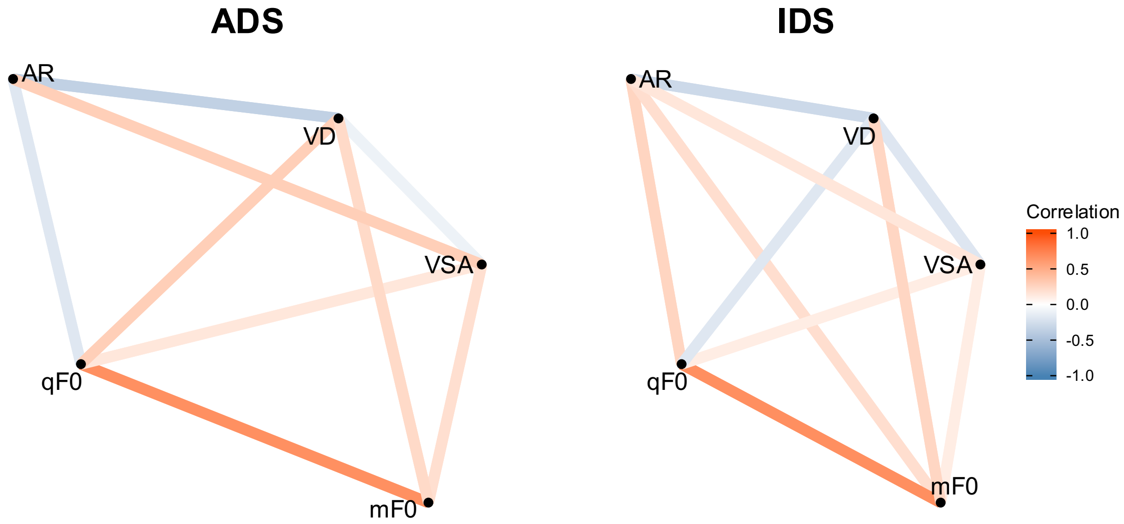


Figure S4.1: Correlation network plot for the acoustic measures in each of the two speech styles, where AR is articulation rate, VD is vowel duration, qF0 is interquartile range of *f*₀, VSA is vowel space area, and mF0 is median *f*₀. These plots indicate the extent to which each acoustic feature correlates with other features at the level of participant. Orange indicates a positive correlation, while blue indicates a negative correlation. These network plots indicate that the acoustic features exhibit a high degree of interdependence both of the speech styles under investigation.

S5: Precision Analysis for the Sample Size under Investigation

In Figure S5.1, we show the results of a simulation of how the estimates and 95% credible intervals change with a range of different sample sizes. We simulated a dataset with a standardised mean difference of 0.5, a by-subject random intercept standard deviation of 0.3, a by-subject random slope standard deviation of 0.3, a correlation between intercept and slope of 0.3 and a residual error of 1. We used a standardized mean difference of 0.5, as this approximates the smallest effect size in a recent meta-analysis of the acoustic features of IDS (Cox et al., 2022). By simulating datasets for a range of different sample sizes and building hierarchical models of each dataset (cf., S1 for more information about the models), we can assess the extent to which the width of the credible interval changes as a function of sample size. The plot in Figure S5.1 shows different estimates of effect sizes for different sample sizes, and the plot in Figure S5.2 shows model estimates from 20 simulations of datasets. The sample of 26 participants investigated in the current study appears to allow us to draw meaningful conclusions from the data for moderate effect sizes; that is, the average standard deviation in effect size is approximately 0.05, and we therefore have sufficient power to detect the expected effects of moderate sizes.

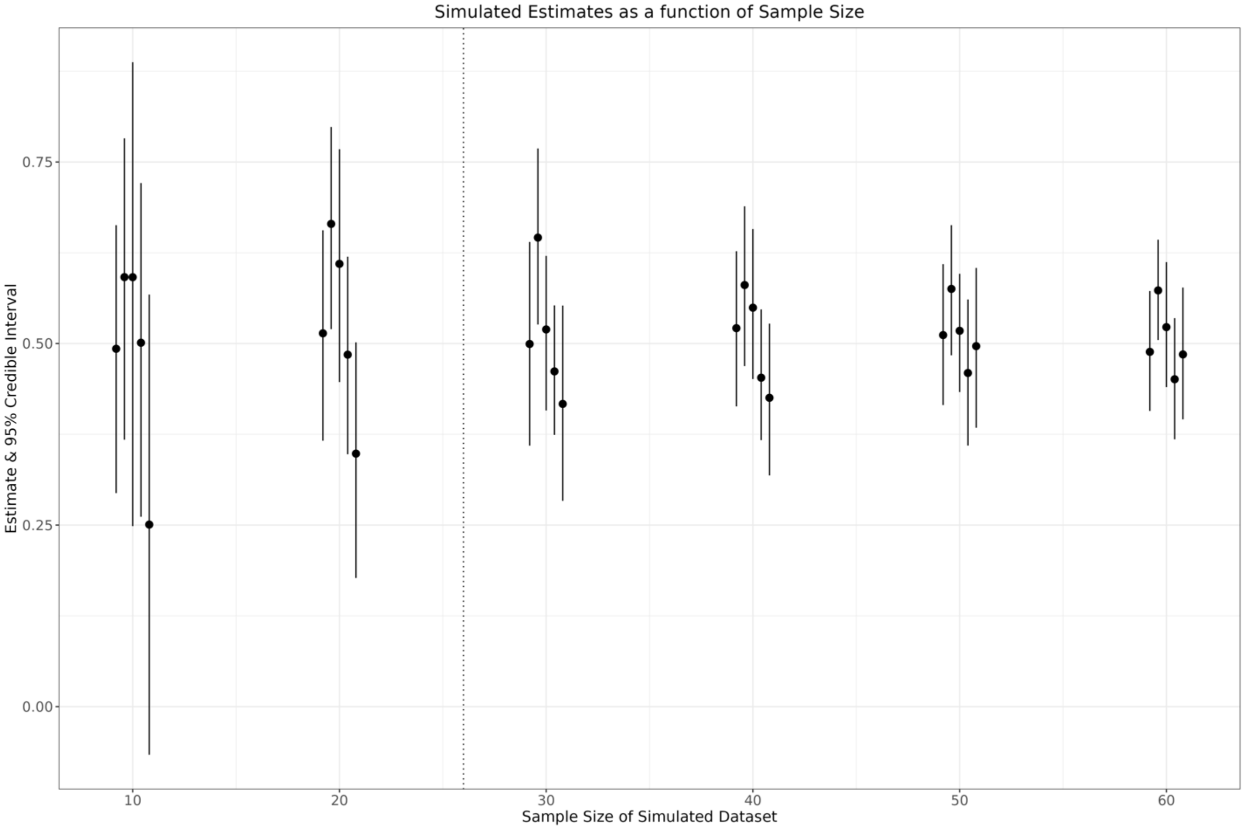


Figure S5.1: Simulation of expected values for the estimates and credible intervals of an effect size of mean = 0.5, sd = 0.2), given a simulation of the expected data structure and analysis. The vertical dashed line shows the sample size used in the current study. The analysis shows that with 26 participants, we have plenty of participants to capture the expected effect (cf., Cox et al., 2022).

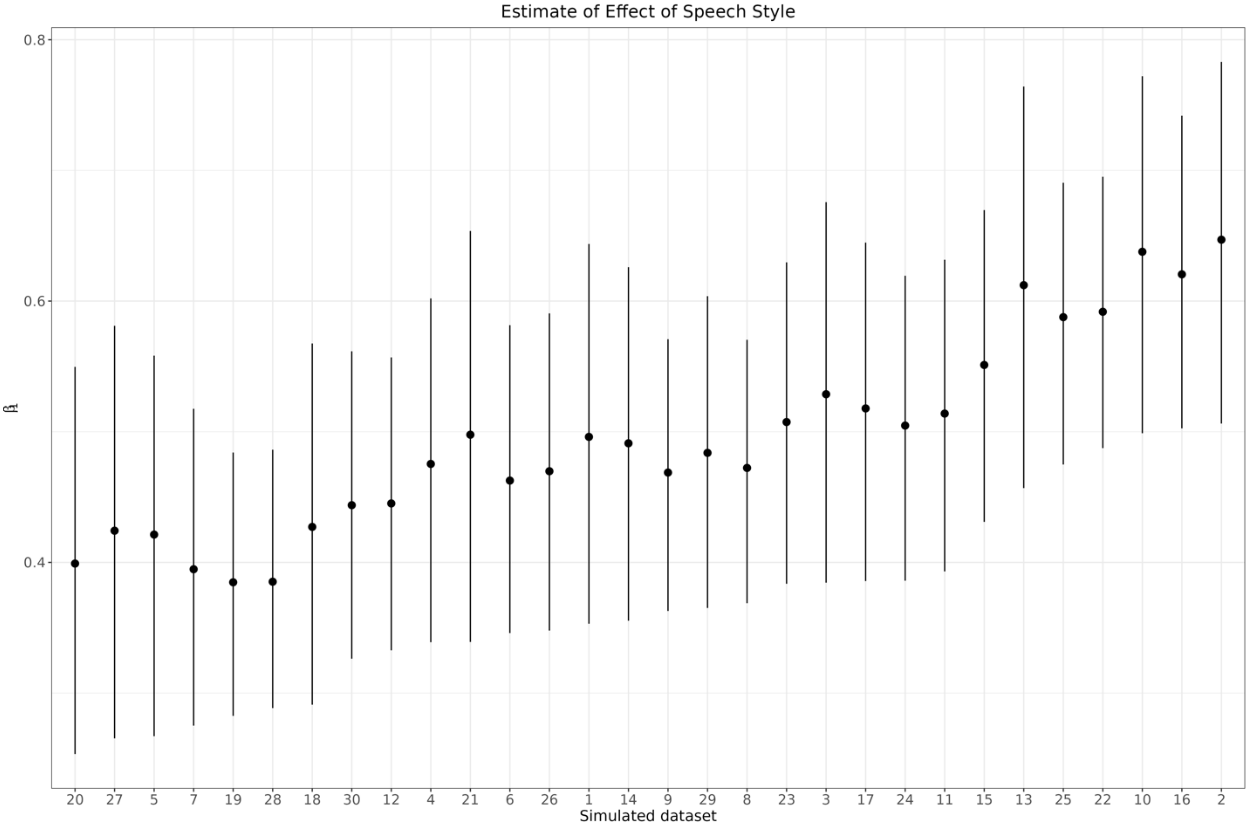


Figure S5.2: Plot of 30 simulations for the expected mean effect size of 0.5 and standard deviation of 0.2. The simulations were run with repeated measures and 26 participants. The analysis shows that with 26 participants, we have plenty of participants to capture the assumed effect based on meta-analytic evidence (cf., Cox et al., 2022).

S6: Control Analysis of Estimates across Contexts

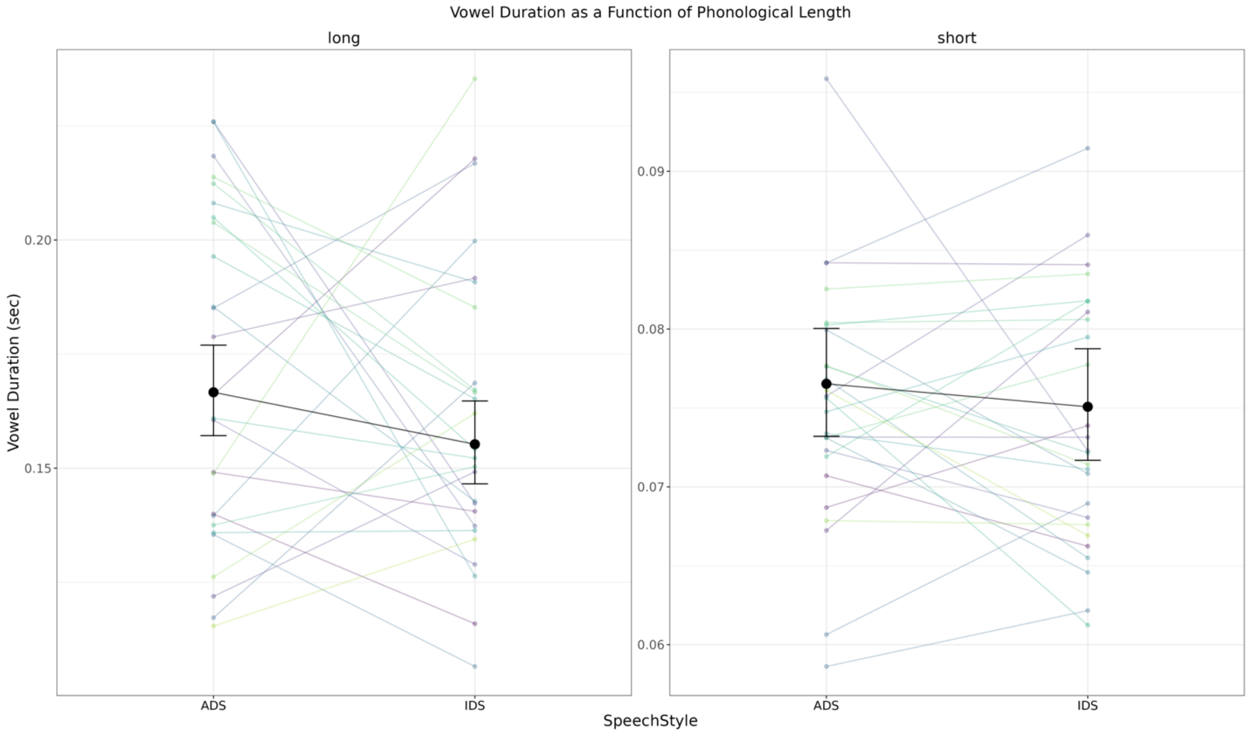


Figure S6.1: Control analysis of the extent to which vowel duration differs across the binary context of phonological length; that is, whether vowel duration is longer for phonologically long vowels vs. short vowels. The data indicate no robust differences between the speech styles in terms of short vowels; however, the phonologically long vowels in ADS are slightly longer than long vowels in IDS. Note the different scales on each of the y-axes.

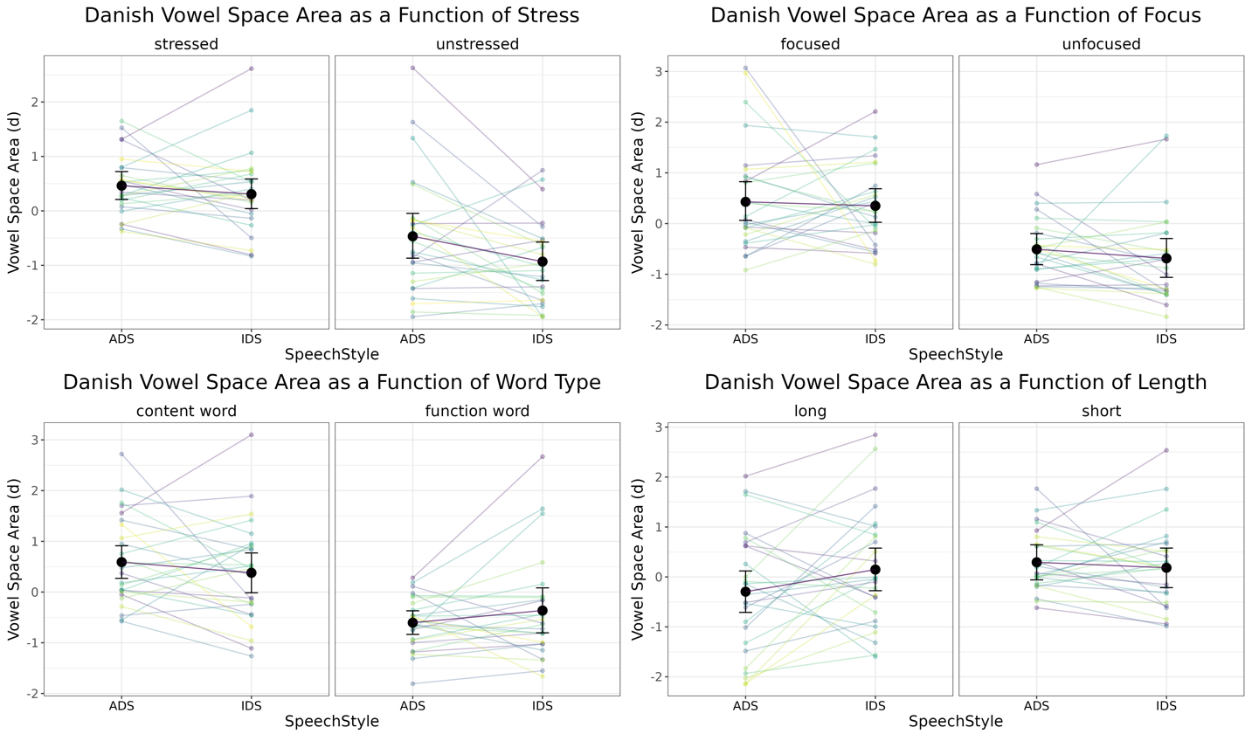


Figure S6.2: Control analysis of the extent to which vowel space area with all of the border vowels differs across different binary context; that is, whether the vowel space area is larger for content words vs. function words, for focused vs unfocused constituents, for stressed vs. unstressed syllables or long vs. short vowels. The control analysis indicated that vowel space area with all border vowels exhibited similar values across different contexts of focus (estimate = 0.08 [-0.36; 0.54], evidence ratio = 5.91, credibility score = 0.86) and non-focus (estimate = 0.18 [-0.16; 0.51], evidence ratio = 4.47, credibility score = 0.82), content words (estimate = 0.21 [-0.2; 0.61], evidence ratio = 4.01, credibility score = 0.80) and non-content words (estimate = -0.24 [-0.64; 0.16], evidence ratio = 3.4, credibility score = 0.77), as well as stressed words (estimate = 0.16 [-0.2; 0.51], evidence ratio = 5.46, credibility score = 0.85), short vowels (estimate = 0.11 [-0.37; 0.59], evidence ratio = 5.52, credibility score = 0.85).

S7: Formant estimates across contexts

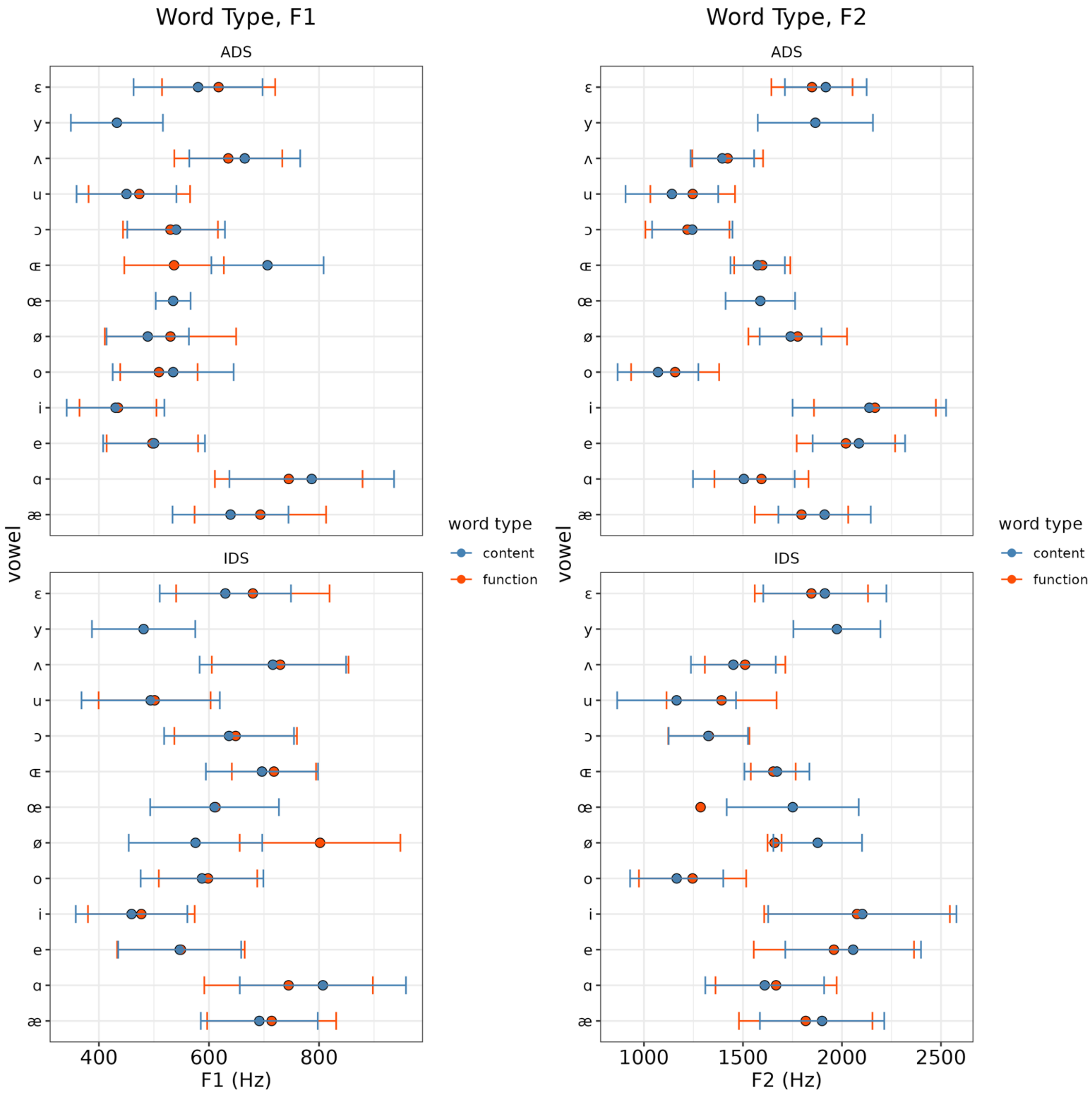


Figure S7.1: Plot of the first (left) and second (right) vowel formants for each of the vowel categories in IDS (bottom) and ADS (top) across content words (in blue) and function words (in orange). The points denote the mean F1 and F2, and the error bars show the standard deviation of the vowel formants in Hertz. Where there is no error bar, this means that there is only one vowel token within this context.

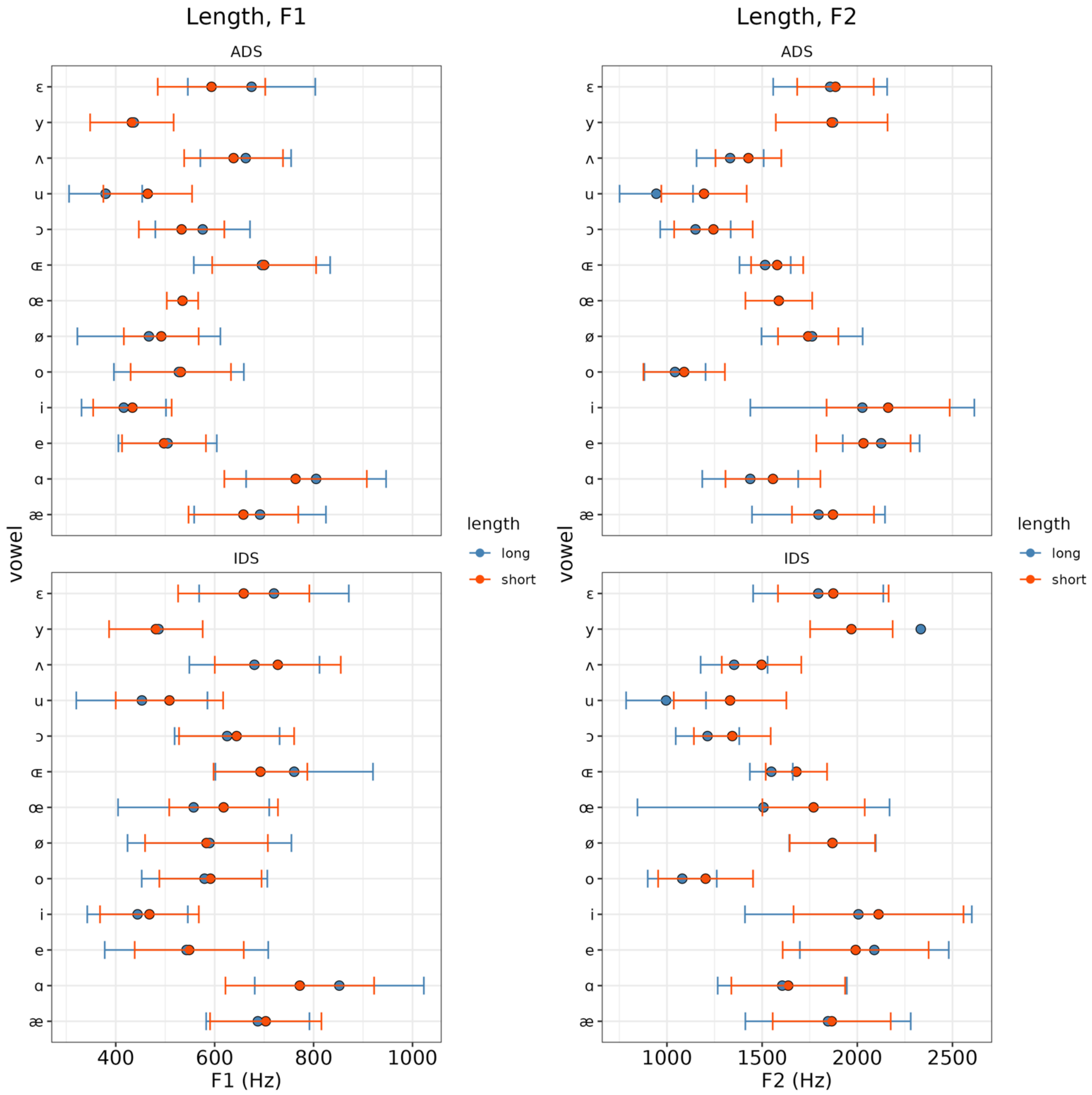


Figure S7.2: Plot of the first (left) and second (right) vowel formants for each of the vowel categories in IDS (bottom) and ADS (top) across long vowels (in blue) and short vowels (in orange). The points denote the mean F1 and F2, and the error bars show the standard deviation of the vowel formants in Hertz. Where there is no error bar, this means that there is only one vowel token within this context.

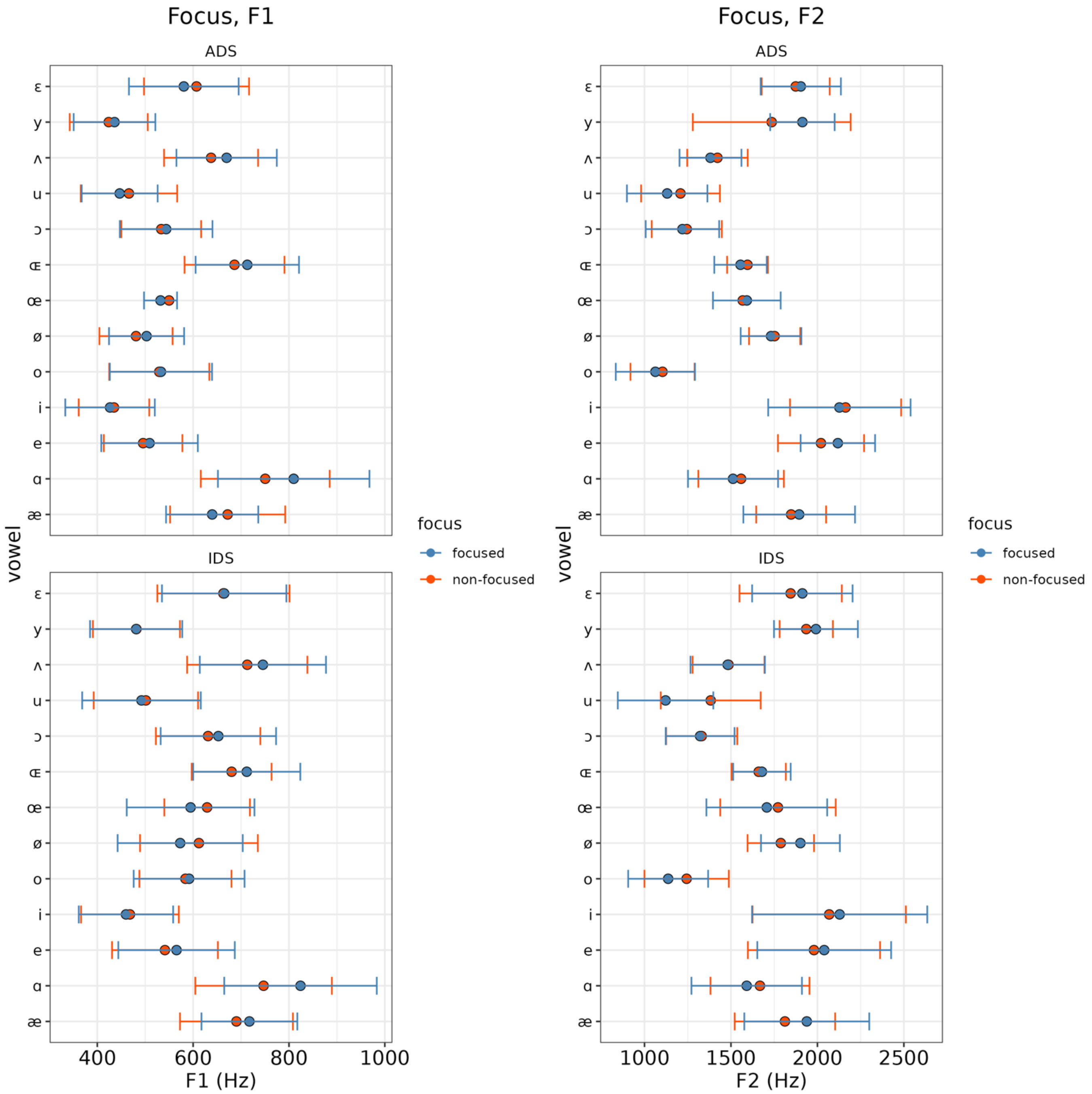


Figure S7.3: Plot of the first (left) and second (right) vowel formants for each of the vowel categories in IDS (bottom) and ADS (top) across focused words (in blue) and non-focused words (in orange). The points denote the mean F1 and F2, and the error bars show the standard deviation of the vowel formants in Hertz. Where there is no error bar, this means that there is only one vowel token within this context.

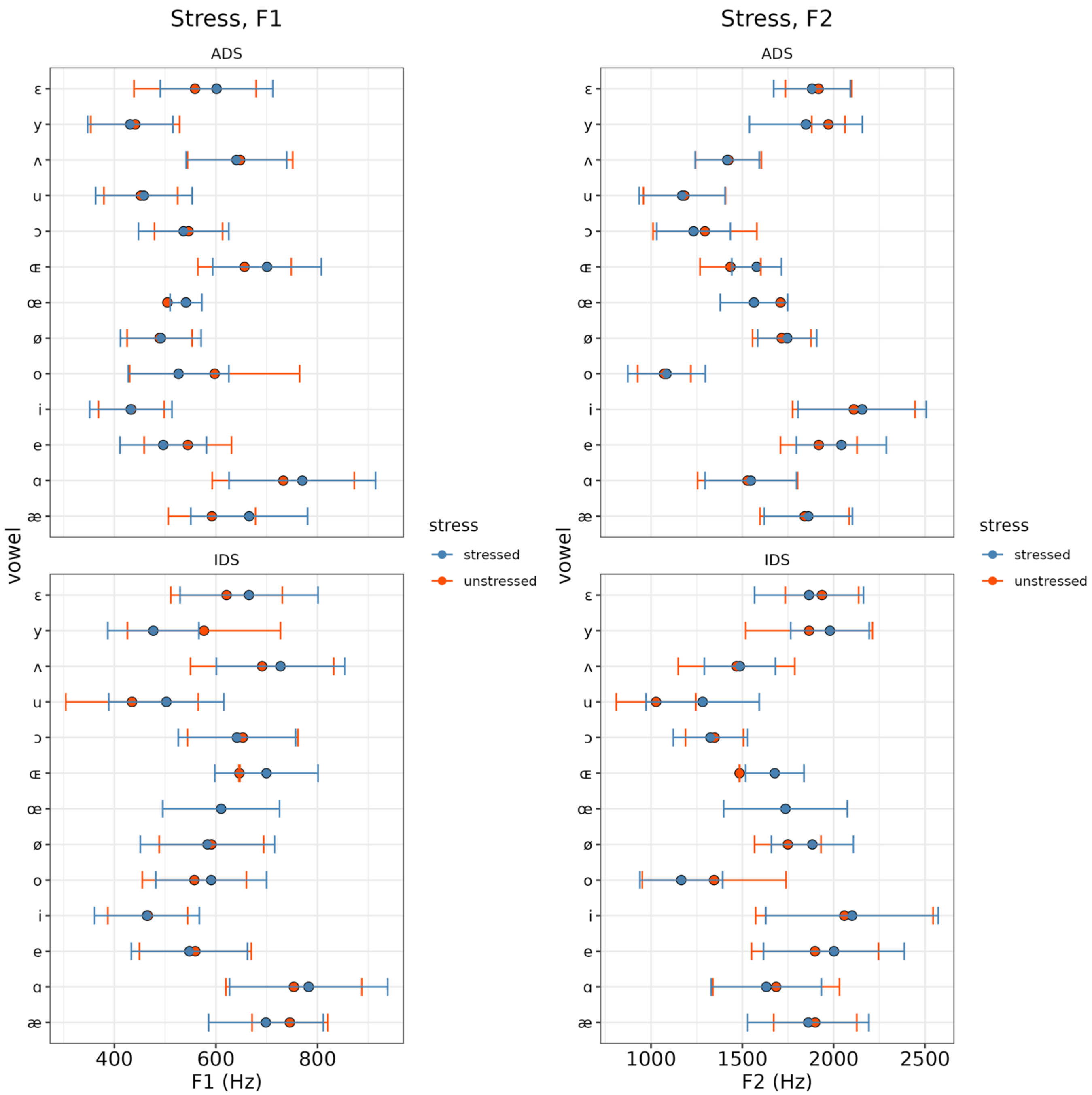


Figure S7.4: Plot of the first (left) and second (right) vowel formants for each of the vowel categories in IDS (bottom) and ADS (top) across stressed vowels (in blue) and unstressed vowels (in orange). The points denote the mean F1 and F2, and the error bars show the standard deviation of the vowel formants in Hertz. Where there is no error bar, this means that there is only one vowel token within this context.

S8: Table of formant estimates across contexts

Speech Style	Vowel	Word Type	F1	SD of F1	F2	SD of F2	N
ADS	æ	function	693.3974	119.4999	1795.735	235.8255	70
ADS	æ	content	639.1782	105.3925	1912.291	233.2941	88
ADS	ɑ	function	744.9603	134.1928	1593.435	237.5675	562
ADS	ɑ	content	786.7457	149.5976	1504.47	256.7515	672
ADS	e	function	497.1471	83.04408	2019.989	248.5293	491
ADS	e	content	500.1988	92.48543	2085.468	233.3468	181
ADS	i	function	434.6724	69.93453	2166.51	307.8303	197
ADS	i	content	430.2256	88.75982	2138.205	387.5633	195
ADS	o	function	509.0581	70.33213	1157.406	222.2465	14
ADS	o	content	534.9755	109.9166	1070.871	203.8264	78
ADS	ø	function	530.037	119.3571	1776.491	249.0785	5
ADS	ø	content	488.7915	74.81098	1740.577	156.0494	82
ADS	œ	content	534.9411	31.69412	1587.705	175.4251	6
ADS	ɶ	function	536.6075	90.45745	1597.298	141.8256	4
ADS	ɶ	content	706.3699	101.9885	1574.411	137.369	99
ADS	ɔ	function	530.0335	86.28304	1219.452	212.2321	60
ADS	ɔ	content	540.2948	88.75462	1244.061	202.9351	121
ADS	u	function	473.46	92.2535	1246.047	213.8298	30
ADS	u	content	450.1307	90.72914	1141.139	233.9394	71
ADS	ʌ	function	635.1246	98.10778	1422.803	178.8052	699
ADS	ʌ	content	665.052	100.7828	1395.776	160.4893	166
ADS	y	content	432.7002	83.64375	1865.623	290.8352	56
ADS	ɛ	function	617.5053	102.8158	1848.497	205.0493	287
ADS	ɛ	content	580.2201	117.1909	1918.284	206.4128	289
IDS	æ	function	713.9325	117.2408	1816.952	337.1669	65
IDS	æ	content	691.4303	106.273	1899.536	314.0437	83
IDS	ɑ	function	744.6941	153.1138	1667.24	306.0038	482
IDS	ɑ	content	807.0389	150.9783	1609.848	300.0666	686
IDS	e	function	549.0696	115.8999	1959.266	404.6288	441
IDS	e	content	546.8653	111.709	2056.481	342.7183	283
IDS	i	function	477.1016	96.98782	2076.081	468.871	86
IDS	i	content	459.3724	101.4069	2102.804	475.4138	219
IDS	o	function	598.3625	89.46514	1245.539	270.8651	23
IDS	o	content	587.3763	111.4958	1165.253	235.2585	156

IDS	ø	function	801.8679	145.9888	1659.669	35.41305	4
IDS	ø	content	575.5309	121.2757	1877.519	223.8219	101
IDS	œ	function	611.9918	*NA*	1286.071	*NA*	1
IDS	œ	content	610.1776	116.9715	1751.339	333.4454	30
IDS	ɶ	function	718.1258	76.49746	1652.972	113.3481	3
IDS	ɶ	content	696.3303	101.8706	1671.433	164.1	52
IDS	ɔ	function	648.5072	111.4196	1327.748	204.9877	148
IDS	ɔ	content	636.5182	117.974	1325.024	200.3356	188
IDS	u	function	501.2584	101.7373	1391.842	277.9603	68
IDS	u	content	494.0725	125.6897	1164.793	300.0936	89
IDS	ʌ	function	729.4248	124.2241	1510.802	203.2183	383
IDS	ʌ	content	716.1258	133.0774	1451.483	214.1366	307
IDS	y	content	481.2412	93.89571	1974.432	219.8036	66
IDS	ɛ	function	679.7629	139.3954	1845.601	285.8541	529
IDS	ɛ	content	629.7819	119.2437	1913.546	310.9381	251

Table S8.1: Overview of formant measures for each vowel across content words and function words in each speech style. F1 refers to the mean first formant frequency while SD of F1 refers to the standard deviation of the first formant frequency. N refers to the number of vowel tokens.

Speech Style	Vowel	Length	F1	SD of F1	F2	SD of F2	N
ADS	æ	long	691.7567	133.2161	1796.153	349.3509	24
ADS	æ	short	658.0846	110.8768	1872.205	215.5128	134
ADS	ɑ	long	805.078	141.5012	1437.562	251.9817	123
ADS	ɑ	short	763.579	144.009	1556.881	249.3134	1111
ADS	e	long	504.8426	99.47522	2125.492	201.7569	38
ADS	e	short	497.5571	84.80102	2032.359	247.6327	634
ADS	i	long	416.1725	85.36808	2026.499	588.361	28
ADS	i	short	433.7133	79.32507	2162.117	323.405	364
ADS	o	long	527.644	131.3459	1042.12	160.8452	12
ADS	o	short	531.5396	101.4228	1090.327	214.0964	80
ADS	ø	long	467.0001	144.5877	1762.507	265.653	3
ADS	ø	short	492.0248	75.62651	1741.931	158.4691	84
ADS	œ	short	534.9411	31.69412	1587.705	175.4251	6
ADS	ɶ	long	695.647	137.8419	1515.736	134.5299	6
ADS	ɶ	short	700.0327	105.0595	1578.984	136.8754	97
ADS	ɔ	long	575.7416	95.69509	1149.836	185.1288	16

ADS	ɔ	short	533.1262	86.42276	1244.249	206.2974	165
ADS	u	long	379.5614	73.89569	944.0158	192.9315	9
ADS	u	short	464.6416	89.65659	1194.632	224.2525	92
ADS	ʌ	long	662.8162	91.76101	1331.834	176.2613	93
ADS	ʌ	short	638.2238	99.86747	1427.951	172.8632	772
ADS	y	long	436.4784	*NA*	1872.514	*NA*	1
ADS	y	short	432.6315	84.41308	1865.498	293.5143	55
ADS	ɛ	long	674.5737	128.8339	1857.462	299.3388	37
ADS	ɛ	short	593.5963	108.6779	1885.3	201.0394	539
IDS	æ	long	687.2229	104.394	1846.354	434.402	18
IDS	æ	short	703.264	112.5784	1865.608	310.0075	130
IDS	ɑ	long	851.9571	170.9532	1605.816	339.1747	135
IDS	ɑ	short	772.0785	150.3002	1637.154	298.7658	1033
IDS	e	long	543.0274	165.296	2089.392	391.0836	44
IDS	e	short	548.5432	110.2559	1991.304	383.4118	680
IDS	i	long	444.0705	101.6476	2005.779	595.8415	47
IDS	i	short	468.0697	99.85521	2111.572	446.4572	258
IDS	o	long	579.4912	126.9496	1080.374	181.2975	40
IDS	o	short	591.4632	103.2928	1202.964	249.25	139
IDS	ø	long	589.5127	165.6676	1869.881	227.2589	13
IDS	ø	short	583.396	124.1256	1869.127	224.2503	92
IDS	œ	long	557.471	152.8013	1507.292	662.1801	4
IDS	œ	short	618.0532	109.888	1770.262	268.723	27
IDS	ɶ	long	760.7456	159.3105	1548.374	112.816	4
IDS	ɶ	short	692.5602	94.75505	1679.999	161.0955	51
IDS	ɔ	long	625.1437	106.1391	1213.032	166.8069	44
IDS	ɔ	short	644.3087	116.374	1343.28	201.6898	292
IDS	u	long	452.8323	132.5891	995.3194	209.7965	32
IDS	u	short	508.5391	108.5702	1331.693	295.6881	125
IDS	ʌ	long	680.369	131.5588	1352.863	175.8069	60
IDS	ʌ	short	727.6161	127.3473	1496.938	208.8887	630
IDS	y	long	486.4788	*NA*	2333.354	*NA*	1
IDS	y	short	481.1606	94.62413	1968.91	216.8517	65

Table S8.2: Overview of formant measures for each vowel across long and short vowels in each speech style. F1 refers to the mean first formant frequency while SD of F1 refers to the standard deviation of the first formant frequency. N refers to the number of vowel tokens.

Speech Style	Vowel	Focus	F1	SD of F1	F2	SD of F2	N
ADS	æ	non-focused	671.988	120.1195	1847.981	202.1024	115
ADS	æ	focused	639.6949	96.21375	1894.541	322.8539	43
ADS	ɑ	non-focused	750.4415	134.4818	1558.553	247.3272	875
ADS	ɑ	focused	809.8175	158.1074	1511.925	260.5362	359
ADS	e	non-focused	495.6042	81.96654	2020.486	249.0024	554
ADS	e	focused	509.072	100.7275	2118.092	215.2296	118
ADS	i	non-focused	434.8228	73.62092	2162.7	321.3353	279
ADS	i	focused	426.6275	93.38671	2127.073	411.7022	113
ADS	o	non-focused	529.3123	104.4661	1104.195	185.4739	47
ADS	o	focused	532.8272	106.63	1062.987	229.055	45
ADS	ø	non-focused	480.7661	76.34396	1752.417	148.0597	46
ADS	ø	focused	502.8255	78.34313	1731.672	175.4208	41
ADS	œ	non-focused	549.8123	*NA*	1567.896	*NA*	1
ADS	œ	focused	531.9669	34.48624	1591.667	195.8309	5
ADS	ɶ	non-focused	686.2881	104.2195	1595.677	118.0924	51
ADS	ɶ	focused	713.0069	107.8528	1555.315	151.619	52
ADS	ɔ	non-focused	533.4345	83.21991	1244.354	202.71	120
ADS	ɔ	focused	543.6974	96.63935	1219.279	212.4625	61
ADS	u	non-focused	466.057	100.5951	1208.159	227.7426	54
ADS	u	focused	446.7234	79.23457	1131.1	232.6994	47
ADS	ʌ	non-focused	637.4156	98.1001	1421.917	174.8914	773
ADS	ʌ	focused	669.874	104.7427	1381.481	179.0156	92
ADS	y	non-focused	423.8397	81.4209	1735.679	456.4525	15
ADS	y	focused	435.9419	85.20206	1913.164	186.37	41
ADS	ɛ	non-focused	607.029	109.6294	1874.362	196.6527	397
ADS	ɛ	focused	580.5426	114.4824	1903.805	231.9392	179
IDS	æ	non-focused	690.3457	117.8445	1812.036	290.4317	88
IDS	æ	focused	717.3985	99.9968	1938.404	361.3332	60
IDS	ɑ	non-focused	747.0091	142.4732	1667.687	286.1535	648
IDS	ɑ	focused	824.0565	159.1478	1590.969	319.4802	520
IDS	e	non-focused	541.1413	110.4096	1980.005	382.2811	513
IDS	e	focused	565.389	121.5179	2039.231	386.9472	211
IDS	i	non-focused	468.0639	101.9878	2068.176	443.2819	168
IDS	i	focused	459.8436	98.46587	2128.493	506.6662	137
IDS	o	non-focused	583.899	96.09575	1243.547	244.1094	65

IDS	o	focused	591.5754	115.6712	1136.81	231.1317	114
IDS	ø	non-focused	612.0733	122.8932	1788.101	192.1582	30
IDS	ø	focused	572.9853	130.5449	1901.668	228.0586	75
IDS	œ	non-focused	629.1234	89.29656	1771.784	333.9684	14
IDS	œ	focused	594.6819	133.2347	1707.132	349.2703	17
IDS	ɶ	non-focused	680.289	83.32309	1660.158	157.1186	25
IDS	ɶ	focused	711.8775	111.7382	1678.982	166.3741	30
IDS	ɔ	non-focused	631.2071	109.0074	1330.734	205.5216	170
IDS	ɔ	focused	652.6462	120.4206	1321.605	199.0471	166
IDS	u	non-focused	501.388	108.8559	1382.865	288.993	85
IDS	u	focused	492.2228	123.7357	1121.782	276.1947	72
IDS	ʌ	non-focused	712.8979	125.4384	1486.019	207.8055	466
IDS	ʌ	focused	745.5797	131.6657	1481.06	215.154	224
IDS	y	non-focused	481.6275	90.77743	1935.335	153.991	20

Table S8.3: Overview of formant measures for each vowel across focused and non-focused vowels in each speech style. F1 refers to the mean first formant frequency while SD of F1 refers to the standard deviation of the first formant frequency. N refers to the number of vowel tokens.

Speech Style	Vowel	Stress	F1	SD of F1	F2	SD of F2	N
ADS	æ	unstressed	591.9822	85.8633	1840.613	243.9984	5
ADS	æ	stressed	665.5267	115.0122	1861.307	241.4575	153
ADS	ɑ	unstressed	732.5975	139.8691	1528.7	273.8751	80
ADS	ɑ	stressed	770.1499	144.2819	1546.116	250.5313	1154
ADS	e	unstressed	544.6913	86.0304	1918.374	209.4574	24
ADS	e	stressed	496.2386	85.1901	2042.042	246.3674	648
ADS	i	unstressed	433.3079	64.76582	2110.326	335.2323	30
ADS	i	stressed	432.3901	80.97854	2155.919	350.9962	362
ADS	o	unstressed	597.5321	167.3034	1072.249	145.4289	6
ADS	o	stressed	526.3919	99.068	1084.862	212.0859	86
ADS	ø	unstressed	489.0613	63.91154	1715.544	159.9726	9
ADS	ø	stressed	491.4043	79.41434	1745.767	161.7519	78
ADS	œ	unstressed	504.5481	*NA*	1709.095	*NA*	1
ADS	œ	stressed	541.0197	31.28141	1563.427	184.5178	5
ADS	ɶ	unstressed	656.3955	91.81939	1434.728	166.3674	2
ADS	ɶ	stressed	700.6363	106.885	1578.083	135.7815	101
ADS	ɔ	unstressed	546.0205	67.12649	1295.175	284.4041	9
ADS	ɔ	stressed	536.4157	88.91528	1232.802	201.5054	172

ADS	u	unstressed	452.0112	72.63891	1182.857	224.61	17
ADS	u	stressed	458.082	95.03133	1170.164	234.8938	84
ADS	ʌ	unstressed	647.6942	103.5791	1423.602	180.4725	51
ADS	ʌ	stressed	640.4402	99.04505	1417.242	175.4747	814
ADS	y	unstressed	441.0181	87.36477	1970.936	90.91283	8
ADS	y	stressed	431.3139	83.88458	1848.071	309.1104	48
ADS	ɛ	unstressed	558.8238	120.237	1917.335	181.8241	33
ADS	ɛ	stressed	601.2274	110.8527	1881.456	209.9854	543
IDS	æ	unstressed	745.5833	74.41031	1899.051	227.3206	9
IDS	æ	stressed	698.4466	112.9702	1860.949	331.7245	139
IDS	ɑ	unstressed	753.4765	133.864	1685.095	346.4202	51
IDS	ɑ	stressed	782.5819	155.6974	1631.178	301.5904	1117
IDS	e	unstressed	559.4699	110.1446	1897.74	347.5387	29
IDS	e	stressed	547.7381	114.4262	2001.418	385.4509	695
IDS	i	unstressed	465.6907	78.62054	2058.54	485.7155	38
IDS	i	stressed	464.1837	103.1818	2100.497	471.8183	267
IDS	o	unstressed	557.5424	102.5439	1345.456	393.4053	10
IDS	o	stressed	590.6368	109.1194	1165.517	226.5608	169
IDS	ø	unstressed	591.1855	102.8538	1748.738	182.0418	11
IDS	ø	stressed	583.3304	132.2034	1883.319	224.4764	94
IDS	œ	stressed	610.2361	115.0059	1736.33	338.3233	31
IDS	ɶ	unstressed	646.2156	1.093867	1484.421	1.369944	2
IDS	ɶ	stressed	699.4551	101.6069	1677.445	159.821	53
IDS	ɔ	unstressed	652.8721	108.8264	1347.401	158.513	11
IDS	ɔ	stressed	641.4242	115.465	1325.507	203.5663	325
IDS	u	unstressed	434.7131	130.3369	1027.577	217.7774	12
IDS	u	stressed	502.3549	113.2878	1282.627	310.0856	145
IDS	ʌ	unstressed	690.9476	141.1262	1467.758	319.0793	70
IDS	ʌ	stressed	727.1838	126.3845	1486.289	194.2644	620
IDS	y	unstressed	576.4189	150.6721	1865.276	347.3006	3
IDS	y	stressed	476.7089	89.73003	1979.63	214.8406	63

Table S8.4: Overview of formant measures for each vowel across stressed and unstressed vowels in each speech style. F1 refers to the mean first formant frequency while SD of F1 refers to the standard deviation of the first formant frequency. N refers to the number of vowel tokens.

S9: Control Analysis with Rate-Normalized Vowel Length

We ran a control analysis of the vowel duration data by normalizing vowel durations according to the inverse of the median articulation rate of each speaker. This captures the implication that speaking faster might make a vowel of the same length seem longer (as they cover a bigger proportion of the spoken time). Accordingly, faster speech (higher articulation rate and therefore lower inverse of the articulation rate) would make vowel length relatively longer than slower speech (lower articulation rate, and therefore higher inverse), because it divides it by a smaller number. We found that normalized vowel length is shorter in IDS than in ADS, as shown in Figure S9.1, and that this applies to both short and long vowels, as shown in Figure S9.2. In other words, even taking the articulation rate into account, vowels seem shorter in IDS.

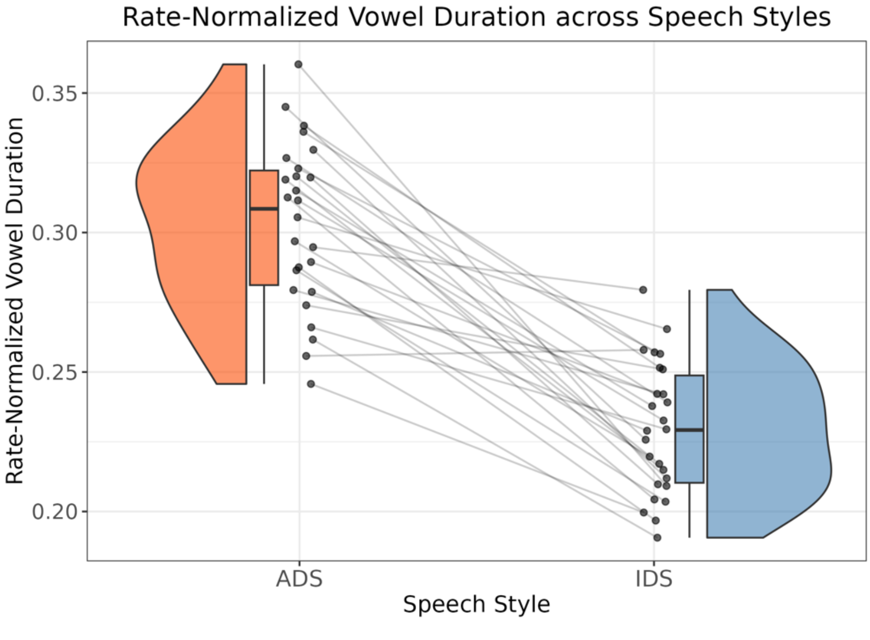


Figure S9.1: Overview of the extent to which rate-normalised vowel duration differs across the two speech styles. Each point in each speech style indicates one subject. The points for each subject are connected across the two speech styles with a line. The population-level estimates and 95% credible intervals from the model showed values for ADS as 0.30 [0.29; 0.31] and for IDS as 0.23 [0.22; 0.24].

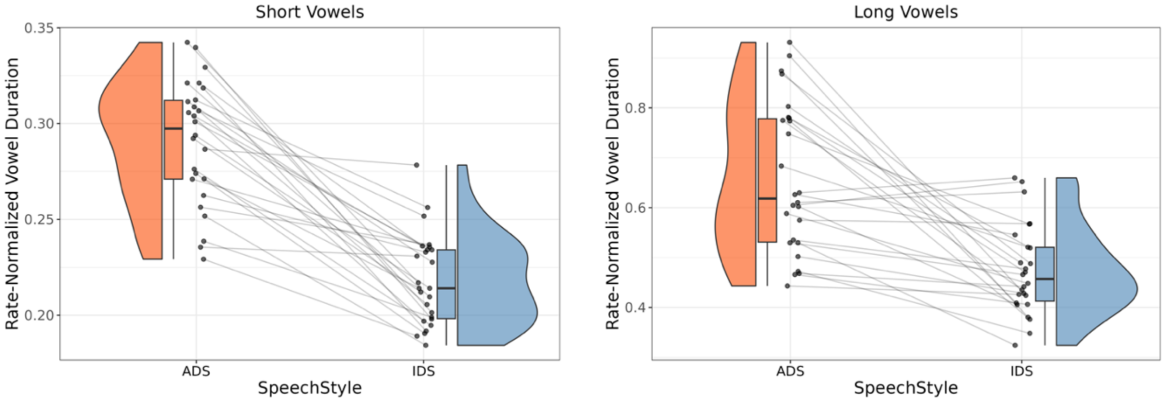


Figure S9.2: Panel of plots showing the extent to which rate-normalised vowel duration differs across the two speech styles for phonologically long and short vowels. Each point in each speech style indicates one subject. The points for each subject are connected across the two speech styles with a line. Note the different y-axis scales for the two plots. The population-level estimates and 95% credible intervals from the model showed values for short vowels in ADS as 0.29 [0.28; 0.31] and short vowels in IDS as 0.22 [0.21; 0.23] and for long vowels in ADS as 0.63 [0.58;0.68] and long vowels in IDS as 0.46 [0.42; 0.50].

S10: Reliability Analyses for ALICE and Manual Transcriptions

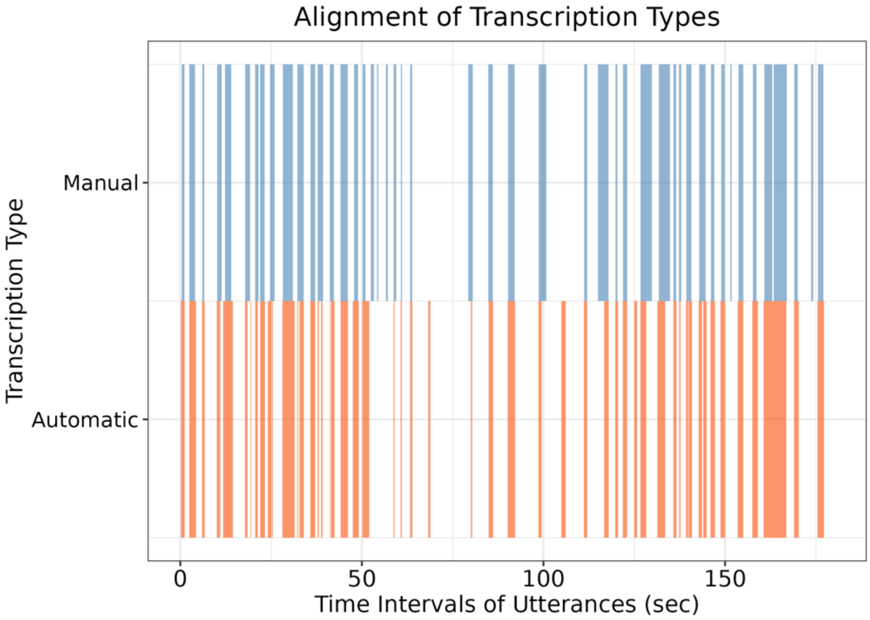


Figure S10.1: Representative plot showing alignment between the manual and automatic (i.e., ALICE) transcription in a three-minute random sample for participant AF. Each rectangle indicates a time period within which either ALICE or the transcriber has indicated that the female caregiver is speaking. The plot indicates a high degree of alignment between the two types of transcription.

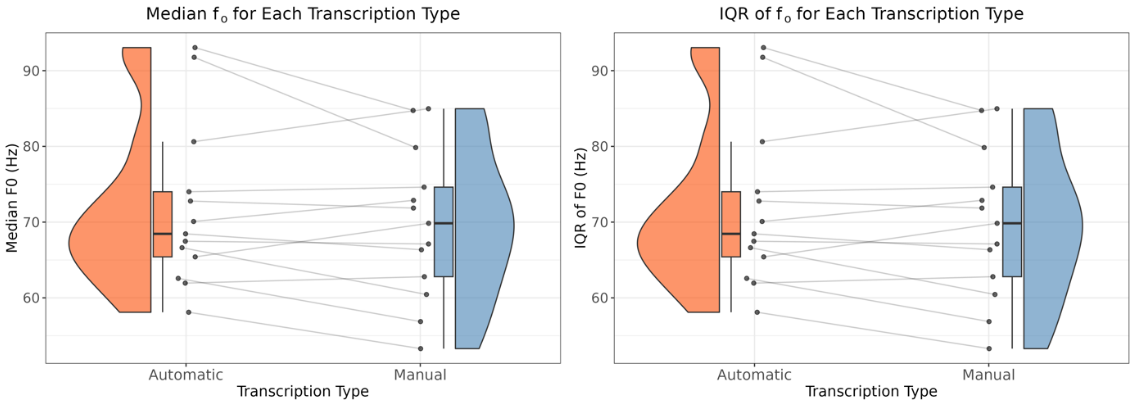
To estimate the reliability of the timecodes, we divided the audio recordings into 100ms segments and compared the manual and automated coding of the presence or absence of an adult female voice using the Kappa coefficient. Our findings show a substantial degree of agreement between the two methods (Kappa = 0.72 [0.65; 0.84]) for each subject, as shown in Table S10.1 below:

Subject	Kappa ($\varkappa$) for Start Times	Kappa ($\varkappa$) for End Times
AF	0.736	0.736
AN	0.790	0.782
DM	0.6450	0.650
IA	0.857	0.850

LK	0.716	0.718
LM	0.811	0.807
MPS	0.674	0.689
PV	0.699	0.700
SAF	0.655	0.655
SA	0.764	0.764
SC	0.811	0.809
SS	0.719	0.721

Table S10.1: Overview of Kappa coefficients for the start and end times for each of the 13 subjects for which we have both manual and automatic (i.e., ALICE) transcriptions. The results indicate a substantial degree of agreement for all of the subjects (i.e., $\varkappa > 0.60$), with some reaching perfect agreement (i.e., $\varkappa > 80$).

Apart from comparing the reliability of the actual timecodes, we also evaluated how the subtle differences in timecodes affected the extracted measures. To estimate the reliability of these prosodic measures, we extracted pitch, pitch variability, and articulation rate on two different datasets for the subset of 13 participants. The first dataset relied on the manual timecodes, while the second dataset used the automated timecodes generated with ALICE. We found that the pitch measures had an extremely high subject-level reliability (pearson coefficient > 0.99) and that the articulation rate had a high reliability (pearson coefficient 0.63). Figure S10.2 displays these findings.



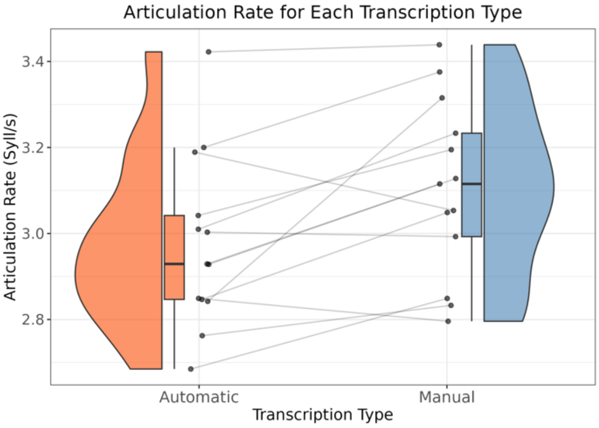


Figure S10.2: Panel of plots showing the extent to which each of the utterance-global acoustic measures differs across the two methods of transcription for the subset of 13 participants for which we have both manual and automatic (i.e., ALICE) transcriptions. Each point in transcription method indicates one subject. The points for each subject are connected across the transcription types with a line. Multivariate models with both median transcription method values as outcomes showed a correlation of 1.00 [1.00; 1.00] for median F0, a correlation of 0.99 [0.98; 1.00] for interquartile range of F0, and a correlation of 0.63 [0.17; 0.89] for articulation rate.

We also modeled the IDS versus ADS contrasts on the acoustic measures extracted from both the manual and automatic timecodes. Figure S10.3 displays the comparison of posterior model estimates, which reveal minute differences in the pitch and pitch variability estimates for IDS and only small differences in the articulation rate estimates across the two datasets. Regardless of transcription type for this subset of 13 participants, the strong differences between the ADS and IDS estimates remain highly robust. These findings are further supported by the population-level estimates reported in Table S10.2.

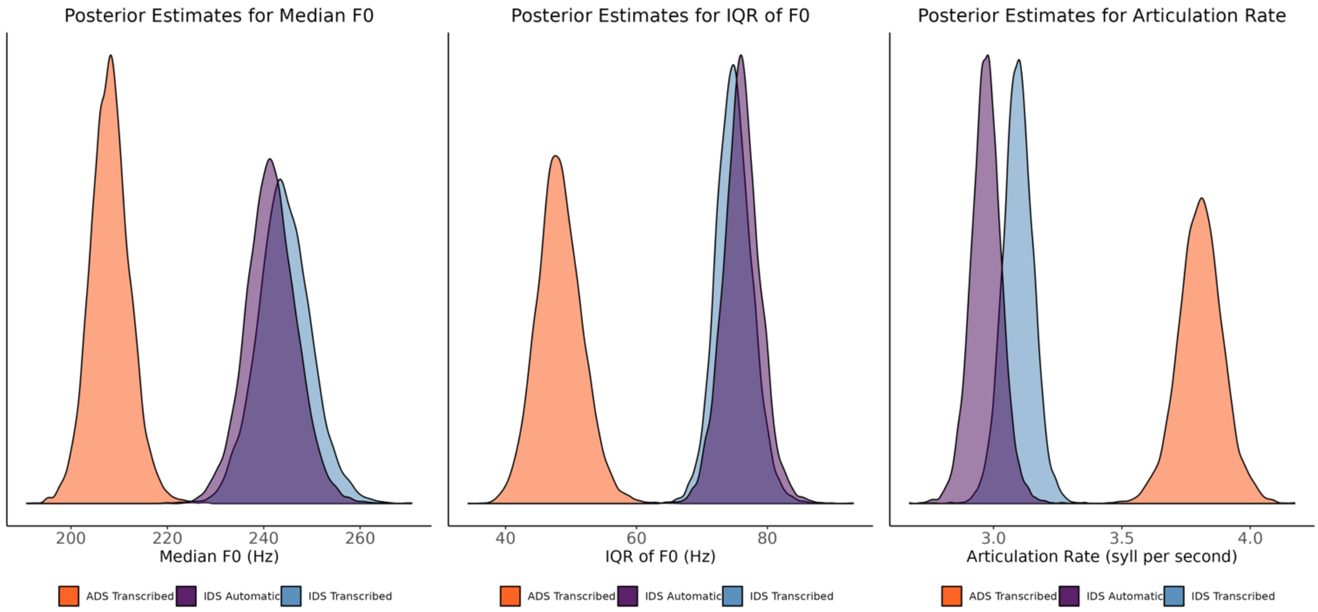


Figure S10.3: Panel of plots to compare posterior estimates for each of the utterance-global acoustic variables for the subset of 13 participants for which we had both transcribed and automatic time codes. The plots show minute differences in the pitch and pitch variability estimates for IDS and small differences in the articulation rate estimates. The main differences between ADS and IDS estimates remain highly robust regardless of the transcription type.

Style	Models	N	Median f_0	IQR of f_0	AR
IDS	**Full Dataset**	26	235.8 [228.5; 243.2]	74.9 [71.1; 79.0]	2.96 [2.89; 3.02]
	Manually Transcribed Dataset	13	244.2 [233.4; 255.9]	74.9 [69.6; 80.6]	3.09 [2.98; 3.21]
	Matched Subset of Full Dataset	13	241.5 [231.0; 252.3]	76.0 [70.4; 82.0]	2.97 [2.85; 3.09]
ADS	**Full Dataset**	26	202.7 [196.0; 209.6]	48.9 [44.5; 53.9]	3.72 [3.61; 3.84]
	Manually Transcribed Dataset	13	208.1 [200.2; 216.8]	48.1 [41.7; 55.5]	3.81 [3.63; 3.98]

Table S10.2: Comparison between model estimates for each of the datasets. The Full Dataset is the dataset used in the paper, with all 26 participants with automatically diarized IDS; the Transcribed Dataset refers to the dataset with a subset of participants for which we have manually transcribed utterances; the Matched Subset of Full Dataset refers to the automatically diarized data, but only for the subset of participants for which we have the manually transcribed data. N refers to the number of participants in each of the datasets. AR refers to Articulation Rate. IQR of f_0 refers to interquartile range of fundamental frequency (f_0).

Study III

A systematic review and Bayesian meta-analysis of the development of turn taking in adult-child vocal interactions

Vivian Nguyen* [1, 2], Otto Versyp* [1, 2], Christopher Cox* [2, 3, 4], Riccardo Fusaroli[2, 3, 5]

* Joint first-authorship

[1] Faculty of Psychology & Educational Sciences, University of Ghent

[2] School of Communication and Culture, Aarhus University,
Jens Chr. Skous Vej 2, Building 1485, 8200 Aarhus, Denmark

[3] Interacting Minds Centre, Aarhus University,
Jens Chr. Skous Vej 4, Building 1483, 8200 Aarhus, Denmark

[4] Department of Language and Linguistic Science, University of York

[5] Linguistic Data Consortium, University of Pennsylvania, PA, USA

Nguyen, V., Versyp, O., Cox, C., & Fusaroli, R. (2022). A systematic review and Bayesian meta-analysis of the development of turn taking in adult–child vocal interactions. *Child Development*, *93*(4), 1181-1200.

Acknowledgements

VN and OV were funded by an Erasmus Internship program. RF is funded by the NIH-NIDCD R01 project “Early Predictors to School Age Language” and seed funding from the Interacting Minds Center. RF has been a paid consultant for F. Hoffman La Roche on related but non-overlapping topics. The data necessary to reproduce the analyses presented here are publicly accessible at the following URL: https://osf.io/wkceb/. The analytic code necessary to reproduce the analyses presented in this paper is publicly accessible at: https://osf.io/wkceb/. All of the computation done for this project was performed on the UCloud interactive HPC system, which is managed by the eScience Center at the University of Southern Denmark. The materials (i.e., videos and audio files) necessary to attempt to replicate the findings presented here are not publicly accessible. The analyses presented here were not preregistered. The manuscript relies on publicly available data (published articles) and has been deemed exempt from the need of ethical approval by the local ethical committee.

Correspondence: chris.mm.gmail.com (Chris Cox)

University of York
York Graduate Research School
Research Degree Thesis Statement of Authorship

Note that where a paper has multiple authors, the statement of authorship can focus on the key contributing/corresponding authors.

Candidate name	Christopher Martin Mikkelsen Cox
Department	Department of Linguistics & Cognitive Science
Thesis title	How Social Contingency Shapes Early Child-Caregiver Interactions

Title of the work (paper/chapter)	A systematic review and Bayesian meta-analysis of the development of turn taking in adult-child vocal interactions	
Publication status	**Published**	X
	Accepted for publication	
	Submitted for publication	
	Unpublished and unsubmitted	
Citation details (if applicable)	Nguyen, V., Versyp, O., Cox, C., & Fusaroli, R. (2022). A systematic review and Bayesian meta-analysis of the development of turn taking in adult–child vocal interactions. *Child Development*, *93*(4), 1181-1200.	

Description of the candidate's contribution to the work*	Conceptualization, Methodology, Software, Validation, Formal analysis, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision, Project administration
Approximate percentage contribution of the candidate to the work	34 %
Signature of the candidate	
Date (DD/MM/YY)	24-11-23

<u>Co-author contributions</u>
By signing this Statement of Authorship, each co-author agrees that:
(i) the candidate has accurately represented their contribution to the work;
(ii) if required, permission is granted for the candidate to include the work in their thesis (note that this is separate from copyright considerations).

Name of co-author	Otto Versyp
Contact details of co-author	otto.versyp@kuleuven.be
Description of the co-author's contribution to the work*	Conceptualization, Methodology, Software, Validation, Formal analysis, Resources, Data Curation, Writing - Original Draft, Visualization, Project administration
Approximate percentage contribution of the co-author to the work	30 %
Signature of the co-author	OTTO
Date (DD/MM/YY)	24-11-23

Name of co-author	Vivian Nguyen
Contact details of co-author	vivian.nguyen@be.ey.com
Description of the co-author's contribution to the work*	Conceptualization, Methodology, Software, Validation, Formal analysis, Resources, Data Curation, Writing - Original Draft, Visualization, Project administration
Approximate percentage contribution of the co-author to the work	30%
Signature of the co-author	Vivian Nguyen Digitally signed by Vivian Nguyen DN: cn=Vivian Nguyen, email=Vivian.Nguyen@be.ey.com Date: 2023.11.24 11:59:19 +01'00'
Date (DD/MM/YY)	24-11-23

Name of co-author	Riccardo Fusaroli
Contact details of co-author	fusaroli@cc.au.dk
Description of the co-author's contribution to the work*	Conceptualization, Methodology, Software, Validation, Formal analysis, Resources, Data Curation, Writing - Original Draft, Visualization, Project administration
Approximate percentage contribution of the co-author to the work	6%
Signature of the co-author	
Date (DD/MM/YY)	24-11-23

A systematic review and Bayesian meta-analysis of the development of turn taking in adult-child vocal interactions

Abstract

Fluent conversation requires temporal organization between conversational exchanges. By performing a systematic review and Bayesian multi-level meta-analysis, we map the trajectory of infants' turn-taking abilities over the course of early development. We synthesize the evidence from 26 studies reporting response latencies in infant-adult dyadic interactions. Infants' response latencies exhibit only weak evidence of an increase until 20 months of age. Infants' response latencies are also related to those of their adult conversational partners. These results highlight the dynamic reciprocity involved in the temporal organization of turn-taking. Based on these results, we provide recommendations for future avenues of enquiry; studies should analyze how turn-by-turn exchanges develop on a longitudinal timescale, with rich assessment of infants' linguistic and social development.

1. Introduction

Conversations involve structured and coordinated interactions between two or more interlocutors who *take turns* in their roles as speakers and listeners (Fusaroli, Rączaszek-Leonardi, et al., 2014). The fluency of these conversational exchanges depends on their dynamic temporal structure where each speaker has to anticipate the end of the other speaker's turn (Sacks et al., 1978). To keep the conversation running smoothly, interlocutors tend to optimize the response latency between their turns by avoiding overlaps in speech and minimizing the pauses between their own utterances and those of their partner. Turn-taking is schematically represented in Figure 1. Note that this turn-taking system represents a coarse-grained, and yet useful description of conversations. At a higher level of detail, conversations involve frequent backchanneling (nodding and supporting the speaker with sounds like "hmm") and other subtle ways of communicating beyond a strict speaker / listener distinction (Cummins, 2013).

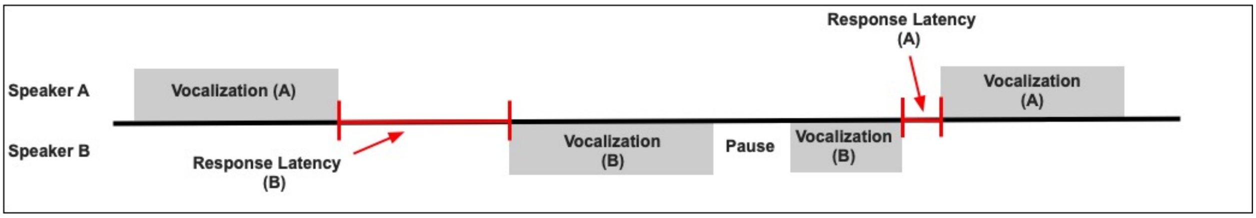


Figure 1 - Visualization of the turn-taking process. The x-axis represents time. Pause refers to within-speaker pauses. Response latency – the focus of the current study - is the period of time which the corresponding speaker requires to answer an utterance by another speaker (i.e., duration of the turn transition to another speaker). Note that response latency could be negative, if the adjacent turns overlap.

This tendency for interlocutors to optimize the latency between their utterances applies in a wide variety of human languages. Stivers et al. (2009), for example, investigate ten typologically distinct languages and show that – in adult-adult conversations - average response latency during turn transitions varies between 0 and 500 milliseconds. The cross-linguistic variation in this corpus is non-trivial, with latencies averaging 7ms in Japanese against the 470ms of Danish. However, the range of acceptable latencies is often overlapping, and conversational turn-taking exhibits similar qualitative structures across languages. Notably, these latencies are most often too short to allow for speech planning to happen after one's interlocutor has finished speaking (Griffin & Bock, 2000; Indefrey, 2011; Indefrey & Levelt, 2004). This indicates that speakers are well-tuned to transitions between turns, to the point that they can anticipate them and start planning before the end of their interlocutor's utterance. Turn-taking has thus been argued to provide the cognitive scaffolding to facilitate social interactions (Stivers et al., 2009).

Further, a growing number of studies is finding signs of communicative turn-taking within different species, often sharing the optimization principles highlighted in human conversation (Pika et al., 2018). For example, nightingales are capable of precisely timing the onset of their song, starting maximally 1 second after the end of a neighbor's song (Todt & Hultsch, 1999). Other studies have identified vocal turn-taking in meerkats (Demartsev et al., 2018) and some primate species (e.g. *Lepilemur edwardsi*, Méndez-Cárdenas & Zimmermann, 2009). These findings have been used to suggest that turn-taking is one of the basic foundations of social communication across a variety of different species (de Reus et al., 2021).

Perhaps because of its foundational basis for social interaction, turn-taking plays an important functional role in human language development. For instance, pre-verbal vocalizations more closely resembling language have been shown to elicit prompter caregiver responses; and in turn, faster caregiver response latencies are shown to relate to more language-like successive vocalizations (the "social feedback loop" model, Lopez et al., 2020; Warlaumont et al., 2010, 2014). Further, at least in adult-adult conversations, smooth and regular turn-taking may play a role in facilitating coordination. For instance, synchrony and shared tempo in joint activities have been shown to create feelings of shared motivation and to lessen the cognitive demands required by the interaction, since it reduces the need to continuously adjust to changing conversational rhythms and to figure out when to talk (Cross et al., 2020; Dideriksen et al., 2020; Fusaroli et al., 2016; Fusaroli, Konvalinka, et al., 2014; Fusaroli, Rączaszek-Leonardi, et al., 2014; Fusaroli & Tylén, 2016; Konvalinka et al., 2011). While this turn-taking process appears to be effortless, it relies on a set of complex skills (e.g., simultaneous language processing and production planning) and coordination between the different interlocutors. The order in which speakers organize their conversations, the length of their turns, and the pauses between those turns all require real-time coordination and cannot be defined in advance. Furthermore, interlocutors need to process the other person's utterance, plan their own response in accordance while simultaneously anticipating when the end of a turn is coming up (Pickering & Garrod, 2021; Riest et al., 2015; Ruiter et al., 2006).

Despite the skills required, infants seem capable of smooth turn-taking at an early point in development (Bateson, 1975; Gratier et al., 2015). However, there is no current overview on how early this might happen and more generally of the developmental trajectory of turn-taking. Reports of response time latencies in parent-child interactions are scattered through the literature, often reported as secondary descriptions of the data. Understanding how infants' turn-taking abilities develop over time would grant us a more complete understanding of language development and the skills underlying it, as well as provide insights as to the cognitive mechanisms necessary for turn-taking.

To provide better foundations for investigations of turn-taking, this paper systematically reviews and meta-analyzes the evidence related to the development of infants' turn-taking abilities. This paper will specifically focus on infant-adult dyadic interactions (i.e., no infant-infant interactions), since adult interlocutors provide more homogeneously skilled interlocutors than other infants and therefore might facilitate comparisons across infant ages. In the following sections, we present models of turn-taking and thereafter summarize the literature on how turn-taking develops over the course of early infant development. We then discuss the factors that likely moderate turn-taking in order to motivate the specific hypotheses to be assessed in the meta-analysis.

1.1 Models of turn-taking

Turn-taking is construed as being organized around two key principles: i) to ensure minimal overlap between speakers' utterances, and ii) to optimize the transition times between speakers' conversational turns (Stivers et al., 2009). Several mechanisms have been suggested to make this possible.

In one of the earliest in-depth analyses of conversational turn-taking, Sacks et al. (1978) developed a descriptive model of turn allocation, that is, of the selection of the next speaker. As an utterance ends, three things can happen: i) the current speaker selects who should speak next; ii) a current listener self-selects and starts speaking; or iii) the current speaker self-selects and resumes speaking. Independently of which of these options is taken, Stivers et al. (2009) have argued for smooth transitions to happen universally across humans by showing modal response time between 0 and 500 milliseconds (ms) across 10 distinct languages. This is argued to indicate a universal mechanism aimed at both avoiding speech overlap and minimizing periods of silence between turns (Levinson & Torreira, 2015). While the general pattern is uncontroversial, it has been argued that a modal range of 0-500 ms also indicates potential non-trivial cross-linguistic differences. For instance, Danish presents response latencies that are 100 ms slower than average, and acceptable response latencies up to 1s, besides other differences (Heinemann, 2010). This has been associated with compensation mechanisms: Danish presents unusually high levels of phonetic reduction making the speech signal potentially more difficult to parse than many other languages (Trecca et al., 2021). However, these cross-linguistic differences have not been systematically investigated further.

Levinson (2006, 2016) takes an explicitly cognitive direction to explain smooth turn-taking, with the Interaction Engine Hypothesis. This model focuses on the abilities necessary for turn-taking, stressing that such interactional abilities are also foundational for language development and use (Levinson, 1995; Warlaumont et al., 2014). These cognitive abilities include the crucial components of the engine: attributing intentions to others (mapping their behavior onto inferred goals); simulating an interlocutor attributing intentions to us (anticipating what an interlocutor thinks we think) (Waade et al., 2020); and producing behaviors aimed at having interlocutors recognize the speaker's intentions (or Gricean intentions). Crucially, this model provides specific predictions, namely that smooth fast turn-taking would be present in pre-verbal infants, but that it would slow down with the acquisition of language (which the author puts at around 12 months of age), as developing linguistic skills would need to be integrated to the existing interactional ones. One study has tested and partially supported this prediction, highlighting a slow-down in turn-taking at around 9 months, that is, before the actual use of words, (Hilbrink et al., 2015).

If the Interaction Engine Hypothesis focuses on the individual cognitive abilities needed to make interactions work, Wilson and Wilson (2005) focus more explicitly on the interpersonal dynamics of turn-taking, with the Coupled Oscillators model. The model states that interlocutors have internal oscillators (e.g., rhythmic patterns of brain activity), which can become progressively coupled across individuals through coordination of response latencies and syllable durations. Turn-taking can be smooth because it is based on the rhythm of these oscillators. For instance, we can imagine interlocutors establishing a shared tempo of 500 milliseconds per oscillation, so that response latencies would only happen at multiples of 500ms and display the same durations: 0 seconds, 500 milliseconds, 1 second, etc. This not only simplifies the complexity of when to start speaking (as only specific timings are preferred), it crucially minimizes the odds of speaking simultaneously due to the fact that the readiness cycles of listeners are counter-phased to those of speakers. This model is supported in a few animal turn-taking studies (Takahashi et al., 2013). However, evidence for the model in human conversations is controversial. In a case study of one conversational pair, O'Dell et al. (2012) failed to detect relations between syllable rate and turn-taking and suggested that slower speech-related rhythms, such as phrasal stress rhythm, might be involved. Alternatively, McFarland (2001) and Rochet-Capellan and Fuchs (2014) explore the possible role of respiration and indicate the presence of at least two respiration regimes that get interpersonally entrained: i) within turns, and ii) during turn transitions. Despite the uncertainty as to the specific details of the mechanisms, growing evidence shows that interlocutors engaged in conversation adjust the temporal aspects of their speech to each other, e.g., the duration of intra-turn pauses (Cappella, 1981); the duration of speech and pause sequences (Fusaroli et al., 2016; Fusaroli & Tylén, 2016); and the tempo of their speech and syllables (Manson et al., 2013); as well as the temporal aspects of non-speech behaviors (Curioni et al., 2019; Duran & Fusaroli, 2017; Fusaroli et al., 2016; Pezzulo et al., 2019).

While there is currently no consensus on the specific mechanisms underlying turn-taking, we should expect cognitive development, speech development and interpersonal dynamics to play a role in both the developmental trajectory and deployment of turn-taking. We now turn our gaze to current evidence as to how turn-taking develops from early infancy.

1.2. The developmental trajectory of turn-taking and its moderators

The earliest described examples of vocal turn-taking in infants are called proto-conversations. Proto-conversation can be described as a structured, nonlinguistic vocal exchange between infants and their peers or adults (Hsu et al., 2001; Hsu & Fogel, 2003; Kaye, 1977), reported already before 3 months of age (Bateson, 1975; Dominguez et al., 2016; Gratier et al., 2015). These proto-conversations are described as shared meaningful activities, and seem to adhere to the temporal structure of adult conversations (Bateson, 1975; Snow, 1977). While proto-conversations have been characterized as analogous in response latencies to adult conversations, older infants already from 9 months of age, and up to 4 years of age have shown to be slower than adults at planning, initiating and completing their responses (Casillas et al., 2016; Hilbrink et al., 2015; Levinson, 2013).

The mechanisms underlying the ability to sustain turn-taking and thus the reported patterns are still under-investigated. There is evidence that infants are sensitive to turn-taking timing in social interactions at least from 9 months of age (Striano et al., 2006; Thorgrimsson et al., 2014), and that at least from the age of one year infants are capable of predicting upcoming turn transitions (Casillas & Frank, 2013). More direct tests that manipulate the degree of predictability of questions directed to older (3;0-6;0) children reinforce this picture: the more predictable the question, the more children shorten their response latency (Lindsay et al., 2019). This finding admits an important role of the contexts and activities at play in scaffolding children's turn-taking abilities. However, very few longitudinal studies of turn-taking have been conducted, and none investigating the joint development of response latencies and their underlying mechanisms (Bateson, 1975; Ginsburg & Kilbourne, 1988; Hilbrink et al., 2015; Snow, 1977).

1.3. Potential factors underlying turn-taking skills

In order to develop a more comprehensive understanding of the development of turn-taking, we reviewed the literature to identify possible factors to be accounted for. This review allowed us to investigate the following moderators in the meta-analysis: i) child age, ii) adult response latencies, iii) familiarity of interlocutors and iv) interactional settings, v) type of activity, vi) language being spoken, and vii) whether the child had a developmental disorder or other special conditions. We discuss each in turn.

Age. The Interaction Engine Hypothesis (Levinson, 2006) predicts turn-taking to be fast at an early age, but to slow down after 12 months. Tomasello (1999) on the contrary talks about a 'nine-month revolution', the age at which infants start recognizing the importance of intentional communication, which might also cause a slowdown (Hilbrink et al., 2015). This slowdown can additionally be attributed to the fact that infants at this age are rapidly acquiring more complex language, which they attempt to incorporate in their turns. The empirical results from longitudinal studies are, however, less straightforward. For example, testing the

Interaction Engine Hypothesis, Hilbrink et al. (2015) found that up until 9 months preverbal infants' response latencies are similar to those of adults, but slow down thereafter, until infants are 18 months of age. Thus, the authors amend the original prediction of a slowdown from 12 months, and place it at 9 month, referring to Tomasello's 9-month revolution (Tomasello, 1999). Another study by Casillas et al. (2016) reported that slowdowns and speed-ups might be relative to the specific context: response latency speeds up earlier for answering simpler questions. However, a cross-sectional study by Gratier et al. (2015) reported a slowdown already from 2-3 month to 4-5 month-old infants. All these findings seem to suggest that age might play a role in turn-taking dynamics, but also that the developmental trajectory of turn-taking is complex and not yet adequately investigated. Based on the aforementioned literature showing that infant response latencies slow down before speeding up again, we expected to find a quadratic relation between the age of infants and their response time latencies.

Adult response latency. As we have argued, turn-taking is an interpersonal activity, where individuals adjust to each other in an interactive fashion (Fusaroli, Rączaszek-Leonardi, et al., 2014). Infants and children would then be likely to adjust to adult response latencies (and vice versa). Accordingly, several studies present evidence for 4- and 9-month-old as well as 4-5-year-old infants adjusting their response latencies to those of adults and vice versa (Beebe et al., 1988; Jasnow & Feldstein, 1986; Welkowitz et al., 1990). Further, direct experimental manipulations showed that when parents delayed their answers by pre-agreed latencies (1 and 3 seconds), their 4 year-old children produced corresponding latencies (Newman & Smit, 1989). These studies lead us to expect an effect of adult response latencies: slower adult response latency will correspond to slower infant response latency and vice versa.

Familiarity of interlocutors and settings. Children have been shown to adjust their response latencies to the predictability of the utterances they're responding to. One could hypothesize that the predictability of the situation (and therefore its familiarity) would also play a role. Indeed, infants are reported to vocalize in a more contingent way when interacting with a familiar partner than with an unfamiliar one (Bigelow, 1998). However, 4-month-old infants have been shown to display shorter response time latencies when they are interacting with strangers than with familiar interlocutors (Jaffe et al., 2001). This latter result would fit with findings in adults, reporting that interactions with unfamiliar interlocutors result in increased vocal activity and more tightly knit turn-taking coordination compared to familiar ones, and that this is even more salient when one of the adults has aphasia (Cappella, 1996; Crown, 1991; Dressler et al., 2009). A possible explanation for the inconsistent findings may include that infants are not reacting to the familiarity of the interlocutor, but to more specific adult behaviors. Indeed, children adjust to the response latencies of slower and faster unfamiliar adults (Bigelow, 1998). The effects of familiarity of other components of the conversational settings - e.g., familiar (i.e., home, daycare) vs. unfamiliar settings (i.e., laboratory, clinic) - have not been directly tested. However, vocal behaviors have been observed to vary: 4- to 5-month-old children produce double the amount of vocalizations when at home compared to laboratory settings (Lewedag et al., 1994), while mothers vocalize more when interacting with their infants in the lab than at home (Belsky, 1980). Further, Jaffe et al. (2001) report that the degree of novelty of the situation predicts the vocal activity (i.e. length of response latencies) on a gradient, meaning that in situations where infants interact with strangers in an unfamiliar environment they are expected to exhibit faster responses than when with strangers at home, which will exhibit faster responses than at home with their parent. These contradictory findings suggest that a closer look at turn-taking differences due to the familiarity of interlocutors and settings is needed.

Activity. Analogously, different kinds of activities might involve different response latencies, as observed for other kinds of conversational behaviors (Dale, 2015; Dideriksen et al., 2020; Duran et al., 2019). While no within-study manipulation was reported in the literature, we can identify at least four different categories of activities during which response latencies were measured. 1) During free play, the adult and the child are generally recorded during spontaneously occurring play (Beebe et al., 1988; Hilbrink et al., 2015; Jaffe et al., 2001; Jasnow & Feldstein, 1986; Kelly, 1994; Kelly & Conture, 1992; Klann-Delius & Hofmeister, 1997; Northrup & Iverson, 2015; Savelkoul et al., 2007; Sawyer et al., 2017; Yaruss, 1997; Zlochower & Cohn, 1996). 2) During more structured everyday activities, the adult and child go through more routinized sequences of events, such as changing diapers (Leonardi et al., 2016; Malin et al., 2011; Marklund et al., 2015; Reissland & Stephenson, 1999). 3) During more focused conversations, the adult and child are focusing more specifically on the vocal interaction within structures at least partially imposed by the experimenter, e.g., by providing a series of objects to be discussed (Bateson, 1975; Carvalho et al., 2019; Gratier et al., 2015; Kondaurova et al., 2020; Lange, 2016; Newman & Smit, 1989; Warlaumont et al., 2010, 2014; Welkowitz et al., 1990) and 4) highly constrained question-answer interactions (Casillas et al., 2016; Clark & Lindsey, 2015). No hypotheses concerning different latencies across these methods have been put forward in the literature.

Language. Cross-linguistic variations observed by Stivers et al (2009) in adult-adult conversations warrant a closer assessment of this factor in child-adult conversations. Indeed, one study of cross-linguistic variation in 2-5-month-old infants found significant differences in response-time latencies for three different languages (US English, French, Hindi) (Gratier, 2003).

Atypical groups. Finally, it might be insightful to study how turn-taking response time latencies vary within various atypical groups. Several studies point towards atypicalities in the development of turn-taking in a variety of different populations – both traditionally defined as atypical (e.g., autistic children) or as special (i.e., insecurely attached children). We refer to all of them as atypical in that they are argued to exhibit potentially atypical turn-taking. Children with cochlear implants as well as insecurely attached infants are slower at responding to their parents compared to more typically developing infants (Klann-Delius & Hofmeister, 1997; Kondaurova et al., 2020); prematurely born infants are faster (Reissland & Stephenson, 1999), and children with a stutter do not show any atypicality (Kelly, 1994; Kelly & Conture, 1992). On the other hand, parents of children with autism, and of prematurely born infants tend to have slower responses to their children, which might affect the child's responses (Reissland & Stephenson, 1999; Warlaumont et al., 2010, 2014). Further, autistic individuals have often been argued to have conversational atypicalities and slower turn-taking dynamics(e.g., Choi et al., 2013; Heeman et al., 2010; Ochi et al., 2019; Ochs et al., 2004). However, no study so far has closely investigated which individual differences within the atypical groups might be related to turn-taking atypicalities, to better reveal which mechanisms are involved (see Weed et al., 2017 for a related example in ASD). From a theoretical perspective, identifying atypicalities in the development of turn-taking in a diverse range of populations might help us understand which mechanisms are indeed involved in turn-taking, for instance if a group presenting slower turn-taking also display a social, but not a linguistic impairment. Further, given the likely foundational role of turn-taking in development, understanding turn-taking atypicalities might help us more effectively tackle linguistic and social irregularities in atypical groups (as in the social feedback loop hypothesis, Warlaumont et al., 2014).

1.4 Summary

The reviewed literature highlights the importance of assessing how infants' cognitive and linguistic development, as well as the interpersonal dynamics inherent in conversations, affect the developmental trajectories of infants' turn-taking abilities. We should note, however, that the studies differ in the questions investigated, with the response latencies often only being reported as a side result. The studies also present conflicting results, perhaps due to confounds between factors (e.g., familiar interactors and free play as a method tend co-occur in most cases, which makes it difficult to disentangle their effects), or perhaps due to the heterogeneous settings, methods and population samples investigated. In the following, we therefore provide a quantitative meta-analysis of the evidence available on the developmental trajectories of turn-taking and the possible factors there involved. We then complement it with a critical discussion of the limitations of the data and analysis, and provide a roadmap for future studies.

2.0 Materials and methods

2.1 Literature Search

We adopted the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Guidelines (PRISMA, Page et al., 2021) to ensure optimal transparency, as shown in Tables S1.1.1-1.1.2 and Figure S1.1 in the supplementary materials. The full process was performed independently by two authors (VN and OV), as well as supervised and double checked by the two other authors (CC and RF). The systematic literature search was conducted on PubMed using the following search term: *"Early mother-infant communication" OR "protoconversation" OR "turn transition" OR "Turn alternation" OR "Turn offsets" OR "Turn timing" OR "Turn allocation" OR "Vocal responsiveness" OR "Vocal contingency" OR "Vocal congruence" OR "Turn-taking" OR "Conversational interaction" AND (infant OR child OR newborn OR baby OR babies)*. The search was not pre-registered. As of March 2021, 232 possibly relevant studies were identified using this search term. In addition to using PubMed, we also performed a backwards literature search (assessing literature cited in our selected papers) and consulted Google Scholar to perform a forward literature search (assessing literature citing our selected papers), which resulted in the identification of an additional 34 studies.

The 266 studies obtained through this process were then screened using the following inclusion criteria: i) studies should include dyadic vocal interactions between a child and an adult (therefore excluding e.g., sign language, since different language modalities have been reported to unfold along different time scales (De Vos et al., 2015)), ii) studies should contain quantitative measures of temporal distance between turns (either adult to child or child to adult), iii) children should be aged between 0 and 96 months. We opted for this 96-month cut-off because our main interest concerned the early development as well as the initial integration of the turn-taking structure and budding linguistic abilities. Both published and unpublished literature was considered for inclusion. 240 papers were accordingly excluded: 156 papers were excluded as unrelated to our topic (e.g., focused on physiological synchrony in the dyad, or alternating moves in a board game); 10 papers which did not include any vocal interaction (5 of which focused on sign language exchanges); 28 papers did not include any child-adult interactions; 46 papers did not contain quantitative measures of response latency. Several studies included multiple age groups or were of longitudinal nature, and therefore contained multiple response-time latencies. The final sample consisted of 26 unique papers and 78 response time latencies. Of these 26 papers, 12 included atypical groups, which constitute 22 of the 78 estimates of response-time latencies.

For more detailed information about the studies, the reader is referred to the supplementary materials. Specifically, Table S1.2 summarizes every study, including characteristics of the samples; Figure S1.3 shows a world map of the languages investigated (from exclusively North American and European sites); and Section S1.4 includes a discussion of possible biases in the search and selection of the studies.

2.2 Citation network

To better explore the literature on infant-adult turn-taking, we extracted the reference list of each included paper using Web of Science, and filtered the lists to only include papers used in the meta-analysis. We used these data to build two different network plots, see Figure 2a.-b. The first, a coupling network, identifies similarities in citing patterns and the discourses the papers refer to. In this network, two papers are connected if they both cite a common third paper. The more similar the bibliographical references are between the two papers, the stronger their connection. The second network, a direct-citation network, identifies which papers cite each other and the coherence among the contributions to the field. In both networks, we assessed the presence of clusters using the Louvain community algorithm, which identifies groups of papers (clusters) that are more likely to be connected to each other than to papers outside of the cluster (Traag & Bruggeman, 2009). To provide the reader with a point of comparison for citation network plots, we have added Figure S2.1 from Cox et al. (2021) in the supplementary materials. The visualizations relied on the *bibliometrix* and the *igraph* packages for R (Aria & Cuccurullo, 2017; Csardi & Nepusz, 2006).

In the coupling network in Figure 2a, the study populations and experimental conditions of the studies appear to be reflected in their citation patterns. The dense, blue cluster of studies, for example, primarily focuses on vocal interactions among atypically developing children and their parents (Kelly, 1994; Kelly & Conture, 1992; Newman & Smit, 1989; Savelkoul et al., 2007; Sawyer et al., 2017; Yaruss, 1997). The green cluster, on the other hand, is comprised of studies of typically-developing infants engaged in spontaneous interaction with their caregivers (Bateson, 1975; Jaffe et al., 2001; Dahlby et al., 2011; Gratier et al., 2015; Jaffe et al., 2001; Jasnow & Feldstein, 1986; Beebe et al., 1988; Klann-Delius & Hofmeister, 1997; Welkowitz et al., 1990; Lange, 2016; Zlochower & Cohn, 1996). Lastly, the orange cluster investigates child-caregiver interaction using more structured experimental activities, such as question-and-answer pairs (Carvalho et al., 2019; Casillas et al., 2016; Clark & Lindsey, 2015; Hilbrink et al., 2015; Kondaurova et al., 2020; Leonardi et al., 2016; Marklund et al., 2015; Reissland & Stephenson, 1999; Warlaumont et al., 2014; Northrup & Iverson, 2015). The direct-citation plot (Figure 2b) reflects a similar picture of a fragmented research field. For instance, the blue cluster of studies on atypical turn-taking exhibits a similar high degree of within-cluster coherence to that in Figure 2a. More generally, these citation plots indicate that the papers under investigation examine a wide variety of relevant aspects of infants' ability to take conversational turns, and that this manifests in the citation patterns of the studies.

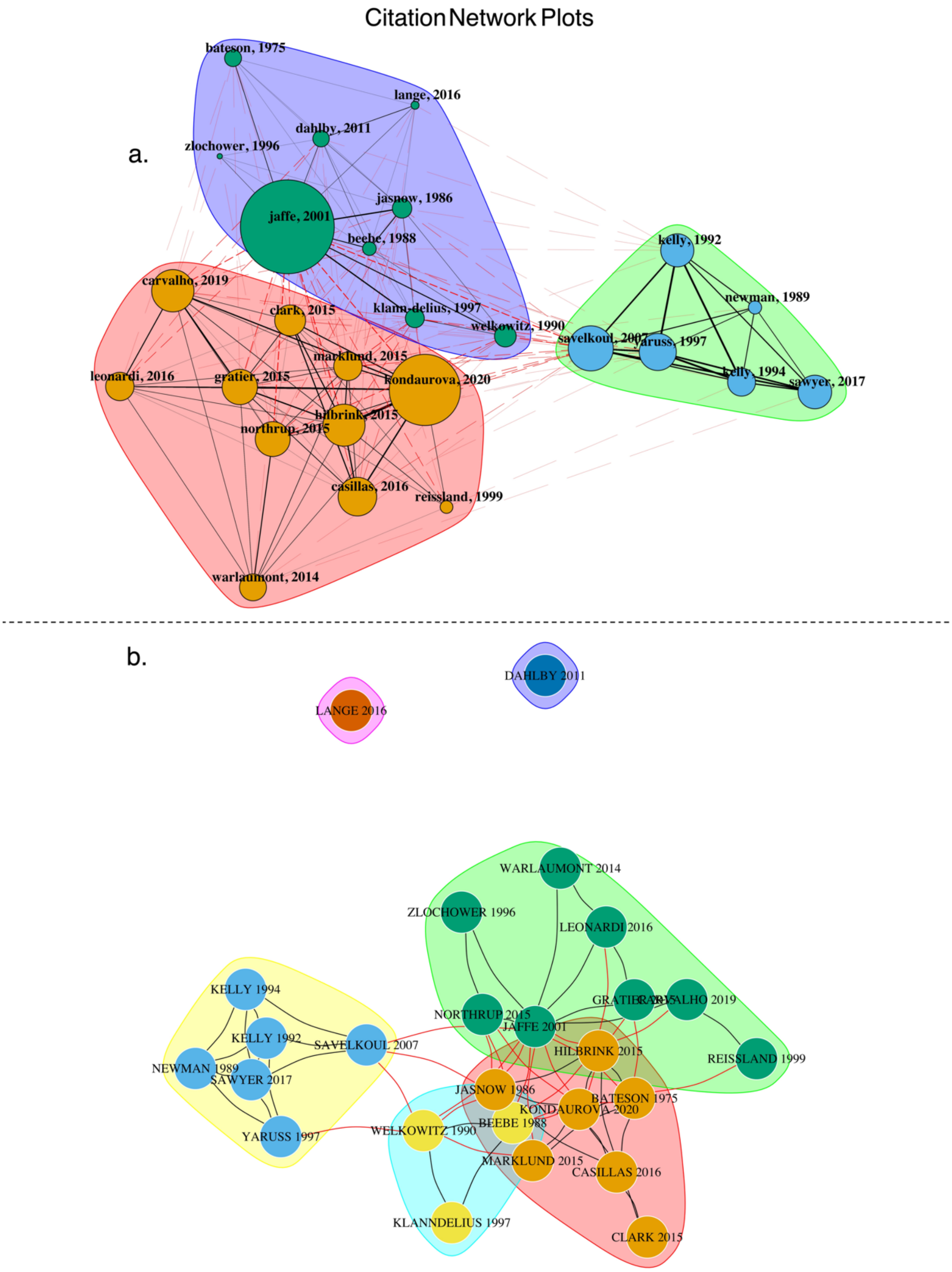


Figure 2. Coupling and direct-citation networks for the studies reviewed. Plot a. above represents a coupling network, where papers are connected according to how much overlap there is in their bibliography. The size of the dots indicates the number and strength of links connected to that dot (paper). The below plot represents a direct-citation network, where the papers are connected according to whether one of them cites the other. Colored clusters in both networks are inferred using a Louvain community detection algorithm and indicate

papers more likely to be connected to each other than to papers outside of the cluster. The two isolates, Lange (2016) and Dahlby (2011), represent grey literature, as they refer to a thesis and conference proceedings, respectively. In both panels, distance between two dots is an approximate measure of how similar (e.g., in connections to other dots) they are.

2.3 Data extraction

In order to perform a meta-analysis of the studies, we extracted the estimates for child and adult response latencies (mean and standard deviation) as well as available values of the following seven variables: i) child age (mean and standard deviation), ii) adult response latencies, iii) familiarity of interlocutors and iv) interactional settings, v) type of activity, vi) language being spoken, and vii) whether the child had a developmental disorder or other relevant conditions. We further extracted recording duration and number of dyads involved in order to facilitate informed missing data imputation.

Response latencies were converted to milliseconds to ensure comparability and logarithmically transformed using the natural base. The transformation accounts for the long-tailed nature of the data, and was possible because no study reported negative average latencies (i.e., overlaps in vocalizations). Log-transformation is a non-linear transformation and that makes the interpretation of the model coefficients difficult. For instance, if an increase of 12 months in age were to relate to a decrease in response latency of 0.1 on a log scale, this increase is impossible to interpret unless we know the intercept. If the intercept (response latency when age is 0) is 7 on a log scale (about 1097 ms), a decrease of 0.1 implies a decrease of circa 104ms (6.9 on a log scale being about 992 ms). If the intercept were 6 on a log scale (circa 403 ms), a decrease of 0.1 would only imply a decrease of 39 ms (5.9 on a log scale being about 365 ms). We therefore report effects of the moderators on the outcome scale to facilitate interpretation. In other words, we exponentiate the expected outcomes at different values of the moderators and report their differences. Age was converted to months, including 2 decimals, and scaled (subtracting the overall mean and dividing by the overall standard deviation) to facilitate model estimation.

Not all selected papers fully reported the relevant numeric estimates for our analyses. 1 paper (Kondaurova et al., 2020) only reported figures, from which we were able to extract the missing estimates using WebPlotDigitizer by Automeris. 1 paper (Hilbrink et al., 2015) only reported the median range of latencies, from which we extracted the average with the following formula: $Average\ response\ latency\ = Max. - \frac{Max. - Min.}{2}$. Given that the range covers the highest and lowest possible values, identifying the central value (half the distance between maximum and minimum values, counting down from the maximum value) provides a rough estimate of the mean with the assumption that the distribution is roughly symmetric.

3 papers only reported the child or the adult response latencies and 8 papers only reported average estimates of response latencies, but not their variance; and these values could not be reconstructed from the papers. We did not contact the authors and chose to impute the missing data, using a multiple multivariate imputation process by chained equations as implemented in the R package *mice,* (Buuren & Groothuis-Oudshoorn, 2011). In order to account for the statistical uncertainty involved in the process of imputation (Azur et al., 2011; Sterne et al., 2009), the missing data were imputed 20 times, thus generating 20 separate datasets, using the following variables (where available): i) child response latency (mean and standard deviation), ii) adult response latency (mean and standard deviation), iii) mean age of infants. Finally, these imputed values were post-processed to ensure realistic values: the ranges of the imputed values were restricted to the interquartile range of the observed data. All analyses in this paper were performed on the resulting twenty datasets in which the missing data were imputed. We also tested the impact of the imputation, i.e., we fitted models on the original and imputed datasets and ensured the results were analogous, see supplementary

materials, S3). The raw data are available on OSF and will be made available on MetaLab (https://langcog.github.io/metalab/) upon publication of the manuscript.

Note that many additional dimensions are likely to impact turn-taking dynamics, but remain largely unreported in the studies: from ethnical and socio-economic variation, to the more fine-grained structure of the interaction (e.g., was the sequence adult- or child-initiated?). These dimensions could not therefore be systematically extracted and considered in this study.

2.4 Statistical modeling

Meta-analytic models are usually deployed to pool the effect size estimates across the whole literature, weighted according to their variance, so as to achieve a more reliable estimate of the effect in the population. In this specific meta-analysis, we aimed at assessing the average child response latency in child-adult conversation and its relevant moderators. In other words, differently from most meta-analyses, our study is not concerned with aggregating effect sizes (e.g., differences between conditions, or existing correlation estimates and their uncertainty) from previous studies, since our questions of interest are mostly still unexplored. This study is concerned with aggregating the actual response latencies as they are reported, sometimes as secondary data. We then use these measures across studies to assess whether they are affected by possible moderators.

Since heterogeneous results are expected due to the differing nature of the methods and population samples in the studies, we relied on a random-effects model (Fernández-Castilla et al., 2020). By means of a multi-level structure, these models are able to address the heterogeneity between studies by assuming that every study and every sample has a different 'true' response time latency, which does not necessarily overlap with the average response latency of the overall population. We modeled the outcome using a Student's t likelihood (that is, modeling the error term of the model as following a Student's t distribution), where the standard deviation was given by the standard error of the effect size estimates, and the long-tails of the distribution allowed outliers to span out with only minimal effects on the estimates (Jylänki et al., 2011). This procedure produces robust inferential results, as single studies are weighted by the inverse of their variance (thus more uncertain effect estimates weigh less) and outliers exert a limited effect on the pooled response latency estimation (as they are accounted for in the long-tail of the t-distribution). Further, a subset of studies reported trimming their data by excluding response latencies above a certain threshold (most commonly between 1 and 3 seconds). We explicitly modeled this censoring to make the estimates more comparable and fully account for available information. Estimates from the model are reported as mean and 95% Compatibility Intervals (CI) of the posterior estimates, that is, the range covered by the middle 95% of the posterior samples.

In order to investigate the developmental trajectory of turn-taking and whether the moderators we previously identified can systematically explain the variance in the response time latencies of infants across studies, we fitted three different models only on the typical population. The first model (**baseline model**) conditions the average response latency (on a log-scale) only to the variance due to articles and studies within articles.

Due to our specific interest in the developmental trajectory of response latency, our second model (**age model**) included the mean age of infants as a predictor. To account for the expected slow-down followed by a speed-up in turn-taking, we included age as a quadratic in order to appropriately model its effects. Age was also included in the random structure to explicitly model potential different effects of age in cross-sectional and longitudinal studies.

The third model (**moderators model**) included the following predictors: i) mean age of the infants, ii) logged adult response latency, iii) familiarity of interlocutors, iv) familiarity of interactional settings, v) type of activity, and vi) language. However, some contrasts might be confounded due to the sparsity of available data (e.g., data for non-English speaking infants coming from only one study with only familiar interlocutors). Therefore, we first checked cross-tabs in order to determine which contrasts should be tested and would allow us to make meaningful inferences (see Figure 3 and Table 1). The resulting meaningful contrasts are reported in Table 2.

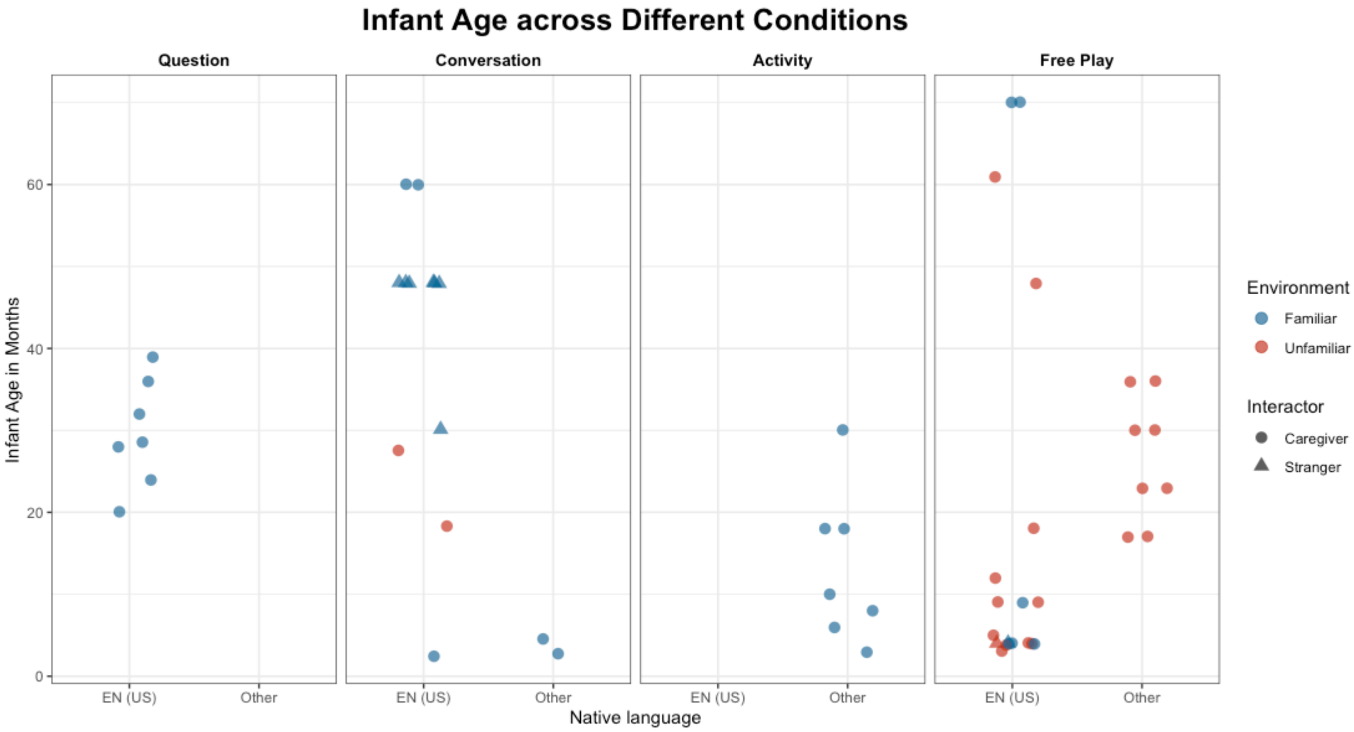


Figure 3: Cross-tab showing the distribution of infant age across different activities, languages, environments and interactors.

The subsequent models included also the data from atypical populations (i.e., autism, preterm, stutterers, insecurely attached and hard of hearing), since systematic variations might provide cues as to which mechanisms modulate the development of turn-taking. The fourth and fifth models (**typical/atypical baseline and age models**) corresponded to the first and second models but we added a categorical variable which distinguished between the typical and five other atypical groups and added separate intercepts for these atypical groups.

The last model (**typical/atypical moderators model**) corresponds to the third moderators model but with a different intercept for every population and moderator. We planned to assess the impact of the same predictors as in model 3. However, when assessing the data available (see table S9.1 in the supplementary materials), we concluded that no meaningful between-group comparisons could be made, as it would be confounded by some conditions only happening in specific atypical groups. Therefore, it was decided that beyond the effect of age and adult response latency, the other planned moderators would not be examined.

Detailed description of the modeling choices, implementation and quality checks are available in the supplementary materials (see S3-S10). Moreover, we present exploratory analyses of how infant sex, country and language affect our effects of interest in Figure S11.1 and Table 11.1 in the supplementary materials.

3.0 Results

3.1 Response latency estimates

The baseline model was fitted to 55 measurements from 21 studies allowing us to estimate the overall intercept for infants' response latency. The overall estimate of infants' response latency was approximately 1 second (1089 ms [870, 1363]). If we include between-study variance, we can expect the mean response latency in a new study to range between 694 and 3559 ms. See Figure 4.

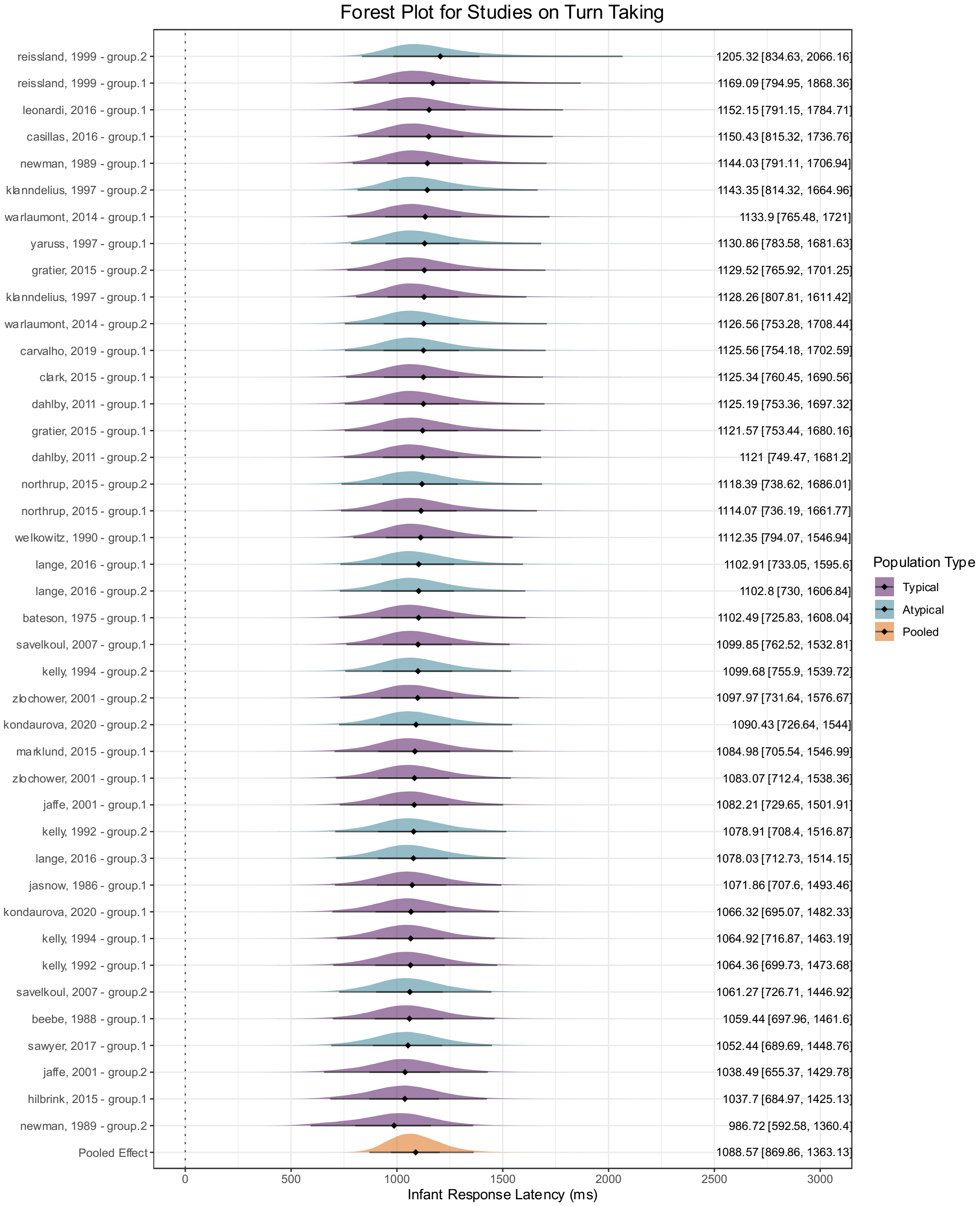


Figure 4 - Forest plot of the studies included in the meta-analysis. Studies of typical populations are indicated in purple, and atypical in light blue. Note that the pooled effect estimate in the plot includes studies on atypical groups and therefore differs from that reported in the main text.

3.2 Moderator Analysis for Heterogeneity

Including age **(age model)** provides an increased explained variance (from 14% to 25%, Bayesian R^2), but also a decrease in model generalizability (stacking weight = 0.053). Response latencies seem to increase with age up to 20 months of age, to then decrease (see Figure 5): At 1 month, the model predicted an infant response latency of 1289 ms [286, 4159], at 11 months of 1386 ms [547, 3148], at 20 months of 1429 ms [576, 3130], at 25 months of 1380 ms [522, 3128], at 30 months of 1307 ms [410, 3293], at 35 of 1251 ms [261, 3797]. However, caution is needed given the model is likely overfit (more explained variance, but lower generalizability). A control analysis including only conditions with a wide range of ages reported gave analogous results and is reported in the appendix (Figure S5.3).

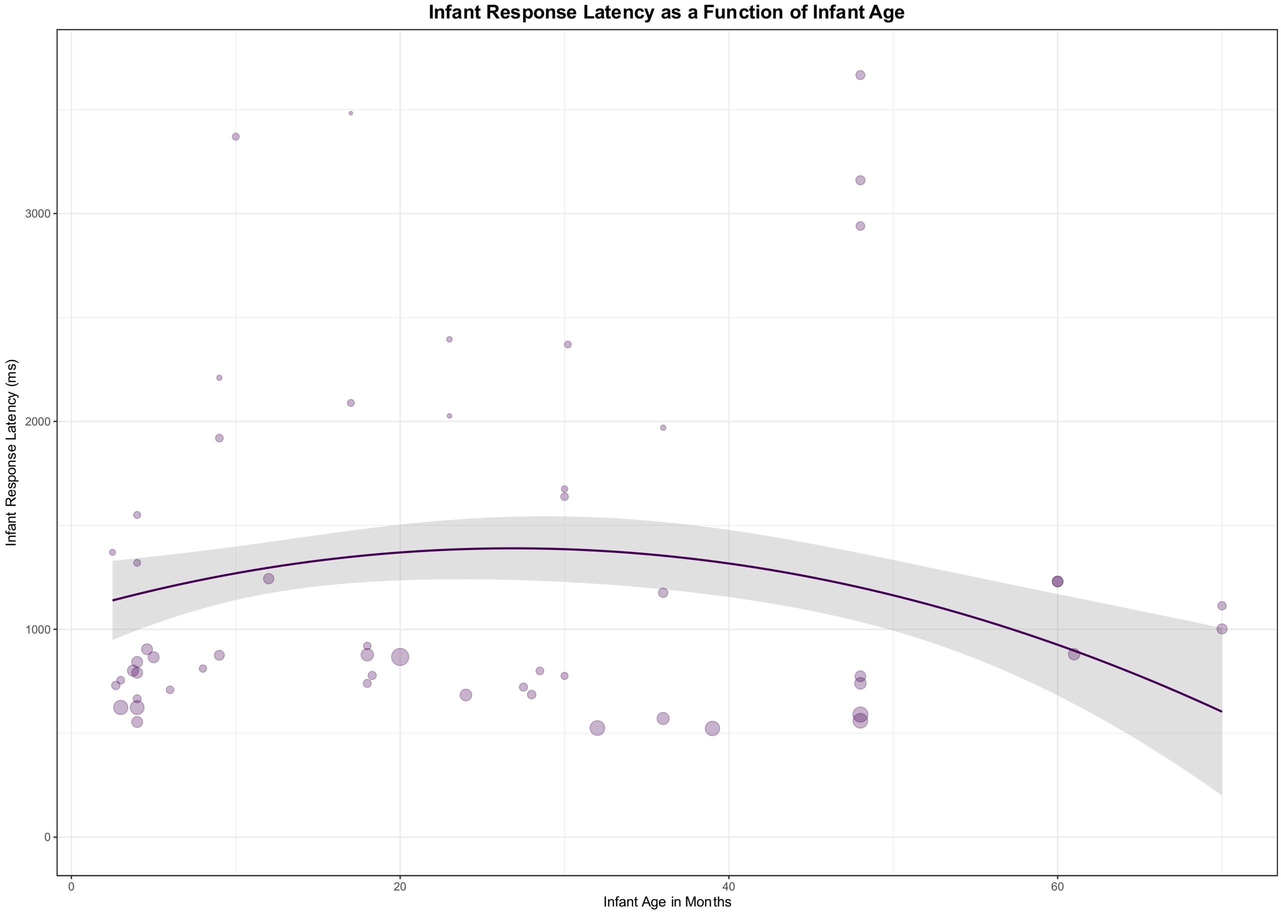


Figure 5. Developmental trajectory of response time latencies: model prediction and actual data. The x-axis indicates the infants' average age in months, the y-axis response latencies in milliseconds. Red points represent

the observed data, blue lines the model's inferred trajectory of response latencies, including shaded standard errors. The size of the points is inversely proportional to the standard error of the estimate (i.e., the larger the point, the smaller the standard error).

Including additional moderators provided at least some improvement in generalizability compared to the age model (stacking weight = 0.614) as well as the baseline model (stacking weight = 0.581). In other words, its estimated generalizability to new data was comparable to that of the age model, but different enough predictions are produced that generalizability would be improved by averaging across the two models' predictions. The evidence available for how response latencies vary with most moderators seems fairly sparse and unreliable, with the exception of adults' response latency, and experimental activity of Question (cf. Table 2). While the infant response latency between studies exhibits heterogeneity, they do seem to be robustly linked to those of adults (beta = 1.13 [0.78, 1.43] on a log scale). Keeping everything else constant, an increase of 1 second in adult response latency (e.g., 400 to 1400 ms) would be related to an increase of over 1 second in infant response latency (from 600 to 1857 ms). However, as adult response latency and infant age are correlated (r = 0.33), caution in interpretation should be exercised. Within the meaningfully comparable contrasts (cf. Table 1; Figure 3), most of the moderators show no reliable influence on infant response latencies (ER < 2 and > .5), with the exception of Question as experimental activity, which elicits slower infant response latencies than the other activities (ER > 10).

		Method	A	C	P	Q
Environment	Interactor	Language				
Familiar	Caregiver	EN (US)	0	3	6	7
		Other	7	2	0	0
	Stranger	EN (US)	0	7	1	0
		Other	0	0	0	0
Unfamiliar	Caregiver	EN (US)	0	2	11	0
		Other	0	0	8	0
	Stranger	EN (US)	0	0	1	0
		Other	0	0	0	0

Table 1 - Cross-table for typical infants across method, environment, interactor and language. Note. The abbreviation A under method represents 'Activity', C represents 'Conversation', P represents 'Free play' and Q represents 'question'.

Moderator	**Hypothesis tested**	**E.R.**
Age	Age < 0	1.62
Age^2	Age^2 < 0	8.77
Adult response latency	Response latency > 0	> 1000
Environment	Familiar > unfamiliar	1.45
Method	Free play > conversation	0.93
Method	Question > conversation	12.39
Method	Question > free play	22.93

Language	Other > English	3.35
Interactor	Caregiver > stranger	1.32

Table 2 - Evidence ratios for moderators. See table S5.1 for the detailed coefficients of each predictor.

3.3 Typical/atypical baseline model

The **typical/atypical baseline model** was fitted to 78 measurements from 26 studies allowing us to estimate and compare the separate intercepts for the typical and different atypical groups (see Table 3). Autistic, preterm, and insecurely attached children exhibited slower response latencies than typically developing children. The typical/atypical baseline model was able to explain 21% of the observed variance.

Population	**Estimates [95% CIs]**	**Range including between study variability**	**Evidence Ratio (compared to Typical)**
Autism	1345 ms [528, 3003]	[212 ms, 11403 ms]	Slower: 2.86
Cochlear Implant	995 ms [445, 2060]	[284 ms, 6556 ms]	Faster: 1.00
Typical	1009 ms [806, 1295]	[604 ms, 3043 ms]	-
Insecurely attached	1238 ms [676, 2200]	[504 ms, 6214 ms]	Slower: 2.97
Preterm	1411 ms [799, 2323]	[630 ms, 5521 ms]	Slower: 6.12
Stutterer	945 ms [637, 1404]	[462 ms, 3171 ms]	Faster: 1.62

Table 3 - Overall estimates of the infants' response time latencies and range for the typical/atypical baseline model

3.4. Typical/atypical moderator models

Including *age* as a predictor of infant response latency did not provide additional generalizability (stacking weight = 0.00), and decreased explained variance (from 21% to 9%). However, the expected effects of age within each population were rather small and uncertain, due to the sparsity of data (see Figure 6).

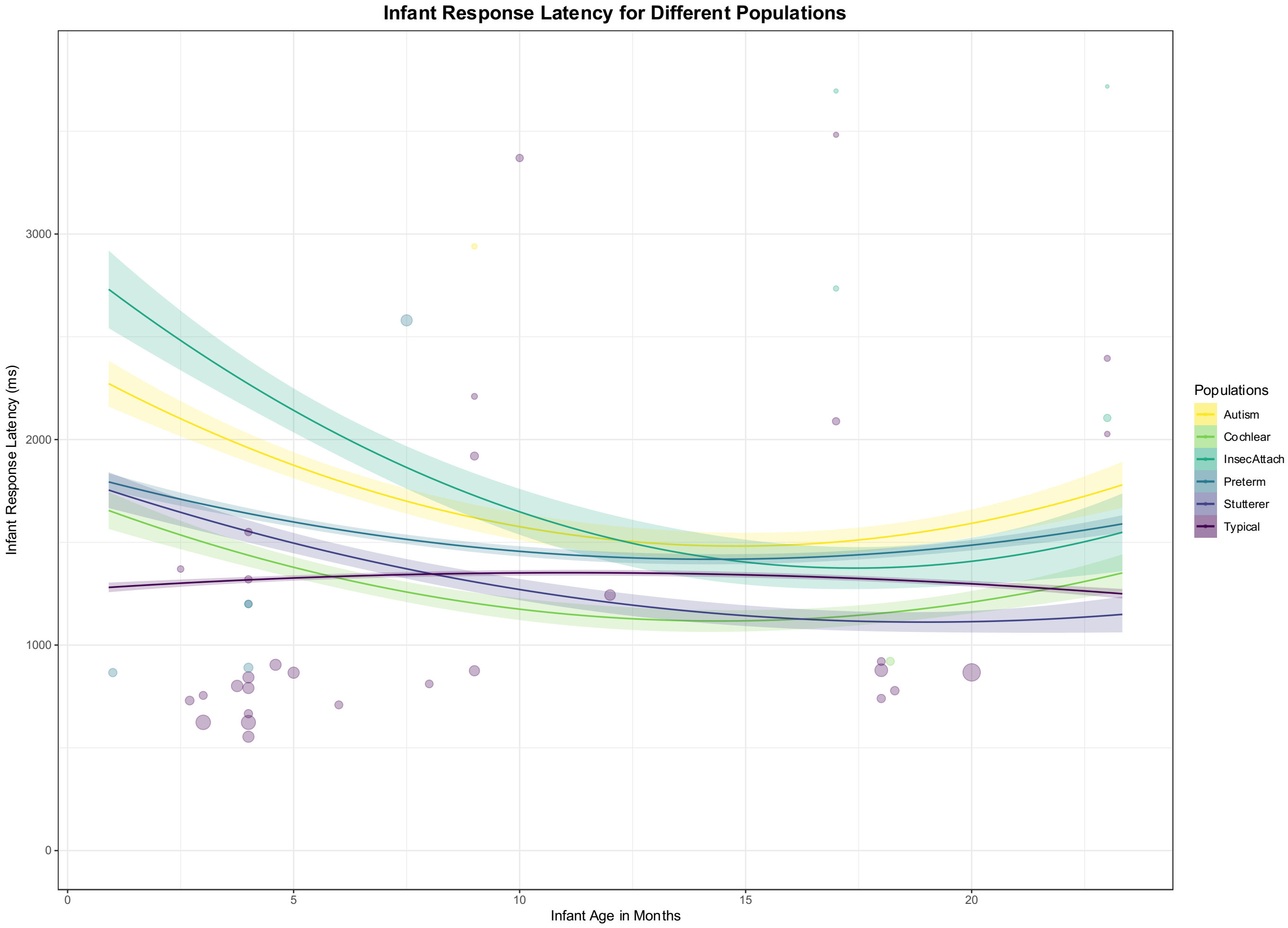


Figure 6 – Estimated effects of age in typical and atypical populations. The shaded area indicates the standard error of the estimated population mean for each given group. The size of the points is inversely proportional to the standard error of the estimate (i.e., the larger the point, the smaller the standard error).

Adding *adult response latency* to the age model resulted in increased generalizability (stacking weight = 0.879) and explained variance (from 9% to 44%). We observe that adult response latency is generally the more robust predictor (except for insecurely attached infants), while age becomes more uncertain (see Tables 4). However, as the data are sparse (see Figure S9.1 in the supplementary materials), much interpretive caution should be exercised.

Population	**Hypothesis tested**	**E.R.**
Autism	Age < 0	1.02
	Age^2 > 0	1.09
	Adult response latency > 0	1.68
Cochlear Implant	Age > 0	1.08
	Age^2 < 0	1.04
	Adult response latency > 0	1.14
Typical	Age < 0	2.31
	Age^2 < 0	11.81
	Adult response latency > 0	> 1000
Insecure Attached	Age < 0	1.56
	Age^2 > 0	1.09
	Adult response latency < 0	1.08
Preterm	Age < 0	1.23
	Age^2 > 0	1.15

	Adult response latency > 0	3.92
Stutterer	Age < 0	1.16
	Age^2 < 0	1.56
	Adult response latency > 0	1.25

Table 4 - Evidence ratios for the moderator effects in the different populations

4.0 Discussion

The current study aimed at better understanding the current state of research on the development of turn-taking behaviors in human infants. In particular, we wanted to investigate the developmental trajectory of infants' turn-taking abilities and identify the key moderators affecting them in order to better constrain the space of potential mechanisms underlying turn-taking.

The above meta-analysis indicates that typically developing children exhibit an average response latency of about 1s and that atypically developing infants exhibit slower response latencies (i.e., particularly preterm infants and infants with autism and insecure attachment). The meta-analysis also provides information about the influence of moderators. Specifically, the results indicate that infant response latencies might exhibit a quadratic trajectory of development in early infancy and an increase until 20 months of age. Furthermore, these latencies exhibit a strong correlation with adult response latencies in both typical and atypical infants. This correlation highlights the interactional nature of turn-taking, as discussed further below. The analysis also indicates that insufficient evidence is available as to how infant response latencies vary according to age, language, methodology, and the familiarity of the environment and interactor. In the following, we contextualize these findings in the wider literature and comment on developmental trajectories, interpersonal dimensions, atypical development and future directions of research on turn-taking.

4.1. Developmental trajectory of turn-taking

A popular model (the Interaction Engine Hypothesis, introduced in Levinson, 2006) of turn-taking suggests a slowdown at around 12 months of age as children integrate their interactional abilities with their growing capacities for language (i.e., their intentional use of words in particular), followed by a speed-up at around 18 months. Earlier studies place the slowdown at 9 months (Hilbrink et al., 2015), or even 4-5 months of age (Gratier et al., 2015), and do not investigate the subsequent speed-up. The meta-analytic pattern is very uncertain and might point to an early slowdown (Gratier et al., 2015) and a speed-up at around 20 months of age. Albeit cautiously, due to the aggregated data analyzed and likely overfitting of the model, we speculate that this pattern is not in contradiction with the Interaction Engine Hypothesis. However, the continuous nature of the processes underlying language acquisition should play a more important role. In particular, the growing recognition of the importance of intentional communication, and the ‘nine-month revolution’ (Tomasello, 1999, as suggested in Hilbrink et al 2015) might not be the one determining factor for response time latencies of infants, or at least should have a much more diffuse and gradual impact. Further, one should more closely consider the content, complexity and contexts of interactions and turn-by-turn exchanges, as we might expect them to affect response latencies. For instance, Casillas and colleagues (2016) have shown that the complexity of questions differentially affect children at different ages, with complex questions being answered relatively more slowly by younger children. The meta-analytic results likewise provide evidence for questions exhibiting slower response time latencies (see Table 2). As meta-analytic data could present multiple confounds, however, these findings emphasize the importance of longitudinal studies including data on both the linguistic and social development of the child, so we can better assess the concurrent development of these measures and response time latencies. This could also help to build more detailed versions of the Interaction Engine Hypothesis and to generate more specific predictions about how and why trajectories might differ on a child-to-child basis.

In any case, the response latencies we found (roughly 1 second) are much slower than those reported for adult-adult conversations (ranging between 7 and 470 milliseconds, see Stivers et al., 2009). This result suggests that even the older children in our data (8-year-olds) fail to display clear signs of anticipating conversational turns in their response latencies and that their developmental trajectories are far from complete. On the other hand, previous work has shown that 3-5 year old children anticipate conversational turns in a controlled task, despite the longer latencies, which might be due to language production skills still being in development (Lindsay et al., 2019).

4.2. Interpersonal adaptation

Turn-taking has often been conceptualized as a shared, cooperative process (Fusaroli, Rączaszek-Leonardi, et al., 2014; Manson et al., 2013; Pickering & Garrod, 2021; Wilson & Wilson, 2005). Interlocutors come to develop a shared tempo in their syllable lengths and response latencies, thus making predictions and alternation in speakers smoother. In line with these ideas, we observe a strong relation between the response latencies of children and adult interlocutors. Given the coarse graining of the data we have available (summary statistics at the study level), it is hard to assess whether both children and adults adapt to each other and whether this happens at a turn-by-turn basis or at a slower timescale. Therefore, one can but speculate as to whether adults are simply adjusting to the longer time required by children to answer as they engage in more and more complex conversations, or are more actively providing scaffolding to the children's development, implicitly helping them to take enough time (Gratier et al., 2015; Hilbrink et al., 2015; Stern et al., 1977). In Figure 7 below, we depict four potential causal structures underlying the interrelatedness among child age, child response latency and adult response latency. Structures B, C and D would apply if the developmental trajectory of infant response latencies depends in part on their relation to caregivers' response time latencies. This process of mutual adaptation may in turn be affected by child age in various plausible ways. For example, i) child age may primarily affect children's response time latencies and adults may then adapt the timing of their own responses accordingly (structure C). Alternatively, ii) if adults adjust their response time latencies to scaffold their child's development, child age may primarily affect adult response time latencies, to which children then adapt their own response latencies (structure D). Finally, iii) child age could also influence both child and adult response latencies, and child-adult adaptation may take place to varying extents (structures A & B).

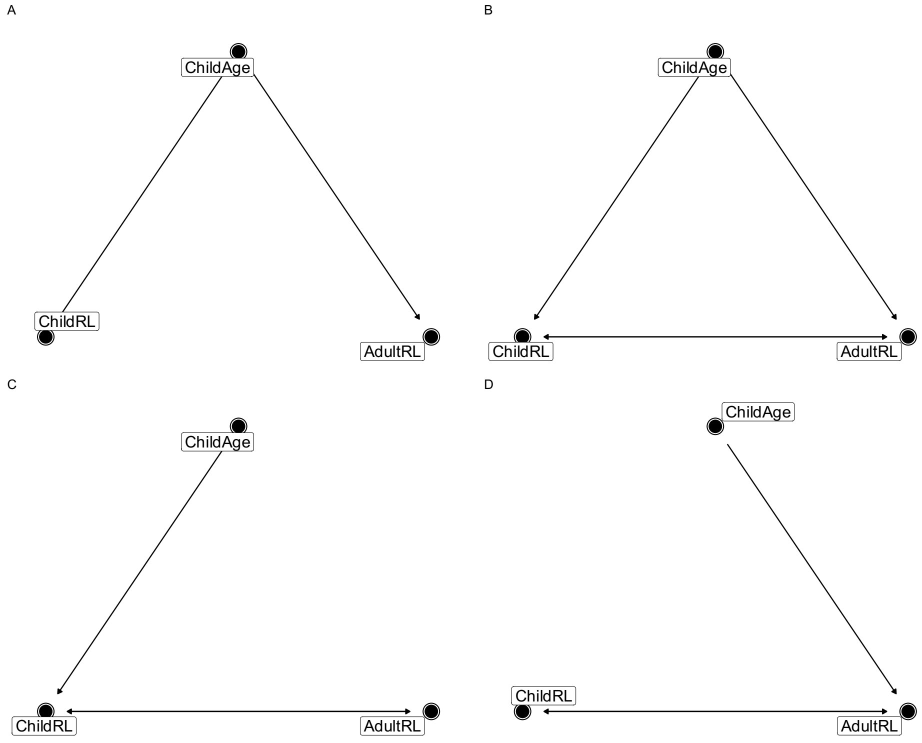


Figure 7 - Potential causal structures underlying the dependency among child age, child response latencies, and adult response latencies. Each node represents a variable, and each arrow a hypothesized causal

relation. Note that in contrast to conventions in the visualization of causal connections, here we occasionally allow for bidirectional causality: infants and adults influencing each other.

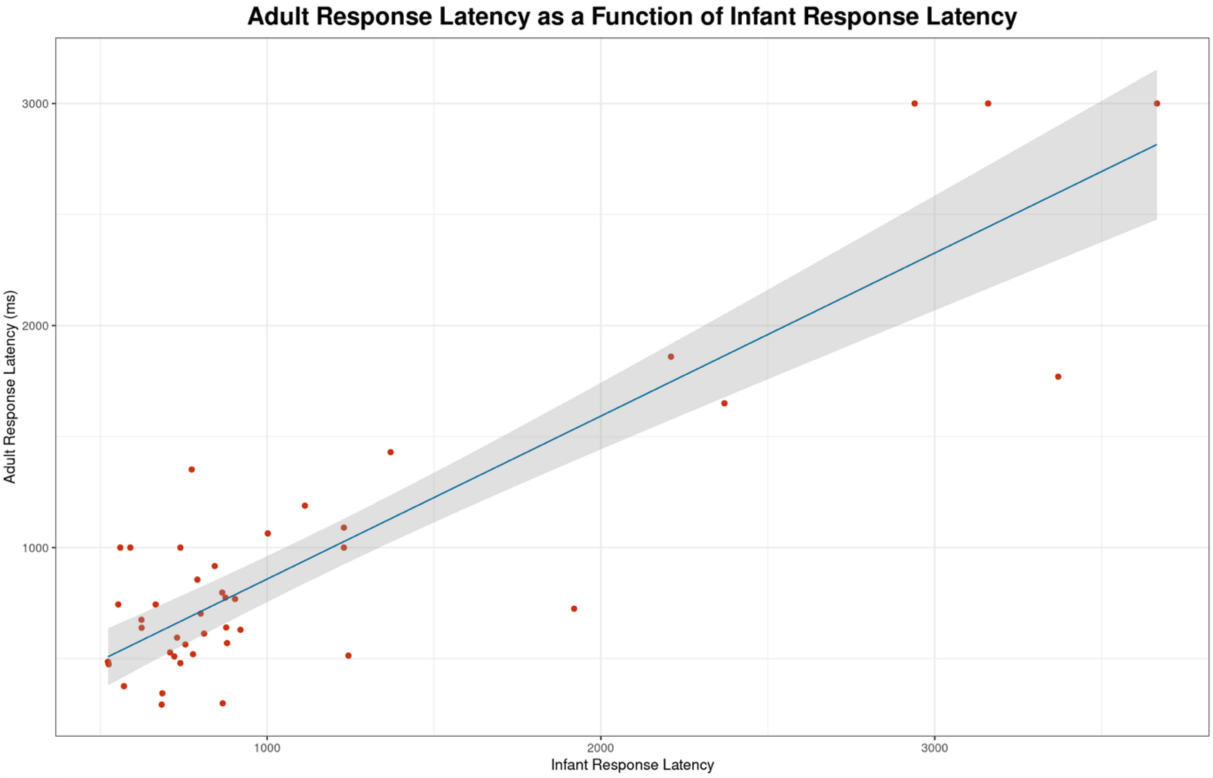


Figure 8 - Plot of adult response latency as a function of infant age for the data on typical infants. The plot shows a positive correlation between the two moderator variables.

In order to better understand the interpersonal dynamics of turn-taking and the reciprocal roles of child and adult response latency, we need both experimental setups manipulating response latencies (Newman & Smit, 1989) as well as fine-grained longitudinal datasets, with turn-by-turn latencies (Weed et al., 2017). Availability of turn-by-turn data would allow more nuanced analysis of how self-regulation (maintaining a certain average tempo) and interpersonal adaptation (following the other's tempo) of responses unfold in real time (Dale et al., 2013; Ferrer & Helm, 2013; Fusaroli et al., 2016; Rast & Ferrer, 2018; Takahashi et al., 2013).

4.3. Typical/atypical populations and other moderators

We included atypical populations with a diverse range of impairments in this meta-analysis, with a view to investigating how these conditions affect response latencies and the developmental trajectories of infants' turn-taking abilities. Preterm infants and infants diagnosed with autism and insecure attachment stood out as having the slowest response latencies. While these groups involve both social and linguistic developmental impairment, which might motivate the slower turn-taking patterns (Klann-Delius & Hofmeister, 1997; Warlaumont et al., 2014), the sparsity of the data make mechanistic explanations untenable. Although the data were too sparse to provide realistic descriptions of the developmental trajectories in these groups, the connection between infant and adult response latencies seems to be held constant across most of the groups.

We do not observe any reliable effects from the familiarity of environment and interactors, method and language, with the exception of the experimental activity of Question exhibiting slower infant response latencies. These null effects and the finding of an influence of constrained question-answer interactions (Casillas et al., 2016; Clark & Lindsey, 2015) on infant response latencies should be interpreted with caution. The data available to contrast these conditions remain sparse, are often confounded across moderators, and aggregate studies with substantial differences within the same condition (e.g., studies focusing only on questions-answers and other studies including all response latencies). Targeted studies with within-subject manipulations are therefore needed to investigate these more subtle effects, for instance by recording the same infants interacting with familiar and unfamiliar interlocutors in familiar and unfamiliar environments.

4.4. Future directions

Systematic reviews and meta-analyses function as moments of self-reflection for a field of research, more aimed at identifying good practices, pitfalls and new directions than state-of-the-art estimates, which are likely inaccurate (see, Fusaroli et al., 2022; Lewis et al., 2020). Our exploratory network analyses indicated a field examining a wide variety of aspects of the development of turn-taking, often reporting infant response latencies as side results to more central research questions. We intend for this systematic review and meta-analysis to serve as common ground amongst researchers interested in the development of turn-taking as well as in neighboring fields of animal communication, adult-adult interactions, and human-computer interactions.

Further, the field could benefit from the development of shared curated resources, both data and computational models. Most of the studies included in the meta-analysis presented several issues for making strong inferences: i) they were cross-sectional, or covered only short longitudinal timespans, ii) they presented roughly aggregated data (e.g., age instead of measures of linguistic and social development); iii) they provided group-level summary statistics instead of individual-level data; iv) they contributed interaction-level estimates, instead of turn-by turn response latencies; and v) the data lacked variability: 51 of the 78 effect sizes reported concerned US English (see also Figure S1.2 for a representation of the geographical reach of the data). In order to advance our understanding of the development of turn-taking, we need access to comprehensive cross-linguistic and more representative (e.g., of socio-economic, cultural and ethnic diversity) datasets that provide turn-by-turn response latencies and the possibility to code for conversational moves (Bergey et al., 2021; Nikolaus et al., 2021), in combination with longitudinal assessments of linguistic and social development (Fusaroli et al., 2019; Naigles & Fein, 2017). A recent paper and R package has also been produced to provide streamlined standard ways of pre-processing the data (Casillas & Scaff, 2021), and other approaches are being developed to identify diverse types of interactional sequences and provide a partially automated coding of them (Bergey et al., 2021; Nikolaus et al., 2021).

Better shared data resources can provide the necessary testing ground for computational models of turn-taking and turn-taking development. We have mentioned the Interaction Engine Hypothesis, Coupled Oscillators models, and dyadic models of self-regulation and adaptation, which are promising but currently underspecified to capture the key patterns we see in the meta-analytic data . By providing computational implementations of these and other models, future research can be more explicit about the assumptions of the theories, as well as compare the predictions of different models and test them on the data available (Guest & Martin, 2021; Navarro, 2021; Rich et al., 2021). Further, the development of better mechanistic models can help the field better capture the meaningful aspects of cross-linguistic, socio-demographic, ethnic and cultural diversity. Better articulating which aspects of diversity might play a role in the mechanisms at stake, we can create more representative samples for those aspects and develop more meaningful causal inference analyses (Christiansen et al., in press; Deffner et al., 2021)

Finally, our understanding of the development of turn-taking could benefit from a more explicit dialogue with the study of turn-taking across language modalities (e.g., sign languages and co-speech gestures), adult-adult turn-taking as well as animal communication. Different language modalities have been reported to unfold at different time scales (De Vos et al., 2015), and could bring important discussions about how timing is adjusted to different contextual constraints. Adult-adult studies can provide a picture of fully developed turn-taking abilities and a target for child-adult trajectories, while animal communication studies can provide both a comparative and evolutionary perspective (Demartsev et al., 2018; Pika et al., 2018; Takahashi et al., 2013, 2016).

5.0. Conclusion

This systematic review and meta-analysis revealed a possible developmental trajectory of infant response latencies: there is uncertain evidence that response latencies might gradually increase up to 20 months, and quite reliable evidence that they exhibit a strong relation to the adult interlocutors’ response latency. Based on these results, we provide recommendations and suggest directions for future studies. We advocate for the development of shared longitudinal diverse corpora with turn-by-turn data and comprehensive assessment of infants' linguistic and social development. We also suggest more explicit definition and testing of computational models in order to inform the cognitive mechanisms underlying turn-taking.

List of References

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959–975.

Azur, M. J., Stuart, E. A., Frangakis, C., & Leaf, P. J. (2011). Multiple imputation by chained equations: What is it and how does it work? *International Journal of Methods in Psychiatric Research*, *20*(1), 40–49.

Bateson, M. C. (1975). Mother-Infant exchanges: The epigenesis of conversational interaction. *Annals of the New York Academy of Sciences*, *263*(1), 101–113.

Beebe, B., Alson, D., Jaffe, J., Feldstein, S., & Crown, C. (1988). Vocal congruence in mother infant play. *Journal of Psycholinguistic Research*, *17*(3), 245–259. https://doi.org/10.1007/BF01686358

Belsky, J. (1980). Mother-infant interaction at home and in the laboratory: A comparative study. *The Journal of Genetic Psychology*, *137*(1), 37–47.

Bergey, C., Marshall, Z., DeDeo, S., & Yurovsky, D. (2021). *Learning communicative acts in children's conversations: A Hidden Topic Markov Model analysis of the CHILDES corpus*.

Bigelow, A. E. (1998). Infants' sensitivity to familiar imperfect contigencies in social interaction. *Infant Behavior and Development*, *21*(1), 149–162.

Buuren, S. van, & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*(1), 1–67. https://doi.org/10.18637/jss.v045.i03

Cappella, J. N. (1981). Mutual influence in expressive behavior: Adult–adult and infant–adult dyadic interaction. *Psychological Bulletin*, *89*(1), 101.

Cappella, J. N. (1996). Dynamic coordination of vocal and kinesic behavior in dyadic interaction: Methods, problems, and interpersonal outcomes. *In J. H. Watt & C. A. VanLear (Eds.), Dynamic Patterns in Communication Processes (353–386). Sage Publications, Inc.*

Carvalho, M. E. S., Justo, J. M. R. M., Gratier, M., Tome, T., Pereira, E., & Rodrigues, H. (2019). Vocal responsiveness of preterm infants to maternal infant-directed speaking and singing during skin-to-skin contact (Kangaroo Care) in the NICU. *Infant Behavior and Development*, *57*. https://doi.org/10.1016/j.infbeh.2019.101332

Casillas, M., Bobb, S. C., & Clark, E. V. (2016). Turn-taking, timing, and planning in early language acquisition. *Journal of Child Language*, *43*(6), 1310–1337. https://doi.org/10.1017/S0305000915000689

Casillas, M., & Frank, M. (2013). The development of predictive processes in children's discourse understanding. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *35*(35).

Casillas, M., & Scaff, C. (2021). Analyzing contingent interactions in R with `chattr`. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*(43). https://escholarship.org/uc/item/4rr848x0

Choi, J., Lee, Y., Choi, J., & Lee, Y. (2013). Conversational turn-taking and topic manipulation skills of children with high-functioning autism spectrum disorders. *Communication Sciences & Disorders*, *18*(1), 12–23.

Christiansen, M., Contreras, P., & Trecca, F. (in press). Towards A Comparative Approach to Language Acquisition. *Current Directions in Psychological Science*.

Clark, E. V., & Lindsey, K. L. (2015). Turn-taking: A case study of early gesture and word use in answering WHERE and WHICH questions. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00890

Cox, C. M. M., Keren-Portnoy, T., Roepstorff, A., & Fusaroli, R. (2021). A Bayesian meta-analysis of infants' ability to perceive audio–visual congruence for speech. *Infancy*. https://doi.org/10.1111/infa.12436

Cross, L., Michael, J., Wilsdon, L., Henson, A., & Atherton, G. (2020). Still want to help? Interpersonal coordination's effects on helping behaviour after a 24 hour delay. *Acta Psychologica*, *206*, 103062. https://doi.org/10.1016/j.actpsy.2020.103062

Crown, C. L. (1991). Coordinated interpersonal timing of vision and voice as a function of interpersonal attraction. *Journal of Language and Social Psychology*, *10*(1), 29–46.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, *1695*(5), 1–9.

Cummins, F. (2013). Towards an enactive account of action: Speaking and joint speaking as exemplary domains. *Adaptive Behavior*, *21*(3), 178–186.

Curioni, A., Vesper, C., Knoblich, G., & Sebanz, N. (2019). Reciprocal information flow and role distribution support joint action coordination. *Cognition*, *187*, 21–31. https://doi.org/10.1016/j.cognition.2019.02.006

Dahlby, M., Irmalm, L., Kytöharju, S., Wallander, L., Zachariassen, H., Ericsson, A., & Marklund, U. (2011). Parent-child interaction: Relationship between pause duration and infant vocabulary at 18 months. *Fonetik 2011*.

Dale, R. (2015). An Integrative Research Strategy for Exploring Synergies in Natural Language Performance. *Ecological Psychology*, *27*(3), 190–201. https://doi.org/10.1080/10407413.2015.1068649

Dale, R., Fusaroli, R., Duran, N. D., & Richardson, D. C. (2013). The Self-Organization of Human Interaction. *Psychology of Learning and Motivation*, *59*, 43–95. https://doi.org/10.1016/B978-0-12-407187-2.00002-2

de Reus, K., Masayo, S., Marianna, A., Gamba, M., de Heer Kloots, M., Miriam, L., Bruno, J. H., Laurel, T., & Ravignani, A. (2021). Rhythm in dyadic interactions. *PsyArXiv*. doi:10.31234/osf.io/9yrkv

De Vos, C., Casillas, M., Crasborn, O., & Levinson, S. C. (2015). Is turn timing dependent on language modality? *The 36th TABU Dag*.

Deffner, D., Rohrer, J. M., & McElreath, R. (2021). *A Causal Framework for Cross-Cultural Generalizability*. PsyArXiv. https://doi.org/10.31234/osf.io/fqukp

Demartsev, V., Strandburg-Peshkin, A., Ruffner, M., & Manser, M. (2018). Vocal turn-taking in meerkat group calling sessions. *Current Biology*, *28*(22), 3661-3666. e3.

Dideriksen, C., Christiansen, M. H., Tylén, K., Dingemanse, M., & Fusaroli, R. (2020). *Quantifying the interplay of conversational devices in building mutual understanding*. PsyArXiv. https://doi.org/10.31234/osf.io/a5r74

Dominguez, S., Devouche, E., Apter, G., & Gratier, M. (2016). The roots of turn-taking in the neonatal period. *Infant and Child Development*, *25*(3), 240–255.

Dressler, R. A., Buder, E. H., & Cannito, M. P. (2009). Rhythmic patterns during conversational repairs in speakers with aphasia. *Aphasiology*, *23*(6), 731–748. https://doi.org/10.1080/02687030802165582

Duran, N. D., & Fusaroli, R. (2017). Conversing with a devil's advocate: Interpersonal coordination in deception and disagreement. *PLOS ONE*, *12*(6), e0178140. https://doi.org/10.1371/journal.pone.0178140

Duran, N. D., Paxton, A., & Fusaroli, R. (2019). ALIGN: Analyzing linguistic interactions with generalizable techNiques-A Python library. *Psychological Methods*, *24*(4), 419–438. https://doi.org/10.1037/met0000206

Fernández-Castilla, B., Jamshidi, L., Declercq, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. (2020). The application of meta-analytic (multi-level) models with multiple random effects: A systematic review. *Behavior Research Methods*, *52*(5).

Ferrer, E., & Helm, J. L. (2013). Dynamical systems modeling of physiological coregulation in dyadic interactions. *International Journal of Psychophysiology*, *88*(3), 296–308.

Fusaroli, R., Bjørndahl, J. S., Roepstorff, A., & Tylén, K. (2016). A heart for interaction: Shared physiological dynamics and behavioral coordination in a collective, creative construction task. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(9), 1297.

Fusaroli, R., Grossman, R., Bilenberg, N., Cantio, C., Jepsen, J. R. M., & Weed, E. (2022). Toward a cumulative science of vocal markers of autism: A cross-linguistic meta-analysis-based investigation of acoustic markers in American and Danish autistic children. *Autism Research*. https://doi.org/10.1002/aur.2661

Fusaroli, R., Konvalinka, I., & Wallot, S. (2014). Analyzing Social Interactions: The Promises and Challenges of Using Cross Recurrence Quantification Analysis. *Springer Proceedings in Mathematics & Statistics*, *103*, 137–155. https://doi.org/10.1007/978-3-319-09531-8_9

Fusaroli, R., Rączaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, *32*, 147–157.

Fusaroli, R., & Tylén, K. (2016). Investigating conversational dynamics: Interactive alignment, Interpersonal synergy, and collective task performance. *Cognitive Science*, *40*(1), 145–171.

Fusaroli, R., Weed, E., Fein, D., & Naigles, L. (2019). Hearing me hearing you: Reciprocal effects between child and parent language in autism and typical development. *Cognition*, *183*, 1–18. https://doi.org/10.1016/j.cognition.2018.10.022

Ginsburg, G. P., & Kilbourne, B. K. (1988). Emergence of vocal alternation in mother-infant interchanges. *Journal of Child Language*, *15*(2), 221–235.

Gratier, M. (2003). Expressive timing and interactional synchrony between mothers and infants: Cultural similarities, cultural differences, and the immigration experience. *Cognitive Development*, *18*(4), 533–554.

Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., & Parlato-Oliveira, E. (2015). Early development of turn-taking in vocal interaction between mothers and infants. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpgyg.2015.01167

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*(4), 274–279. https://doi.org/10.1111/1467-9280.00255

Guest, O., & Martin, A. E. (2021). How Computational Modeling Can Force Theory Building in Psychological Science. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *16*(4), 789–802. https://doi.org/10.1177/1745691620970585

Heeman, P. A., Lunsford, R., Selfridge, E., Black, L., & van Santen, J. (2010). Autism and interactional aspects of dialogue. *Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 249–252.

Heinemann, T. (2010). The question–response system of Danish. *Journal of Pragmatics*, *42*(10), 2703–2725. https://doi.org/10.1016/j.pragma.2010.04.007

Hilbrink, E. E., Gattis, M., & Levinson, S. C. (2015). Early developmental changes in the timing of turn-taking: A longitudinal study of mother-infant interaction. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01492

Hsu, H.-C., & Fogel, A. (2003). Stability and Transitions in Mother-Infant Face-to-Face Communication During the First 6 Months: A Microhistorical Approach. *Developmental Psychology*, *39*(6), 1061–1082. https://doi.org/10.1037/0012-1649.39.6.1061

Hsu, H.-C., Fogel, A., & Messinger, D. S. (2001). Infant non-distress vocalization during mother-infant face-to-face interaction: Factors associated with quantitative and qualitative differences. *Infant Behavior & Development*, *24*(1), 107–128. https://doi.org/10.1016/S0163-6383(01)00061-3

Indefrey, P. (2011). The Spatial and Temporal Signatures of Word Production Components: A Critical Update. *Frontiers in Psychology*, *2*, 255. https://doi.org/10.3389/fpsyg.2011.00255

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1–2), 101–144. https://doi.org/10.1016/j.cognition.2002.06.001

Jaffe, J., Beebe, B., Feldstein, S., Crown, C., & Jasnow, M. (2001). *Rhythms of dialogue in infancy: Coordinated timing in development* (Vol. 66).

Jasnow, M., & Feldstein, S. (1986). Adult-like temporal characteristics of mother-infant vocal interactions. *Child Development*, *57*(3), 754–761.

Jylänki, P., Vanhatalo, J., & Vehtari, A. (2011). Robust Gaussian Process Regression with a Student-t Likelihood. *Journal of Machine Learning Research*, *12*(11), 31.

Kaye, K. (1977). Toward the origin of dialogue. *Studies in Mother–Infant Interaction*, 89–117.

Kelly, E. (1994). Speech rates and turn taking behaviors of children who stutter and their fathers. *Journal of Speech and Hearing Research*, *37*(6), 1284–1294. https://doi.org/10.1044/jshr.3706.1284

Kelly, E., & Conture, E. (1992). Speaking rates, response-time latencies and interrupting behaviors of young stutterers, non-stutterers, and their mothers. *Journal of Speech and Hearing Research*, *35*(6), 1256–1267. https://doi.org/10.1044/jshr.3506.1256

Klann-Delius, G., & Hofmeister, C. (1997). The development of communicative competence of securely and insecurely attached children in interactions with their mothers. *Journal of Psycholinguistic Research*, *26*(1), 69–88. https://doi.org/10.1023/a:1025012221407

Kondaurova, M. V., Smith, N. A., Zheng, Q., Reed, J., & Fagan, M. K. (2020). Vocal Turn-Taking Between Mothers and Their Children With Cochlear Implants. *Ear and Hearing*, *41*(2), 362–373. https://doi.org/10.1097/AUD.0000000000000769

Konvalinka, I., Xygalatas, D., Bulbulia, J., Schjødt, U., Jegindø, E.-M., Wallot, S., Orden, G. V., & Roepstorff, A. (2011). Synchronized arousal between performers and related spectators in a fire-walking ritual. *Proceedings of the National Academy of Sciences*, *108*(20), 8514–8519. https://doi.org/10.1073/pnas.1016955108

Lange, A. E. (2016). *Vocal Rhythm Coordination and Preterm Infants: Rhythms of Dialogue in a High-Risk NICU Sample* (Vol. 1). City University of New York.

Leonardi, G., Nomikou, I., Rohlfing, K. J., & Raczaszek-Leonardi, J. (2016). Vocal Interactions at the Dawn of Communication: The Emergence of Mutuality and Complementarity in Mother-Infant Interaction. *Joint IEEE International Conference on Development and Learning and Epigenetic Robotics ICDL-EpiRob*, 288–293.

Levinson, S. (1995, March). *Interactional biases in human thinking*. Social Intelligence and Interaction: Expressions and Implications of the Social Bias in Human Intelligence; Cambridge University Press. https://doi.org/10.1017/CBO9780511621710.014

Levinson, S. (2006). Cognition at the heart of human interaction. *Discourse Studies*, *8*(1), 85–93. https://doi.org/10.1177/1461445606059557

Levinson, S. (2013). Action Formation and Ascription. In *The Handbook of Conversation Analysis* (pp. 101–130). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118325001.ch6

Levinson, S. (2016). Turn-taking in Human Communication—Origins and Implications for Language Processing. *Trends in Cognitive Sciences*, *20*(1), 6–14. https://doi.org/10.1016/j.tics.2015.10.010

Levinson, S., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, *6*, 731. https://doi.org/10.3389/fpsyg.2015.00731

Lewedag, V., Oller, D., & Lynch, M. (1994). Infants' vocalization patterns across home and laboratory environments. *First Language*, *14*(42–43). https://journals.sagepub.com/doi/10.1177/014272379401404204

Lewis, M., Mathur, M., VanderWeele, T., & Frank, M. C. (2020). *The puzzling relationship between multi-lab replications and meta-analyses of the rest of the literature*. PsyArXiv. https://doi.org/10.31234/osf.io/pbrdk

Lindsay, L., Gambi, C., & Rabagliati, H. (2019). Preschoolers Optimize the Timing of Their Conversational Turns Through Flexible Coordination of Language Comprehension and Production. *Psychological Science*, *30*(4), 504–515. https://doi.org/10.1177/0956797618822802

Lopez, L., Walle, E., Pretzer, G., & Warlaumont, A. (2020). Adult responses to infant prelinguistic vocalizations are associated with infant vocabulary: A home observation study. *PloS One*, *15*(11), e0242232. https://doi.org/10.1371/journal.pone.0242232

Malin, D., Ludvig, I., Satu, K., Linnea, W., Helena, Z., Anna, E., & Ulrika, M. (2011). Parent-child interaction: Relationship between pause duration and infant vocabulary at 18 months. In *Fonetik 2011* (Vol. 1, Issue 1, pp. 101–104). Proceedings from Fonetik 2011: Speech, Music and Hearing; Quarterly Progress and Status Report.

Manson, J. H., Bryant, G. A., Gervais, M. M., & Kline, M. A. (2013). Convergence of speech rate in conversation predicts cooperation. *Evolution and Human Behavior*, *34*(6), 419–426. https://doi.org/10.1016/j.evolhumbehav.2013.08.001

Marklund, U., Marklund, E., Lacerda, F., & Schwarz, I.-C. (2015). Pause and utterance duration in child-directed speech in relation to child vocabulary size. *Journal of Child Language*, *42*(5), 1158–1171. https://doi.org/10.1017/S0305000914000609

McFarland, D. H. (2001). Respiratory markers of conversational interaction. *Journal of Speech, Language, and Hearing Research*, *44*(1), 128–143. https://doi.org/10.1044/1092-4388(2001/012)

Méndez-Cárdenas, M. G., & Zimmermann, E. (2009). Duetting—A mechanism to strengthen pair bonds in a dispersed pair-living primate (Lepilemur edwardsi)? *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, *139*(4), 523–532.

Naigles, L. R., & Fein, D. (2017). Looking through their eyes: Tracking early language comprehension in ASD. In *Innovative investigations of language in autism spectrum disorder* (pp. 49–69). Walter de Gruyter GmbH. https://doi.org/10.1037/15964-004

Navarro, D. J. (2021). If Mathematical Psychology Did Not Exist We Might Need to Invent It: A Comment on Theory Building in Psychology. *Perspectives on Psychological Science*, *16*(4), 707–716. https://doi.org/10.1177/1745691620974769

Newman, L., & Smit, A. (1989). Some effects of variations in response-time latency on speech rate, interruptions, and fluency in children's speech. *Journal of Speech and Hearing Research*, *32*(3), 635–644. https://doi.org/10.1044/jshr.3203.635

Nikolaus, M., Maes, J., Auguste, J., Prévot, L., & Fourtassi, A. (2021). *Large-scale study of speech acts' development using automatic labelling*. PsyArXiv. https://doi.org/10.31234/osf.io/j4smd

Northrup, J. B., & Iverson, J. M. (2015). Vocal Coordination During Early Parent-Infant Interactions Predicts Language Outcome in Infant Siblings of Children with Autism Spectrum Disorder. In *Infancy* (Vol. 20, Issue 5, pp. 523–547). WILEY. https://doi.org/10.1111/infa.12090

O' Dell, M., Nieminen, T., & Lennes, M. (2012). Modeling Turn-Taking Rhythms with Oscillators. *Linguistica Uralica*, *48*(3), 218–227. https://doi.org/10.3176/lu.2012.3.08

Ochi, K., Ono, N., Owada, K., Kojima, M., Kuroda, M., Sagayama, S., & Yamasue, H. (2019). Quantification of speech and synchrony in the conversation of adults with autism spectrum disorder. *PLOS ONE*, *14*(12), e0225377. https://doi.org/10.1371/journal.pone.0225377

Ochs, E., Kremer-Sadlik, T., Sirota, K. G., & Solomon, O. (2004). Autism and the Social World: An Anthropological Perspective. *Discourse Studies*, *6*(2), 147–183. https://doi.org/10.1177/1461445604041766

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. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Pezzulo, G., Donnarumma, F., Dindo, H., D'Ausilio, A., Konvalinka, I., & Castelfranchi, C. (2019). The body talks: Sensorimotor communication and its brain and kinematic signatures. *Physics of Life Reviews*, *28*, 1–21. https://doi.org/10.1016/j.plrev.2018.06.014

Pickering, M., & Garrod, S. (2021). *Understanding Dialogue*. Cambridge University Publisher. https://www.cambridge.org/us/academic/subjects/languages-linguistics/psycholinguistics-and-neurolinguistics/understanding-dialogue-language-use-and-social-interaction, https://www.cambridge.org/us/academic/subjects/languages-linguistics/psycholinguistics-and-neurolinguistics

Pika, S., Wilkinson, R., Kendrick, K. H., & Vernes, S. C. (2018). Taking turns: Bridging the gap between human and animal communication. *Proceedings of the Royal Society B*, *285*(1880), 20180598.

Rast, P., & Ferrer, E. (2018). A Mixed-Effects Location Scale Model for Dyadic Interactions. *Multivariate Behavioral Research*, *53*(5), 756–775. https://doi.org/10.1080/00273171.2018.1477577

Reissland, N., & Stephenson, T. (1999). Turn-taking in early vocal interaction: A comparison of premature and term infants' vocal interaction with their mothers. *Child: Care, Health and Development*, *25*(6), 447–456. https://doi.org/10.1046/j.1365-2214.1999.00109.x

Rich, P., Haan, R. de, Wareham, T., & Rooij, I. van. (2021). *How hard is cognitive science?* PsyArXiv. https://doi.org/10.31234/osf.io/k79nv

Riest, C., Jorschick, A. B., & de Ruiter, J. P. (2015). Anticipation in turn-taking: Mechanisms and information sources. *Frontiers in Psychology*, *6*, 89. https://doi.org/10.3389/fpsyg.2015.00089

Rochet-Capellan, A., & Fuchs, S. (2014). Take a breath and take the turn: How breathing meets turns in spontaneous dialogue. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1658), 20130399. https://doi.org/10.1098/rstb.2013.0399

Ruiter, J., Mitterer, H., & Enfield, N. (2006). Projecting the End of a Speaker's Turn: A Cognitive Cornerstone of Conversation. *Language*, *82*(3), 515–535. https://doi.org/doi:10.1353/lan.2006.0130

Sacks, H., Schegloff, E. A., & Jefferson, G. (1978). A Simplest Systematics for the Organization of Turn Taking for Conversation. In J. Schenkein (Ed.), *Studies in the Organization of Conversational Interaction* (pp. 7–55). Academic Press. https://doi.org/10.1016/B978-0-12-623550-0.50008-2

Savelkoul, E. M., Zebrowski, P. M., Feldstein, S., & Cole-Harding, S. (2007). Coordinated interpersonal timing in the conversations of children who stutter and their mothers and fathers. *Journal of Fluency Disorders*, *32*(1), 1–32. https://doi.org/10.1016/j.jfludis.2006.12.001

Sawyer, J., Matteson, C., Ou, H., & Nagase, T. (2017). The Effects of Parent-Focused Slow Relaxed Speech Intervention on Articulation Rate, Response Time Latency, and Fluency in Preschool Children Who Stutter. In *JOURNAL OF SPEECH LANGUAGE AND HEARING RESEARCH* (Vol. 60, Issue 4, pp. 794–809). AMER SPEECH-LANGUAGE-HEARING ASSOC. https://doi.org/10.1044/2016_JSLHR-S-16-0002

Snow, C. E. (1977). The development of conversation between mothers and babies*. *Journal of Child Language*, *4*(1), 1–22. https://doi.org/10.1017/S0305000900000453

Stern, D., Beebe, B., Jaffe, J., & Bennett, S. (1977). The infant's stimulus world during social interaction: A study of caregiver behaviors with particular reference to repetition and timing. In *Studies in mother-infant interaction*.

Sterne, J. A. C., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M., & Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *BMJ (Clinical Research Ed.)*, *338*, b2393. https://doi.org/10.1136/bmj.b2393

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., De Ruiter, J. P., & Yoon, K.-E. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587–10592.

Striano, T., Chen, X., Cleveland, A., & Bradshaw, S. (2006). Joint attention social cues influence infant learning. *European Journal of Developmental Psychology*, *3*(3), 289–299. https://doi.org/10.1080/17405620600879779

Takahashi, D. Y., Fenley, A. R., & Ghazanfar, A. A. (2016). Early development of turn-taking with parents shapes vocal acoustics in infant marmoset monkeys. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *371*(1693), 20150370. https://doi.org/10.1098/rstb.2015.0370

Takahashi, D. Y., Narayanan, D. Z., & Ghazanfar, A. A. (2013). Coupled oscillator dynamics of vocal turn-taking in monkeys. *Current Biology: CB*, *23*(21), 2162–2168. https://doi.org/10.1016/j.cub.2013.09.005

Thorgrimsson, G. B., Fawcett, C., & Liszkowski, U. (2014). Infants' expectations about gestures and actions in third-party interactions. *Frontiers in Psychology*, *5*, 321. https://doi.org/10.3389/fpsyg.2014.00321

Todt, D., & Hultsch, H. (1999). How nightingales develop their vocal competence. *Proc. 22nd Int. Ornithological Congress*, 193–215.

Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Harvard University Press. https://doi.org/10.2307/j.ctvjsf4jc

Traag, V. A., & Bruggeman, J. (2009). Community detection in networks with positive and negative links. *Physical Review E*, *80*(3), 036115. https://doi.org/10.1103/PhysRevE.80.036115

Trecca, F., Tylén, K., Højen, A., & Christiansen, M. (2021). The puzzle of Danish: Implications for language learning and use. *Language Acquisition*.

Waade, P. T., Enevoldsen, K. C., Simonsen, A., & Fusaroli, R. (2020). *Introducing tomsup: Theory of Mind Simulations Using Python*. PsyArXiv. https://doi.org/10.31234/osf.io/zcgkv

Warlaumont, A., Oller, D., Dale, R., Richards, J. A., Gilkerson, J., & Xu, D. (2010). Vocal interaction dynamics of children with and without autism. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, 121–126.

Warlaumont, A., Richards, J., Gilkerson, J., & Oller, D. (2014). A social feedback loop for speech development and its reduction in autism. *Psychological Science*, *25*(7), 1314–1324. https://doi.org/10.1177/0956797614531023

Weed, E., Fusaroli, R., Tranbjerg, J., Fein, D., & Naigles, L. (2017, May 10). *Conversational Dynamics in a Longitudinal Corpus of Caregiver-Child Interactions*.

Welkowitz, J., Bond, R., Feldman, L., & Tota, M. (1990). Conversational time patterns and mutual influence in parent-child interactions—A time-series approach. *Journal of Psycholinguistic Research*, *19*(4), 221–243. https://doi.org/10.1007/BF01077258

Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review*, *12*(6), 957–968. https://doi.org/10.3758/bf03206432

Yaruss, J. (1997). Utterance timing and childhood stuttering. *Journal of Fluency Disorders*, *22*(4), 263–286. https://doi.org/10.1016/S0094-730X(97)00023-5

Zlochower, A., & Cohn, J. (1996). Vocal timing in face-to-face interaction of clinically depressed and nondepressed mothers and their 4-month-old infants. *Infant Behavior and Development*, *19*(3), 371–374. https://doi.org/10.1016/S0163-6383(96)90035-1

Supplementary Materials

S1. Additional information on systematic search and included studies

S1.1 PRISMA Flow Chart & Check Lists

For the systematic search we followed as much as possible the PRISMA Guidelines (see the flow diagram in Figure S1.1 and the checklists in Table S1.1).

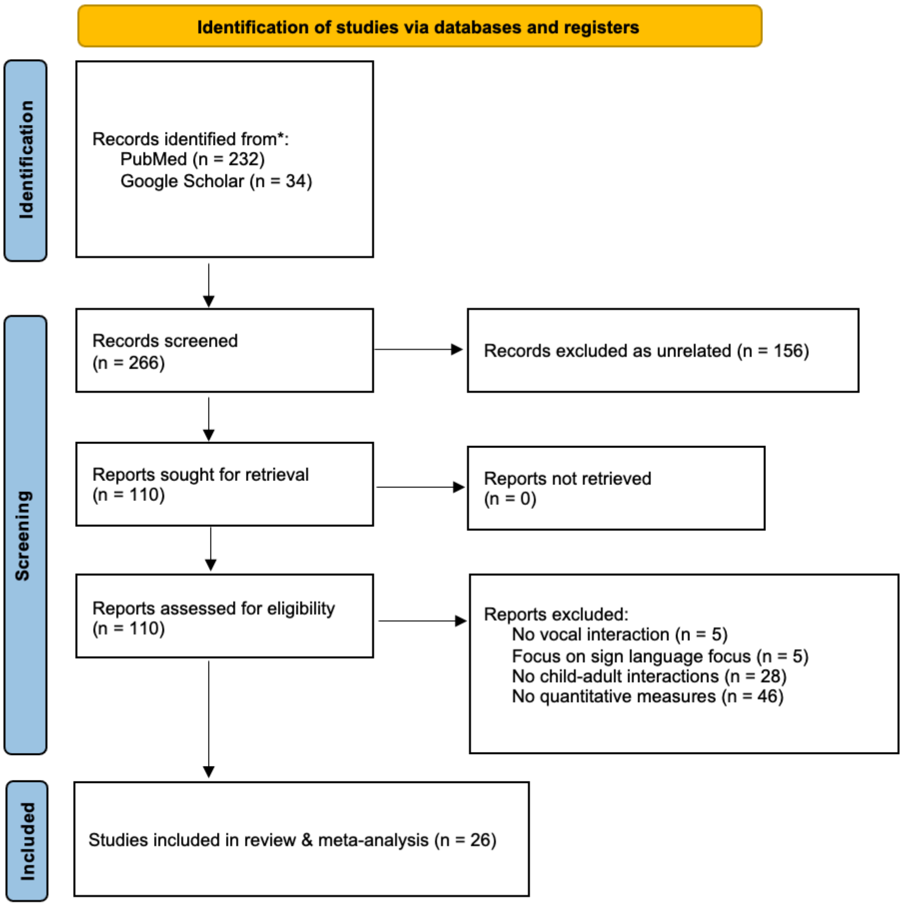


Figure S1.1 – Prisma flow chart of the systematic review

Table S1.1.1 – PRISMA Checklist

Topic	**No.**	**Item**	**Location where item is reported**
TITLE			
Title	1	Identify the report as a systematic review.	Page 1
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist	
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Page 9, Section 1.2
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Page 9, Section 1.2
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Page 14-15, Section 2.1
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Page 14-15, Section 2.1
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Page 14-15, Section 2.1

Topic	No.	Item	Location where item is reported
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Page 14-15, Section 2.1
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Page 14-15, Section 2.1
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Page 18-19, Section 2.3
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Page 18-19, Section 2.3
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Page 14-15, Section 2.1 & S1, Supplementary Materials
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	Page 18-19, Section 2.3

Topic	**No.**	**Item**	**Location where item is reported**
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item 5)).	Page 21-22, Section 2.4
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	Page 19, Section 2.3
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Page 20, Section 2.4
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Page 20-23, Section 2.4
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	Page 20-23, Section 2.4
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	S3-S10 in Supplementary Materials
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	S3-S10 in Supplementary Materials
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Page 20-23, Section 2.4
RESULTS			

Topic	**No.**	**Item**	**Location where item is reported**
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Page 14-15, Section 2.1 & S1 in Supplementary Materials
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	NA
Study characteristics	17	Cite each included study and present its characteristics.	Table S1.2 in Supplementary Materials
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	NA
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Table S1.2 in Supplementary Materials
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Discussion in S1 of Supplementary Materials
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	S4-S9 in Supplementary Materials
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	S4-S9 in Supplementary Materials

Topic	**No.**	**Item**	**Location where item is reported**
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	S4-S9 and S10 in Supplementary Materials
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	Discussion in S1 of Supplementary Materials
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	S4-S9 in Supplementary Materials
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Page 30-31, Section 4.1
	23b	Discuss any limitations of the evidence included in the review.	Page 35, Section 4.3
	23c	Discuss any limitations of the review processes used.	S1 of Supplementary Materials
	23d	Discuss implications of the results for practice, policy, and future research.	Page 35-37, Section 4.4
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	NA
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Page 14, Section 2.1

Topic	No.	Item	Location where item is reported
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	NA
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	Page 1
Competing interests	26	Declare any competing interests of review authors.	Page 1
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Page 1

Table S1.1.2 – PRISMA Abstract Checklist

Topic	No.	Item	Reported?
TITLE			
Title	1	Identify the report as a systematic review.	Yes
BACKGROUND			
Objectives	2	Provide an explicit statement of the main objective(s) or question(s) the review addresses.	Yes
METHODS			
Eligibility criteria	3	Specify the inclusion and exclusion criteria for the review.	No, but see Page 14-15, Section 2.1

Topic	No.	Item	Reported?
Information sources	4	Specify the information sources (e.g. databases, registers) used to identify studies and the date when each was last searched.	No, but see Page 14-15, Section 2.1
Risk of bias	5	Specify the methods used to assess risk of bias in the included studies.	No, but see S1 in Supplementary Martials
Synthesis of results	6	Specify the methods used to present and synthesize results.	Yes
RESULTS			
Included studies	7	Give the total number of included studies and participants and summarise relevant characteristics of studies.	Yes
Synthesis of results	8	Present results for main outcomes, preferably indicating the number of included studies and participants for each. If meta-analysis was done, report the summary estimate and confidence/credible interval. If comparing groups, indicate the direction of the effect (i.e. which group is favoured).	Yes
DISCUSSION			
Limitations of evidence	9	Provide a brief summary of the limitations of the evidence included in the review (e.g. study risk of bias, inconsistency and imprecision).	Yes
Interpretation	10	Provide a general interpretation of the results and important implications.	Yes
OTHER			
Funding	11	Specify the primary source of funding for the review.	Yes
Registration	12	Provide the register name and registration number.	Yes

S1.2 Information about Included Studies

Table S1.2 – Detailed information about the included studies. Activity indicates Free Play (P), structured questions (Q), focused conversation (C), and structured activity (A). The raw data are available on OSF (https://osf.io/wkceb/) and will be made available on MetaLab (https://langcog.github.io/metalab/) upon publication of the manuscript.

Article	**Population**	**Interactor**	**Sex (proportion male)**	**Environment**	**Activity**	**Native language (dialect)**	**Number of dyads**	**Mean age infants**	**Infant RL**	**Adult RL**
Hilbrink, Gattis & Levinson (2015)	Typical	Caregiver	0.583	Unfamiliar	P	EN (US)	12	3	624	639
	Typical	Caregiver	0.583	Unfamiliar	P	EN (US)	12	4	623.5	675
	Typical	Caregiver	0.583	Unfamiliar	P	EN (US)	12	5	865.5	797.5
	Typical	Caregiver	0.583	Unfamiliar	P	EN (US)	12	9	1919.75	725
	Typical	Caregiver	0.583	Unfamiliar	P	EN (US)	12	12	1243.5	513.5
	Typical	Caregiver	0.583	Unfamiliar	P	EN (US)	12	18	877.5	640.2
Jaffe et al. (2001)	Typical	Caregiver	0.466	Familiar	P	EN (US)	84	4	843	917
	Typical	Stranger	0.466	Familiar	P	EN (US)	82	4	666	730

	Typical	Caregiver	0.466	Unfamiliar	P	EN (US)	52	4	780	850
	Typical	Stranger	0.466	Unfamiliar	P	EN (US)	51	4	540	730
Clark & Lindsey (2015)	Typical	Caregiver	0	Familiar	Q	EN (US)	1	28.5	800	
Carvalho et al., 2019	Preterm	Caregiver	0.44	Unfamiliar	C	PT	36	1	866	
Casillas, Bob & Clark (2016)	Typical	Caregiver	0.6	Familiar	Q	EN (US)	5	20	867	299
	Typical	Caregiver	0.6	Familiar	Q	EN (US)	5	24	684	293
	Typical	Caregiver	0.6	Familiar	Q	EN (US)	5	28	686	344
	Typical	Caregiver	0.6	Familiar	Q	EN (US)	5	32	525	475
	Typical	Caregiver	0.6	Familiar	Q	EN (US)	5	36	571	376
	Typical	Caregiver	0.6	Familiar	Q	EN (US)	5	39	523	486
Bateson (1975)	Typical	Caregiver	0	Familiar	C	EN (US)	1	2.5	1370	1430
Kelly & Conture (1992)	Typical	Caregiver	0	Unfamiliar	P	EN (US)	13	48	774	1352
	Stutterer	Caregiver	0	Unfamiliar	P	EN (US)	13	48	798	1268
Savelkoul, Zebrowski, Feldstein & Cole-Harding (2007)	Typical	Caregiver	NA	Familiar	P	EN (US)	10	70	1002	1064
	Typical	Caregiver	NA	Familiar	P	EN (US)	10	70	1113	1189
	Stutterer	Caregiver	NA	Familiar	P	EN (US)	10	70	837	881
	Stutterer	Caregiver	NA	Familiar	P	EN (US)	10	70	904	972

Kelly (1994)	Typical	Caregiver	0	Unfamiliar	P	EN (US)	11	61	880	570
	Stutterer	Caregiver	0	Unfamiliar	P	EN (US)	11	61	1050	720
Beebe (1988)	Typical	Caregiver	1	Unfamiliar	P	EN (US)	15	3.75	801	703
Jasnow & Feldstein (1986)	Typical	Caregiver	0.55	Unfamiliar	P	EN (US)	29	9	875	775
Welkowitz, Bond, Feldman & Tota (1990)	Typical	Caregiver	0.5	Familiar	C	EN (US)	20	60	1230	1090
	Typical	Caregiver	0.5	Familiar	C	EN (US)	20	60	1230	1000
Newman & Smit (1989)	Typical	Stranger	0	Familiar	C	EN (US)	2	48	2940	3000
	Typical	Stranger	0	Familiar	C	EN (US)	2	48	740	1000
	Typical	Stranger	0	Familiar	C	EN (US)	2	48	3666	3000
	Typical	Stranger	0.5	Familiar	C	EN (US)	2	48	560	1000
	Typical	Stranger	0.5	Familiar	C	EN (US)	2	48	3160	3000
	Typical	Stranger	0.5	Familiar	C	EN (US)	2	48	590	1000
Marklund, Marklund, Lacerda & Schwarz (2015)	Typical	Caregiver	0.533	Familiar	A	SE	5	30		736.67
Klann-Delius & Hofmeister (1997)	Typical	Caregiver	0.44	Unfamiliar	P	DE	17	17	3483	
	Typical	Caregiver	0.44	Unfamiliar	P	DE	17	23	2395	
	Typical	Caregiver	0.44	Unfamiliar	P	DE	17	30	1638	
	Typical	Caregiver	0.44	Unfamiliar	P	DE	17	36	1970	
	Typical	Caregiver	0.44	Unfamiliar	P	DE	17	17	2089	
	Typical	Caregiver	0.44	Unfamiliar	P	DE	17	23	2027	

	Typical	Caregiver	0.44	Unfamiliar	P	DE	17	30	1675	
	Typical	Caregiver	0.44	Unfamiliar	P	DE	17	36	1176	
	Insecure Attachment	Caregiver	0.44	Unfamiliar	P	DE	8	17	3696	
	Insecure Attachment	Caregiver	0.44	Unfamiliar	P	DE	8	23	3718	
	Insecure Attachment	Caregiver	0.44	Unfamiliar	P	DE	8	30	2618	
	Insecure Attachment	Caregiver	0.44	Unfamiliar	P	DE	8	36	3244	
	Insecure Attachment	Caregiver	0.44	Unfamiliar	P	DE	8	17	2735	
	Insecure Attachment	Caregiver	0.44	Unfamiliar	P	DE	8	23	2105	
	Insecure Attachment	Caregiver	0.44	Unfamiliar	P	DE	8	30	1340	
	Insecure Attachment	Caregiver	0.44	Unfamiliar	P	DE	8	36	1497	
Leonardi, Nomikou, Rohlfing & Raczaszek-Leonardi (2016)	Typical	Caregiver	NA	Familiar	A	DE	12	3	755	564
	Typical	Caregiver	NA	Familiar	A	DE	12	6	709	528
	Typical	Caregiver	NA	Familiar	A	DE	12	8	811	613
Kondaurova, Smith, Zheng, Reed & Fagan (2020)	Typical	Caregiver	0.4166	Unfamiliar	C	EN (US)	12	18.3	778	520

	Typical	Caregiver	0.4166	Unfamiliar	C	EN (US)	12	27.5	722	510
	Cochlear	Caregiver	0.166	Unfamiliar	C	EN (US)	12	18.2	921	690
	Cochlear	Caregiver	0.166	Unfamiliar	C	EN (US)	12	27.4	983	510
Dahlby et al., 2011	Typical	Caregiver	0.4	Familiar	A	SE	5	18	920	630
	Typical	Caregiver	0.4	Familiar	A	SE	5	18	740	480
Northrup & Iverson (2015)	Typical	Caregiver	0.5	Familiar	P	EN (US)	10	9	2210	1860
	Autism	Caregiver	0.48	Familiar	P	EN (US)	25	9	2940	2650
Yaruss (1997)	Stutterer	Caregiver	0	Unfamiliar	P	EN (US)	12	55.17	820	
Warlaumont et al. (2014)	Typical	Stranger	0.154	Familiar	C	EN (US)	78	30.2	2370	1650
	Autism	Stranger	0.154	Familiar	C	EN (US)	26	30.2	2700	2320
Sawyer, Matteson, Ou & Nagase (2017)	Stutterer	Caregiver	0.118	Unfamiliar	P	EN (US)	17	47.9	770	520
Reissland & Stephenson (1999)	Typical	Caregiver	0.5	Familiar	A	EN (UK)	8	10	3370	1770
	Preterm	Caregiver	0.4	Familiar	A	EN (UK)	5	7.5	2580	3580
Lange (2016)	Preterm	Caregiver	0.5	Unfamiliar	C	EN (US)	15	4	1200	1120
	Preterm	Caregiver	0.5	Unfamiliar	C	EN (US)	13	4	1200	2070
	Preterm	Caregiver	0.5	Unfamiliar	C	EN (US)	6	4	890	910
Zlochower et al. (2001)	Typical	Caregiver	NA	Familiar	P	EN (US)	15	4		660
	Typical	Caregiver	NA	Familiar	P	EN (US)	20	4		440
Gratier et al. (2015)	Typical	Caregiver	0.429	Familiar	C	FR	35	2.7	730	594.6
	Typical	Caregiver	0.533	Familiar	C	FR	15	4.6	904	768.8

S1.3 World Map of Data on Turn Taking

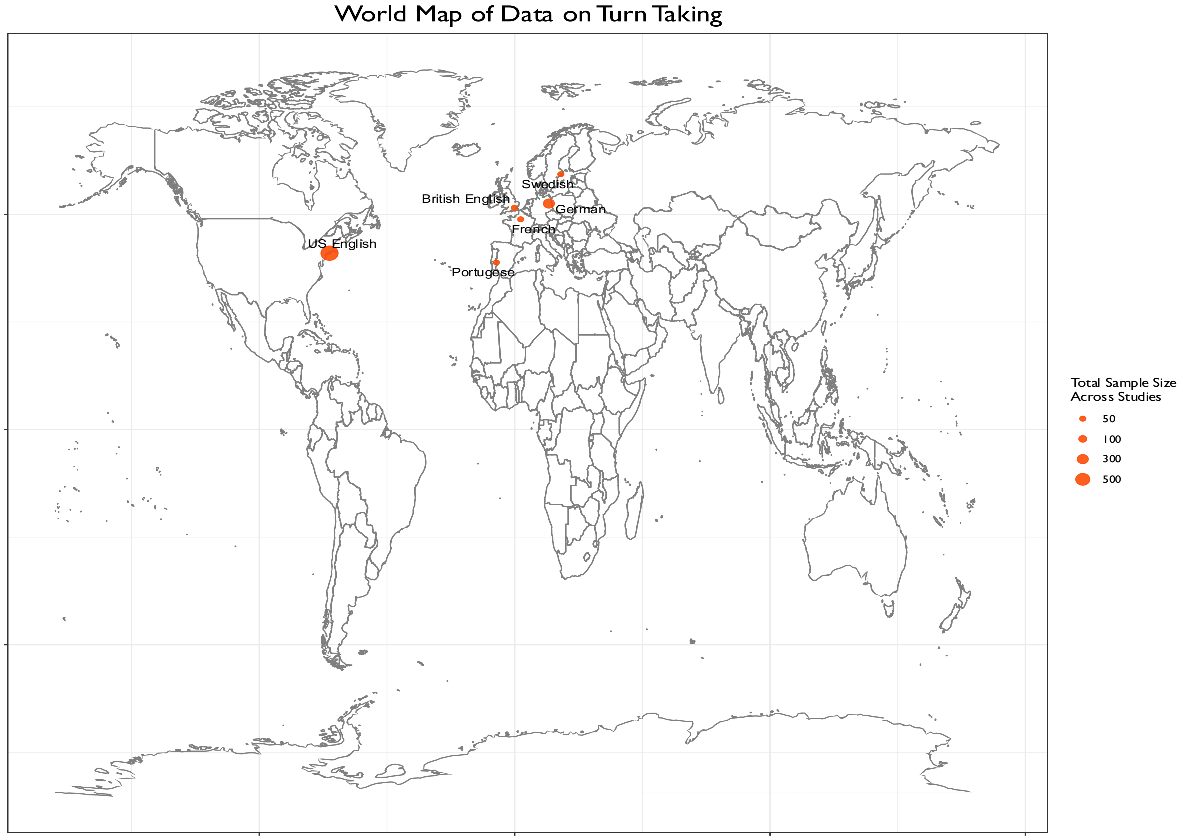


Figure S1.3 - World map of languages investigated by the studies. Each point represents a language, the size of which indicates the cumulative sample size across the studies.

S1.4 Discussion of possible biases in the search and selection of studies

Systematic searches are often argued to be less prone to bias than more subjective overviews of the literature. Nevertheless, these searches and their results are not immune from bias. We here discuss some of the potential sources of biases and how we dealt with them in our work. First, the choice of search terms might select a biased subset of the literature. We strived to make our list of search terms as inclusive as possible (while still not generating too many false positives in the search). We revised our list by conducting initial searches and reading carefully through the papers. We used these initial searchers to include additional terms before we conducted the final systematic search. Nevertheless, we might still have neglected some crucial terms. Second, we might have biased our findings by excluding relevant papers that appeared in the search, especially when not conforming to our hypotheses. However, in this study we did not have favorite hypotheses to assess, besides evaluating how well the data conformed to very vague model predictions. On the contrary, we had strong incentives to include as much data as possible to provide robust estimations of the range of response latencies in turn-taking. Third, we chose to include both published and grey literature (as available on google scholar during our forward literature search). We did include some grey literature (e.g., Lange (2016) is a thesis and Dahlby (2011) is a paper from conference proceedings), and we do not exclude that a broader search and targeted email requests could yield more data. Nevertheless, we would argue that (except for typical-atypical contrasts) our data does not involve effect sizes and statistical significance tests, but only on estimates of response latencies, and we fail to see a strong argument for incentives that would bias those estimate reports or their publication. Fourth, bias might arise as a function of which authors and studies chose to report which estimates (e.g. studies with missing measures of uncertainty being systematically lower quality than the others). We chose not to directly contact the original authors to provide the missing data, but instead to impute them. Most of the papers with missing data were older papers, and from previous work (Fusaroli et al 2017; Parola et al 2020) we know that answers are very unlikely. Further, missing data imputation was shown to not change any of the results, compared to simply excluding the studies with missing data. Fifth, the estimates provided in the studies might be biased themselves. For example, the one study testing the Interaction Engine Hypothesis includes Levinson - the original proponent of the model - as co-author. One could more explicitly model this bias; however, the number of studies involving Levinson is so small compared to the number of studies included that this did not make any noticeable difference on the estimates. Other common biases such as outcome selection (authors only choosing to report outcome measures for which manipulation created a significant effect) do not apply to the specific nature of our data (simple estimates). However, the studies of atypical populations, being focused on a contrast with typical populations, might be biased: studies with significant contrasts might have been more likely to be published. Given the sparsity of the data available, it is difficult to assess the presence of publication bias and we therefore caution against any strong interpretation in the discussion of the results. In general, no analysis can be argued to be completely unbiased, and any scientific project should be part of a larger cumulative self-correcting enterprise. We therefore provide our data openly available as a Community Augmented Meta-Analysis on the metalab website (Tsuji et al., 2014). This makes it possible to critique, integrate and update the selection in a relatively easy way (Cristia et al., 2020).

S2. Comparison with citation networks conducted on another strand of literature

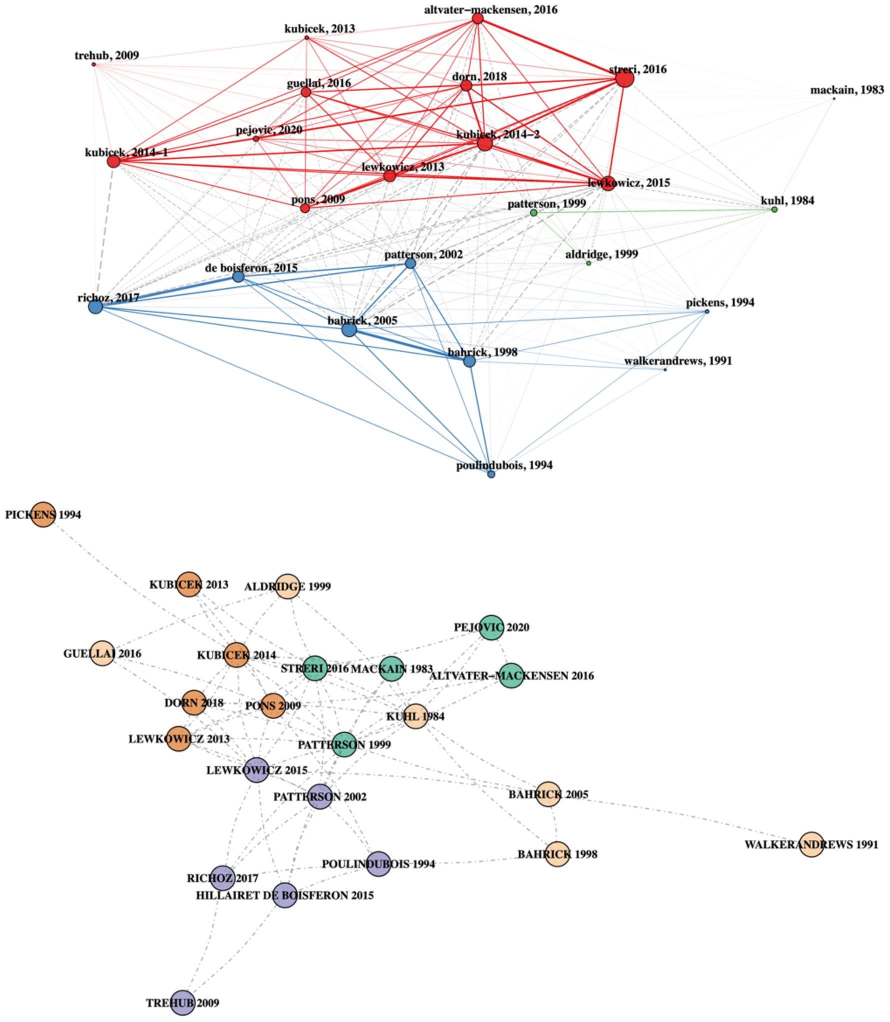


Figure S2.1 - To enable comparison with the citation network plots in the current paper, the above network plot has been included here, with permission from the authors of Cox et al. (2021). Cox, C. M. M., Keren-Portnoy, T., Roepstorff, A., & Fusaroli, R. (2021). A Bayesian meta-analysis of infants' ability to perceive audio–visual congruence for speech. Infancy.

S3. Details on statistical modeling, implementation and quality checks

We fitted the meta-analytic model as a three-level hierarchical Bayesian robust regression model. We used weakly informative priors in order to further avoid the influence of outliers and to regularize the estimated coefficients for possible moderators. The prior on the average response time (on a log scale) was modeled as a normal distribution centered at 7 with a standard deviation of 0.5. This prior implies that we expect average response times in a range between 400 ms and 3 s. The prior on the standard deviation of the average response time by sample and study (that is, the average difference we expect between a given sample or study and the population mean) was a positive truncated normal distribution centered at 0, with a standard deviation of 0.5. This implies we expect single studies to report anything between 150 ms and 8 seconds, and single samples within studies anything between 55 ms and 22 seconds. Lastly, we used a gamma distribution with a shape parameter of 2 and scale parameter of 0.1 for the Student's t-distribution's degrees of freedom (v), which allows for single studies to have deviant estimates, roughly between -25 seconds and 40 seconds, albeit it does not exclude rare more extreme values. See Figure S3.

We performed prior predictive checks to ensure the reasonableness of our priors, that is, that the plausible values of backchannel occurrences predicted by the model on the base of the priors only would exclude implausibly high or low values. The models were fitted using Hamiltonian Monte Carlo samplers, with 2 parallel chains with 10000 iterations each, an adapt delta of 0.99 and a maximum tree-depth of 20. The quality of the models was assessed by: i) ensuring no divergences in the estimation process; ii) visual inspection of the Markov chains to ensure stationarity and overlap between chains; iii) ensuring R-hat statistics to be less than 1.02 and number of effective bulk and tail samples to be above 200; iv) visually ensuring the model had learned from the data by plotting prior against posterior estimates and assessing whether the posteriors had lower variance than the priors, and were not unduly constrained by the prior, and v) performing posterior predictive checks to ensure no obvious issue in the model predictions (Gabry et al., 2019; Gelman et al., 2020; Schad et al., 2020), vi) conducting prior sensitivity analyses (see S10). Estimates from the model are reported as mean and 95% Compatibility Intervals (CI) of the posterior estimates, that is, the range covered by the middle 95% of the posterior samples.

The model was fitted both on the dataset excluding studies with missing data and on the twenty imputed datasets to assess the impact of the imputation. In the latter, each of the datasets was used to fit a model, and the posterior samples of each of the 20 models were pooled together, thus including the full uncertainty across estimates. This yielded the respective estimates of 1043 ms (95% CI [804, 1380], with 45 measures) and 1022 ms (95% CI [812, 1300], with 55 measures). The minimal difference of 21 ms in the mean estimate and the almost complete overlap in the estimated distributions led us to assume no substantial bias on the estimation of the pooled response latency as a result of imputing the missing values. We therefore proceeded using only the multiple imputed datasets.

Once satisfied with the models' quality, we used Pareto Smoothed Importance Sampling to estimate Leave-One-Out-Cross-Validated error (stacking weights) in each model to assess their relative credibility, and ability to generalize to new data (Vehtari et al., 2017). Further, the explained variance for the baseline and age model were assessed in order to determine whether including more moderators was necessary. In the moderator models, evidence ratios were calculated in the form of the posterior probability of the directed hypothesis against the posterior probability of all the alternatives.

Note that usually meta-analyses require a sensitivity analysis in order to determine whether or not there are any indications of a publication bias. Since the data was extracted from a wide range of papers dealing with different research questions and hypotheses, trying to determine the possible publication bias for the response latencies would provide uninformative results.

Note that all the code to implement the analysis is available on OSF (https://osf.io/wkceb/?view_only=9b90de634b9d4defad8da4e04ede7068) and builds on and improves the code used in previous meta-analyses (Fusaroli et al., 2017; Parola et al., 2020; Weed & Fusaroli, 2020).

All analyses were performed in R 4.1.1, *brms* 2.16.1, Stan 2.27, cmdstanr, and RStudio 1.4 (Bürkner, 2017; R Core Team, 2018; RStudio Team, 2020; Stan Development Team, 2021).

S4. Baseline model details

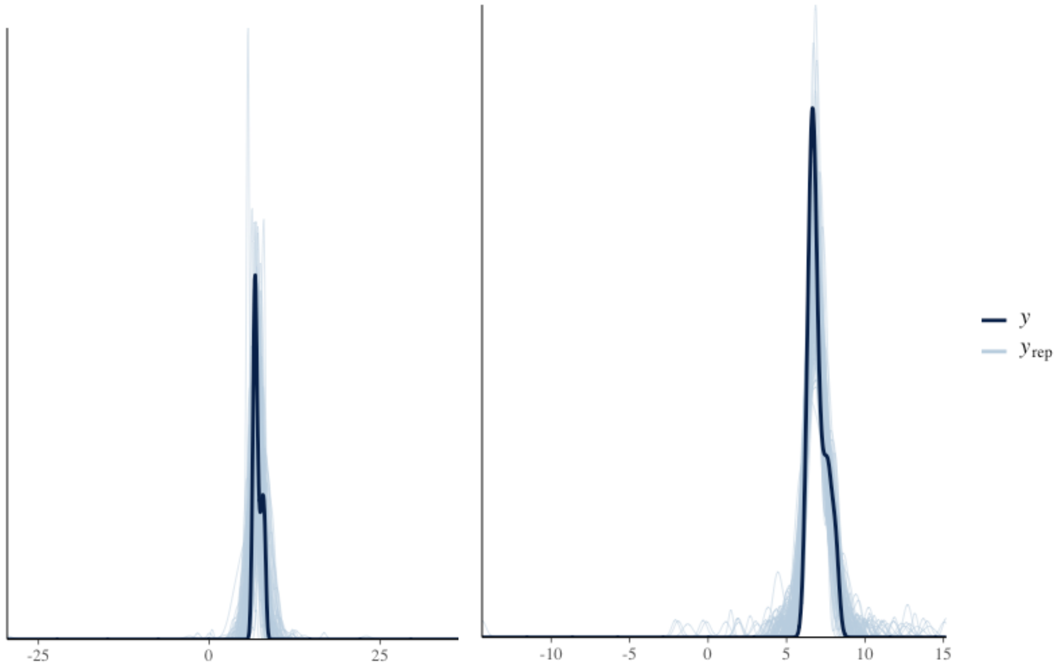


Figure S4.1 – Prior (left) and posterior (right) predictive checks for the baseline model. The black line indicates the actual data, each of the blue lines indicates predictions from

the model (before seeing the data in the left panel, after seeing the data in the right panel). Note the different scales between the two panels.

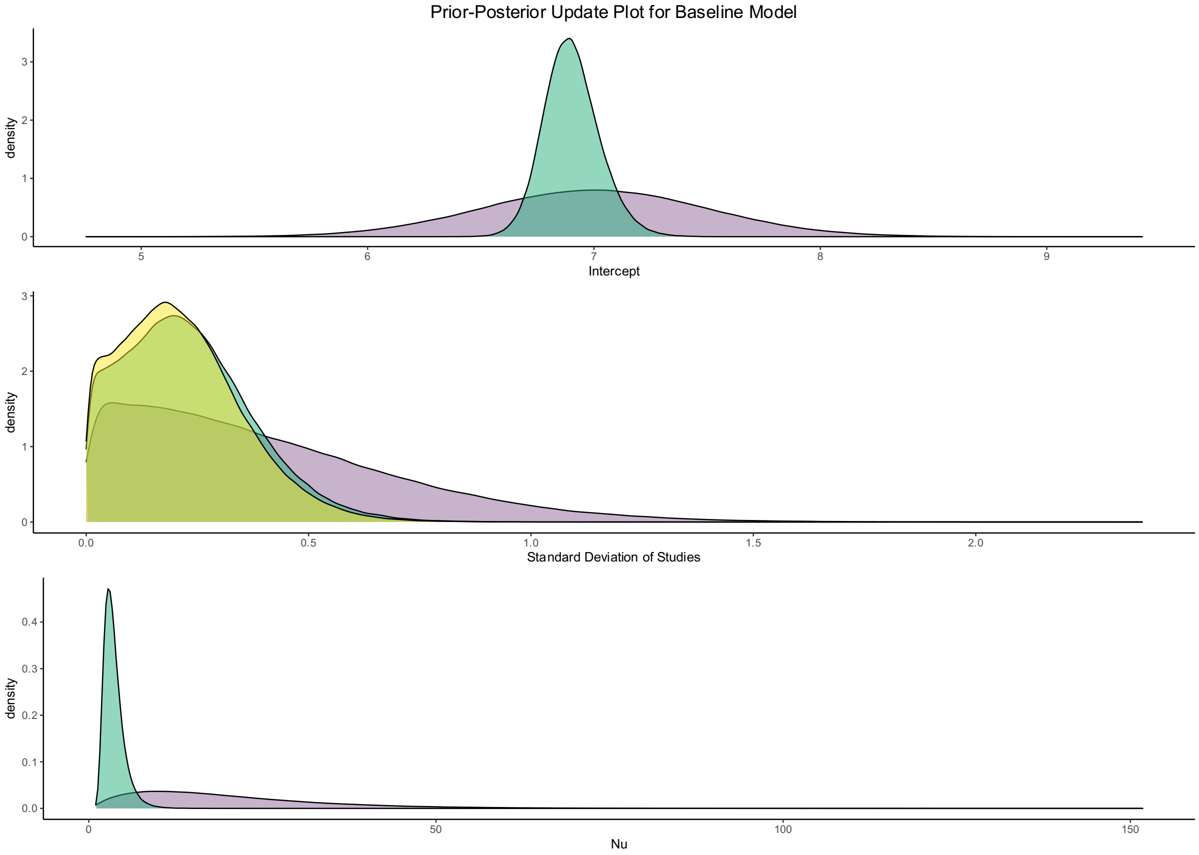


Figure S4.2 – Prior posterior update checks for the key parameters in the baseline model. Violet distributions indicate the prior distributions, other colors posterior estimates.

Table S4.1 – Baseline Model Estimates

	Population-level estimate Mean (95% CIs)	Article-level variance SD (95% CIs)	Study-level variance SD (95% CIs)
Intercept	6.90 (6.68 7.16)	0.23 (0.01 0.55)	0.21 (0.01 0.52)
Nu	3.61 (1.67 7.29)	-	-

S5. Age model details

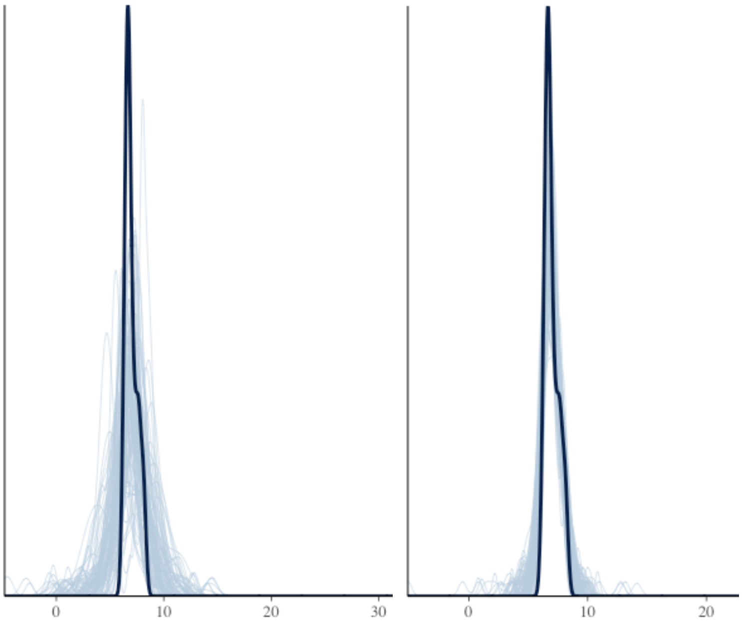


Figure S5.1– Prior (left) and posterior (right) predictive checks for the age model. The black line indicates the actual data, each of the blue lines indicates predictions from the model (before seeing the data in the left panel, after seeing the data in the right panel). Note the different scales between the two panels.

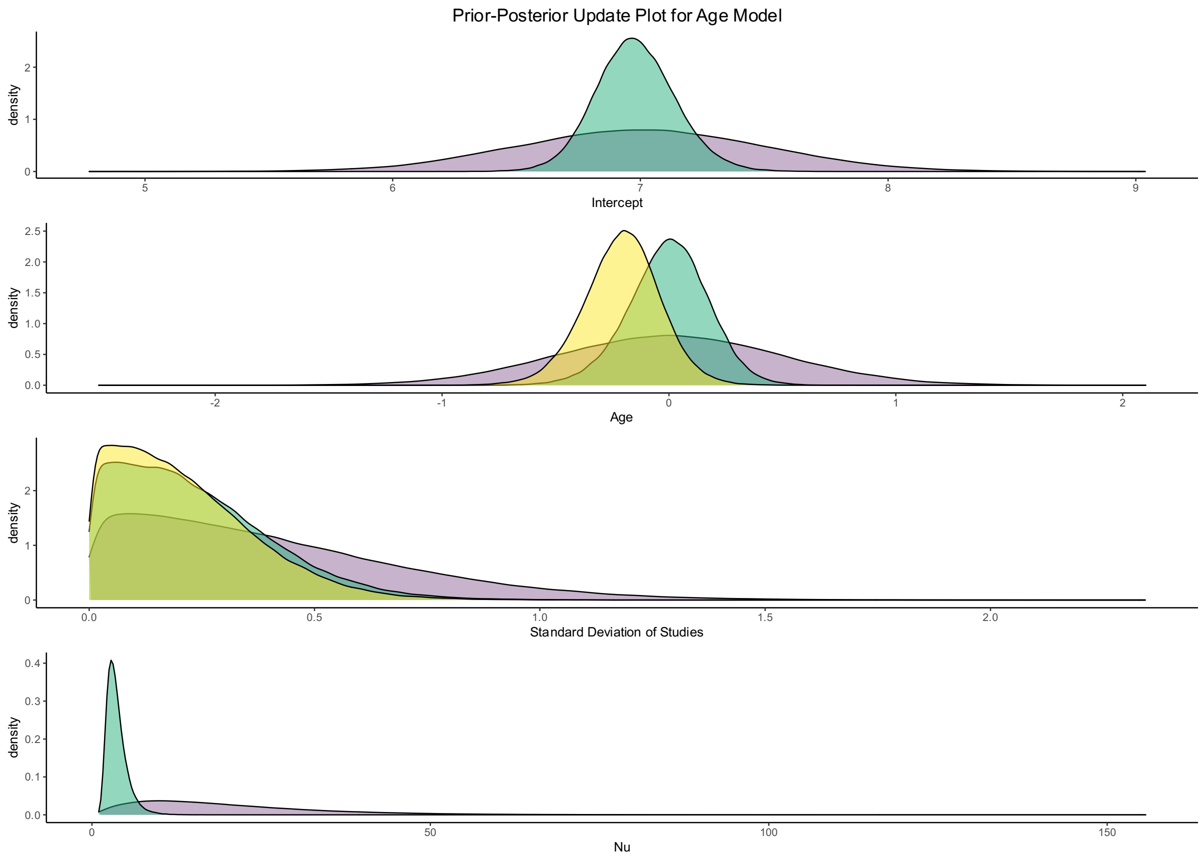


Figure S5.2– Prior posterior update checks for the key parameters in the age model. Violet distributions indicate the prior distributions, other colors posterior estimates.

Table S5.1 – Age model estimates

	Population level estimate Mean (95% CIs)	Article level variance SD (95% CIs)	Study level variance SD (95% CIs)
Intercept	7.18 (6.71 7.45)	0.23 (0.01 0.63)	0.21 (0.01 0.58)
Age (linear)	-0.00 (-0.37 0.34)	0.21 (0.01 0.60)	0.19 (0.01 0.55)
Age (squared)	-0.21 (-0.55 0.12)	0.17 (0.01 0.51)	0.16 (0.01 0.47)
Nu	3.67 (1.68 7.49)	-	-

A control analysis was performed using only the conditions containing a wide representation of different ages (cf. Table 1 and Figure 3): the studies relying on the methods “conversation” (C) and “free play” (P) within the American English samples (“EN(US)”). This control analysis yielded approximately the same pattern, albeit with a steeper predicted increase and overall faster response latencies: 875 ms at 3 months, 942 ms at 12 months, 1024 ms at 21 months and 1046 ms at 30 months (cf. Figure S5.3).

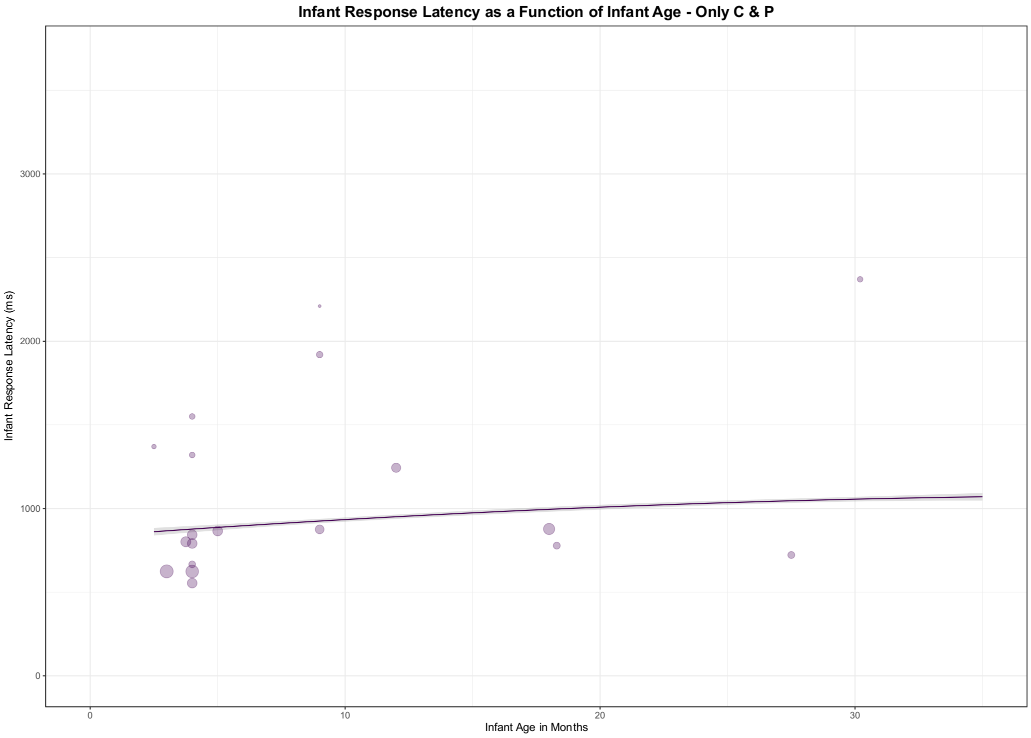


Figure S5.3 – Response latency as a function of age in the control subset of studies

S6. Moderators model details

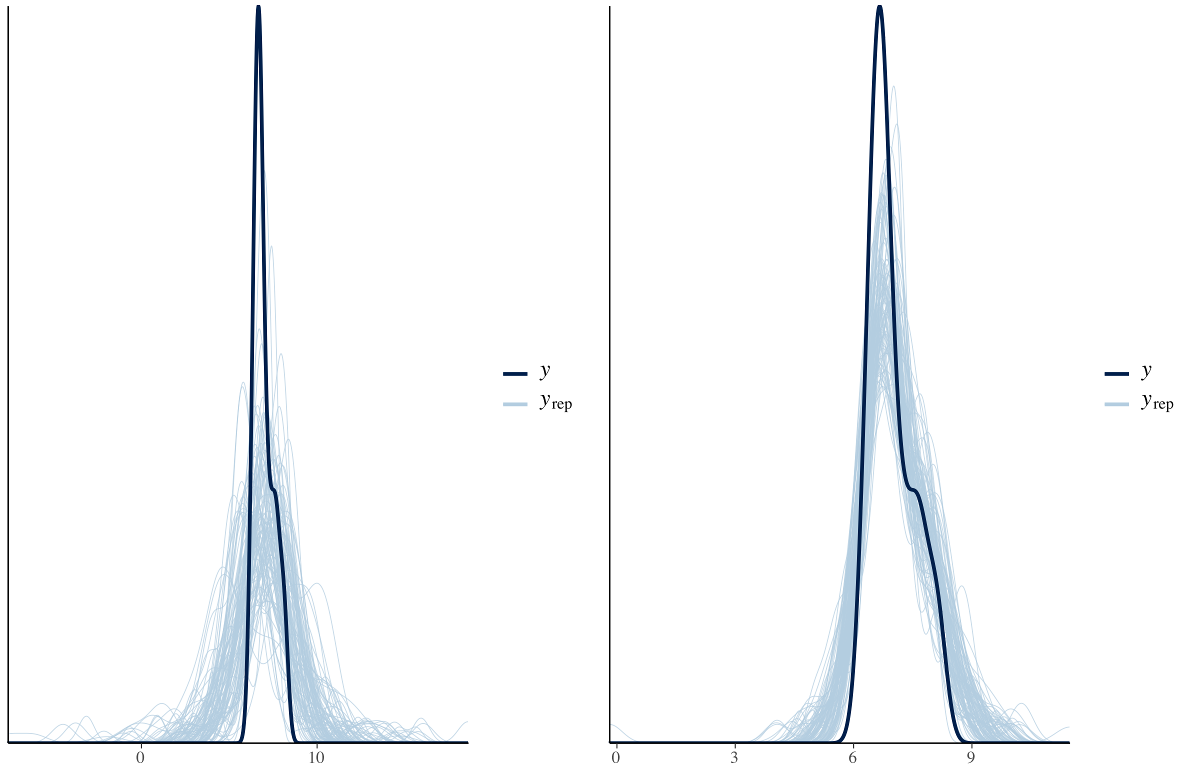


Figure S6.1 – Prior (left) and posterior (right) predictive checks for the moderators model. The black line indicates the actual data, each of the blue lines indicates predictions from the model (before seeing the data in the left panel, after seeing the data in the right panel). Note the different scales between the two panels.

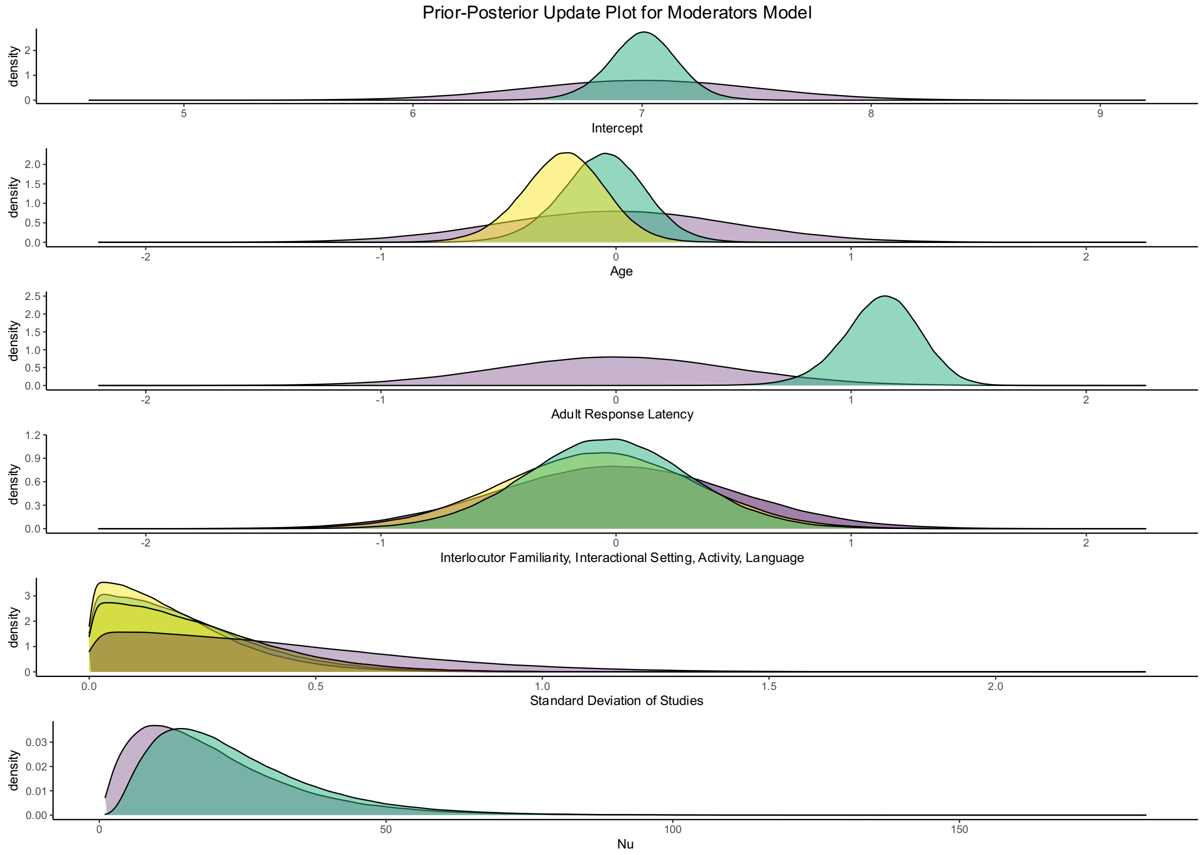


Figure S6.2– Prior posterior update checks for the key parameters in the moderators model. Violet distributions indicate the prior distributions, other colors posterior estimates.

Table S6.1 – Moderators model estimates

	Population level estimate Mean (95% CIs)	Article level variance SD (95% CIs)	Study level variance SD (95% CIs)
Intercept	7.28 (6.75 7.79)	0.23 (0.01 0.67)	0.19 (0.01 0.56)
Age (linear)	-0.06 (-0.43 0.29)	0.21 (0.01 0.59)	0.17 (0.01 0.50)

Age (squared)	-0.22 (-0.59 0.13)	0.18 (0.01 0.54)	0.16 (0.01 0.46)
Adult RTL	1.13 (0.78 1.43)	-	-
MethodA; Familiar Environment; US Eng; Caregiver	0.00 (-0.98 0.98)	-	-
MethodC; Familiar Environment; US Eng; Caregiver	-0.08 (-0.91 0.73)	-	-
MethodP; Familiar Environment; US Eng; Caregiver	-0.14 (-0.82 0.51)	-	-
MethodQ; Familiar Environment; US Eng; Caregiver	0.73 (-0.05 1.47)	-	-
MethodA; Unfamiliar Environment; US Eng; Caregiver	-0.00 (-0.97 0.97)	-	-
MethodC; Unfamiliar Environment; US Eng; Caregiver	-0.13 (-0.96 0.65)	-	-
MethodP; Unfamiliar Environment; US Eng; Caregiver	-0.23 (-0.79 0.33)	-	-
MethodQ; Unfamiliar Environment; US Eng; Caregiver	-0.00 (-0.97 0.97)	-	-
MethodA; Familiar Environment; Other; Caregiver	0.20 (-0.46 0.85)	-	-

MethodC; Familiar Environment; Other; Caregiver	0.27 (-0.62 1.16)	-	-
MethodP; Familiar Environment; Other; Caregiver	-0.00 (-0.98 0.99)	-	-
MethodQ; Familiar Environment; Other; Caregiver	0.00 (-0.98 0.98)	-	-
MethodA; Unfamiliar Environment; Other; Caregiver	-0.00 (-0.98 0.98)	-	-
MethodC; Unfamiliar Environment; Other; Caregiver	-0.00 (-0.98 0.98)	-	-
MethodP; Unfamiliar Environment; Other; Caregiver	-0.04 (-0.74 0.66)	-	-
MethodQ; Unfamiliar Environment; Other; Caregiver	-0.00 (-0.98 0.98)	-	-
MethodA; Familiar Environment; US Eng; Stranger	0.00 (-0.98 0.98)	-	-
MethodC; Familiar Environment; US Eng; Stranger	-0.26 (-0.95 0.45)	-	-
MethodP; Familiar Environment; US Eng; Stranger	-0.13 (-0.92 0.63)	-	-

MethodQ; Familiar Environment; US Eng; Stranger	0.00 (-0.98 0.98)	-	-
MethodA; Unfamiliar Environment; US Eng; Stranger	-0.00 (-0.98 0.98)	-	-
MethodC; Unfamiliar Environment; US Eng; Stranger	0.00 (-0.97 0.98)	-	-
MethodP; Unfamiliar Environment; US Eng; Stranger	-0.20 (-1.02 0.59)	-	-
MethodQ; Unfamiliar Environment; US Eng; Stranger	0.00 (-0.98 0.98)	-	-
MethodA; Familiar Environment; Other; Stranger	-0.00 (-0.98 0.97)	-	-
MethodC; Familiar Environment; Other; Stranger	0.00 (-0.97 0.98)	-	-
MethodP; Familiar Environment; Other; Stranger	0.00 (-0.98 0.98)	-	-
MethodQ; Familiar Environment; Other; Stranger	0.00 (-0.98 0.98)	-	-
MethodA; Unfamiliar Environment; Other; Stranger	0.00 (-0.98 0.98)	-	-

MethodC; Unfamiliar Environment; Other; Stranger	0.00 (-0.97 0.98)	-	-
MethodP; Unfamiliar Environment; Other; Stranger	0.00 (-0.98 0.98)	-	-
MethodQ; Unfamiliar Environment; Other; Stranger	0.00 (-0.98 0.97)	-	-
Nu	23.84 (5.62 59.69)	-	-

S7. Typical/atypical populations baseline model details

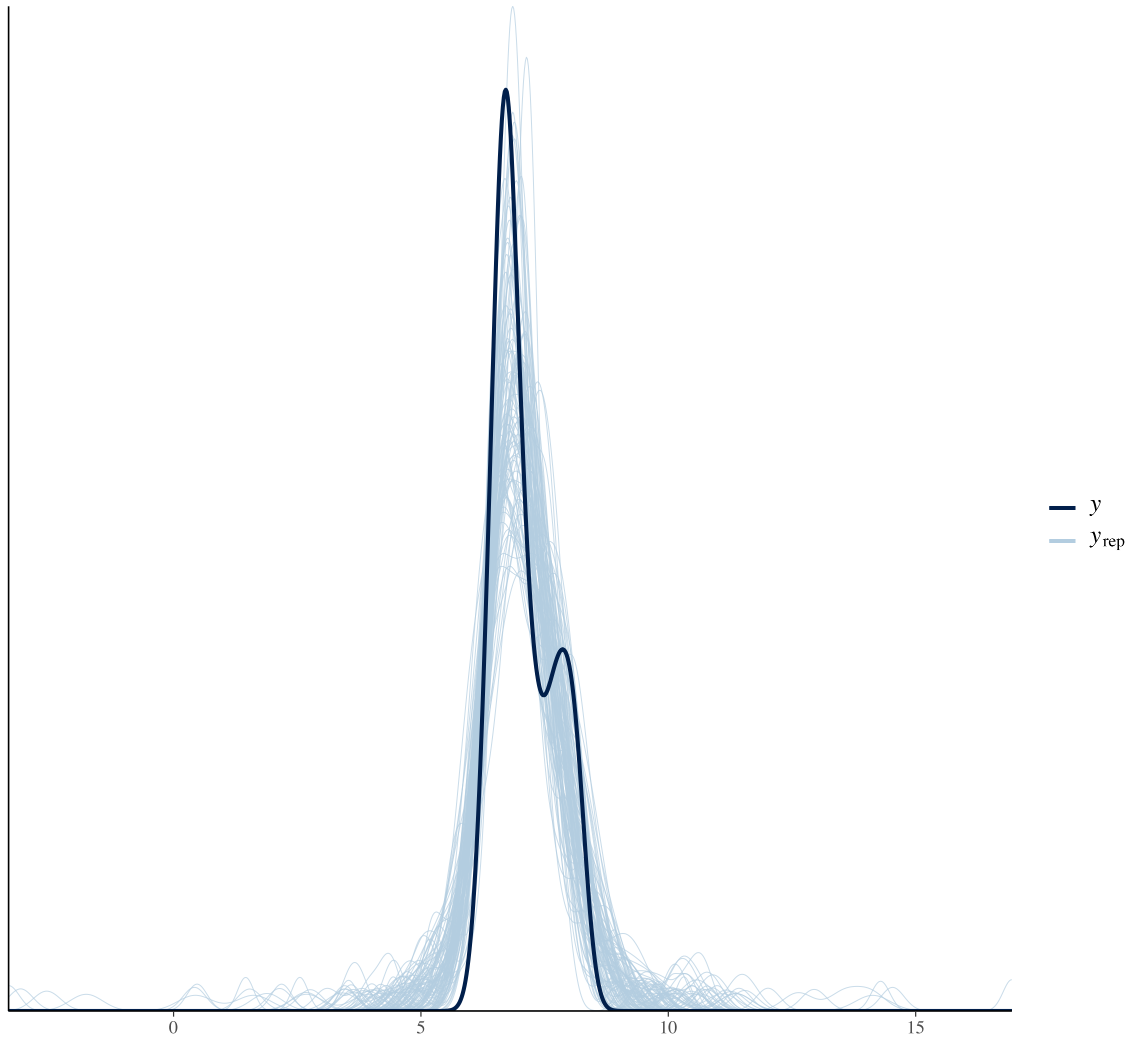


Figure S7.1– ***Posterior predictive checks for the atypical baseline model. The black line indicates the actual data, each of the blue lines indicates predictions from the model.***

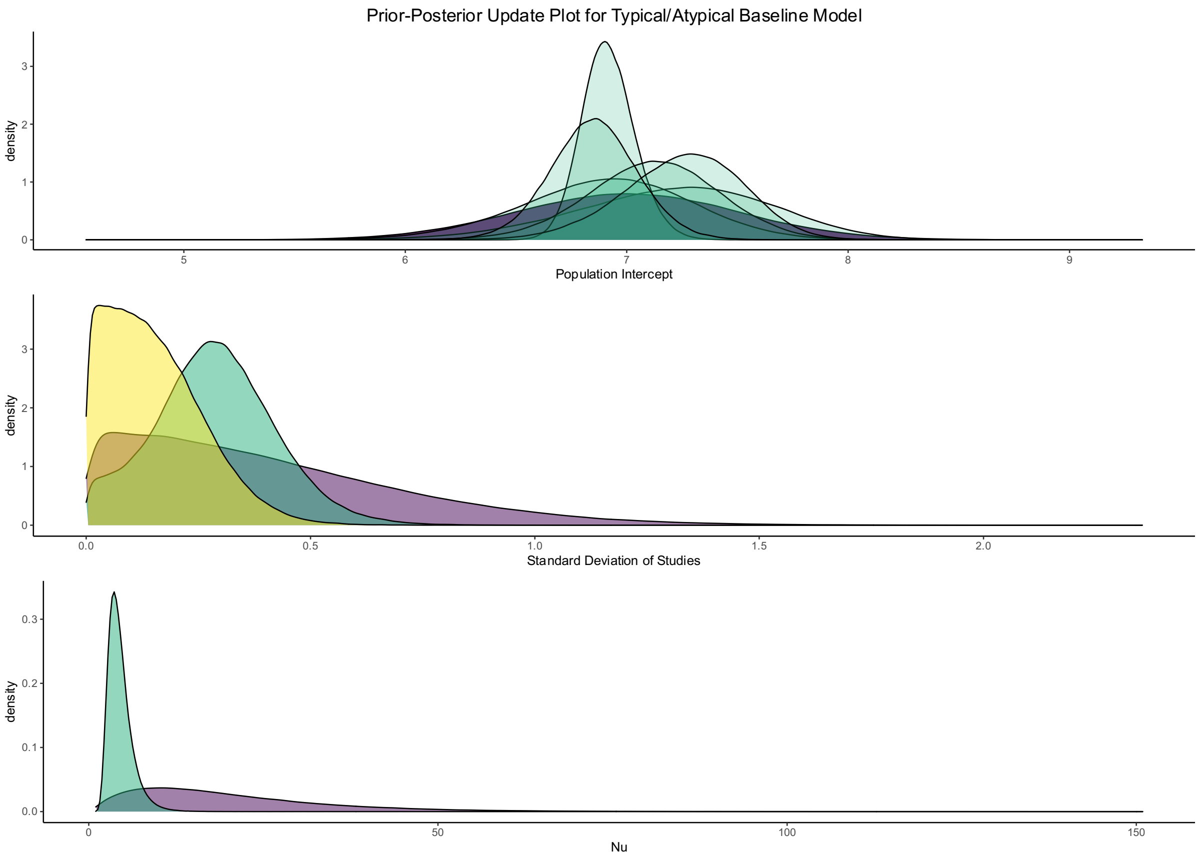


Figure S7.2 – Prior posterior update checks for the key parameters in the atypical baseline model. Violet distributions indicate the prior distributions, other colors posterior estimates.

Table S7.1 – Typical/Atypical baseline model estimates

	Population level estimate Mean (95% CIs)	Article level variance SD (95% CIs)	Study level variance SD (95% CIs)
Intercept	-	0.29 (0.03 0.56)	0.15 (0.01 0.40)
Autism	7.20 (6.27 8.01)	-	-
Cochlear Implant	6.90 (6.10 7.63)	-	-

Typical	6.92 (6.69 7.17)	-	-
Insecure Attached	7.12 (6.52 7.70)	-	-
Preterm	7.25 (6.68 7.75)	-	-
Stutterer	6.85 (6.46 7.25)	-	-
Nu	4.50 (2.17 8.92)	-	-

S8. Typical/atypical populations age model details

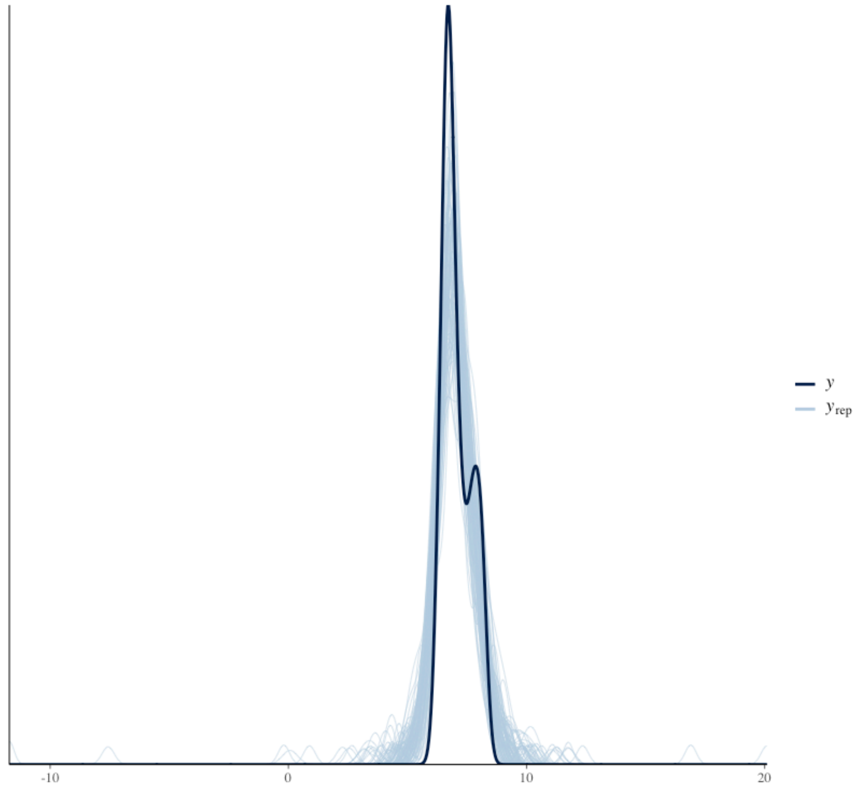


Figure S8.1– Posterior predictive checks for the atypical age model. The black line indicates the actual data, each of the blue lines indicates predictions from the model

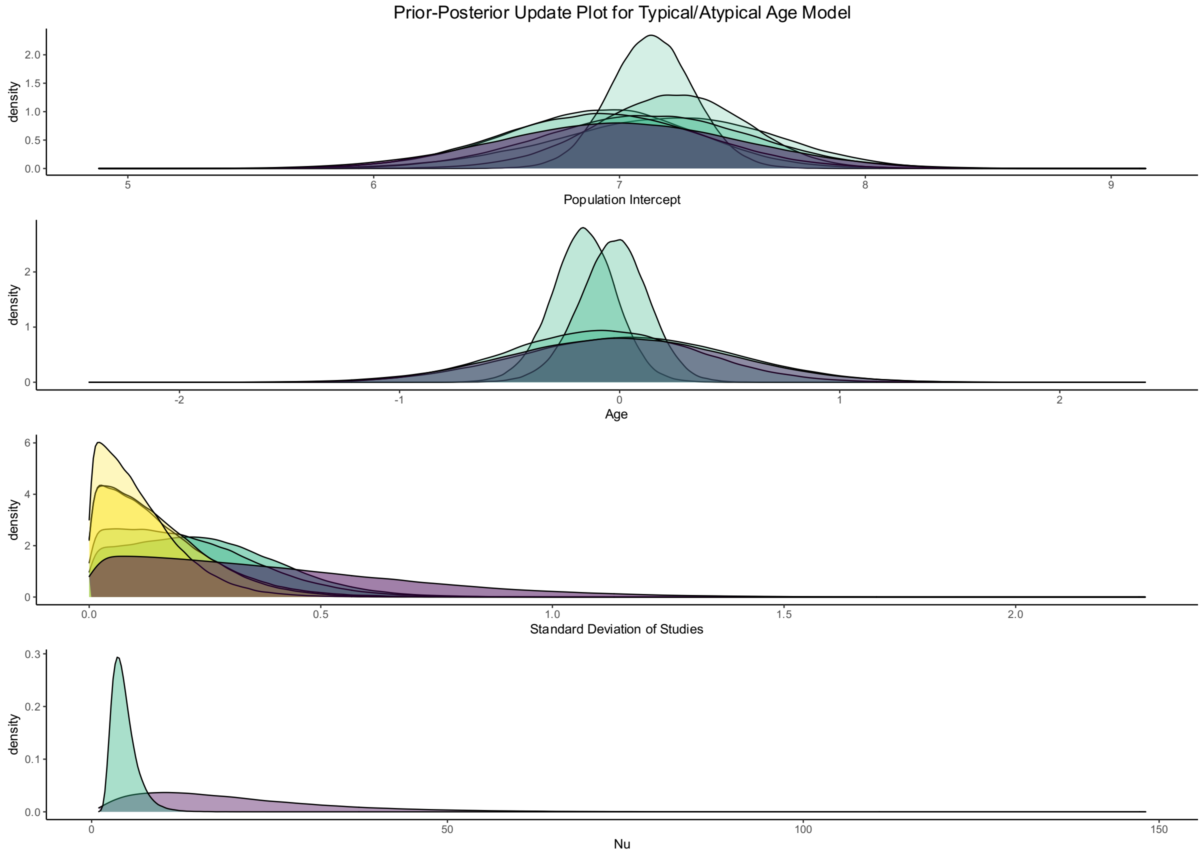


Figure S8.2– Prior posterior update checks for the key parameters in the atypical age model. Violet distributions indicate the prior distributions, other colors posterior estimates.

Table S8.1 – Typical/atypical age model estimates

	Population level estimate Mean (95% CIs)	Article level variance SD (95% CIs)	Study level variance SD (95% CIs)
Intercept	-	0.25 (0.02 0.61)	0.16 (0.01 0.45)
Age (linear)	-	0.22 (0.01 0.57)	0.15 (0.01 0.42)

Age (squared)	-	0.15 (0.01 0.45)	0.11 (0.00 0.34)
Autism	7.17 (6.24 7.98)	-	-
Cochlear Implant	6.91 (6.10 7.65)	-	-
Typical	7.13 (6.79 7.48)	-	-
Insecure Attached	7.20 (6.56 7.78)	-	-
Preterm	7.11 (6.28 7.95)	-	-
Stutterer	6.92 (6.10 7.74)	-	-
Autism:mean_age_infants_months	-0.01 (-0.93 0.94)	-	-
Cochlear Implant:mean_age_infants_months	0.02 (-0.92 0.97)	-	-
Typical:mean_age_infants_months	-0.03 (-0.35 0.28)	-	-
Insecure Attached:mean_age_infants_months	-0.14 (-1.00 0.75)	-	-
Preterm:mean_age_infants_months	-0.08 (-0.91 0.76)	-	-
Stutterer:mean_age_infants_months	-0.04 (-0.87 0.78)	-	-
Autism:Imean_age_infants_monthsE2	0.04 (-0.93 0.98)	-	-
Cochlear	-0.01 (-0.98 0.97)	-	-

Implant:Imean_age_infants_monthsE2			
Typical:Imean_age_infants_monthsE2	-0.16 (-0.46 0.13)	-	-
Insecure Attached:Imean_age_infants_monthsE2	0.03 (-0.94 0.98)	-	-
Preterm:Imean_age_infants_monthsE2	0.04 (-0.77 0.86)	-	-
Stutterer:Imean_age_infants_monthsE2	-0.04 (-0.56 0.47)	-	-
Nu	4.63 (2.19 9.30)	-	-

Table S8.2 - Predictions for the typical/atypical age model

Population	3 months	9 months	12 months	15 months	20 months
Autism	2084 [221, 18659]	1890 [281, 12383]	1823 [298, 10777]	1725 [308, 9508]	1693 [304, 9242]
Cochlear implant	1073 [84, 14129]	1079 [125, 9528]	1083 [142, 8430]	1092 [160, 7650]	1095 [165, 7435]

Typical	1005 [216, 5041]	1074 [261, 4588]	1106 [276, 4584]	1136 [288, 4586]	1153 [291, 4652]
Insecurely attached	2196 [187, 24945]	1942 [248, 15000]	1841 [259, 12754]	1685 [264, 10501]	1615 [259, 9885]
Preterm	1232 [174, 8780]	1250 [198, 7856]	1257 [201, 7715]	1262 [201, 7902]	1259 [197, 8091]
Stutterer	1061 [113, 9977]	1059 [169, 6681]	1057 [194, 5796]	1049 [227, 4900]	1043 [238, 4611]

S9. Typical/Atypical populations moderators model details

Table S9.1 - Contrasts typical-atypical population

			Method	A	C	P	Q
Environment	**Interactor**	**Language**	**Population**				
Familiar	Caregiver	EN (US)	Autism	0	0	1	0
			Cochlear	0	0	0	0
			Typical	0	3	6	7
			Insec. Attach	0	0	0	0
			Preterm	0	0	0	0

			Stutterer	0	0	2	0
		Other	Autism	0	0	0	0
			Cochlear	0	0	0	0
			Typical	7	2	0	0
			Insec. Attach	0	0	0	0
			Preterm	1	0	0	0
			Stutterer	0	0	0	0
	Stranger	EN (US)	Autism	0	1	0	0
			Cochlear	0	0	0	0
			Typical	0	7	1	0
			Insec. Attach	0	0	0	0
			Preterm	0	0	0	0
			Stutterer	0	0	0	0
		Other	Autism	0	0	0	0
			Cochlear	0	0	0	0
			Typical	0	0	0	0
			Insec. Attach	0	0	0	0
			Preterm	0	0	0	0
			Stutterer	0	0	0	0
Unfamiliar	Caregiver	EN (US)	Autism	0	0	0	0
			Cochlear	0	2	0	0
			Typical	0	2	11	0

		Insec. Attach	0	0	0	0
		Preterm	0	3	0	0
		Stutterer	0	0	4	0
	Other	Autism	0	0	0	0
		Cochlear	0	0	0	0
		Typical	0	0	8	0
		Insec. Attach	0	0	8	0
		Preterm	0	1	0	0
		Stutterer	0	0	0	0
Stranger	EN (US)	Autism	0	0	0	0
		Cochlear	0	0	0	0
		Typical	0	0	1	0
		Insec. Attach	0	0	0	0
		Preterm	0	0	0	0
		Stutterer	0	0	0	0
	Other	Autism	0	0	0	0
		Cochlear	0	0	0	0
		Typical	0	0	0	0
		Insec. Attach	0	0	0	0
		Preterm	0	0	0	0
		Stutterer	0	0	0	0

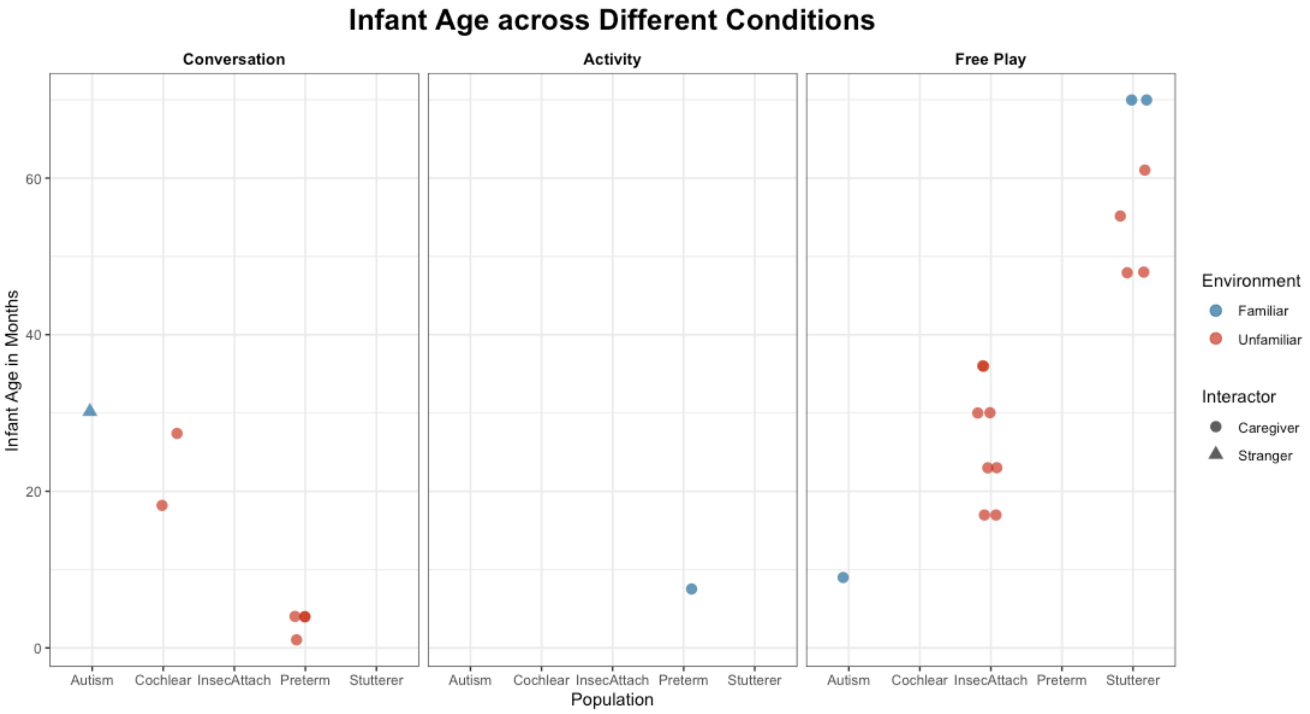


Figure S9.1 - Age distribution of data on atypical populations where InsecAttach refers to insecurely attached infants

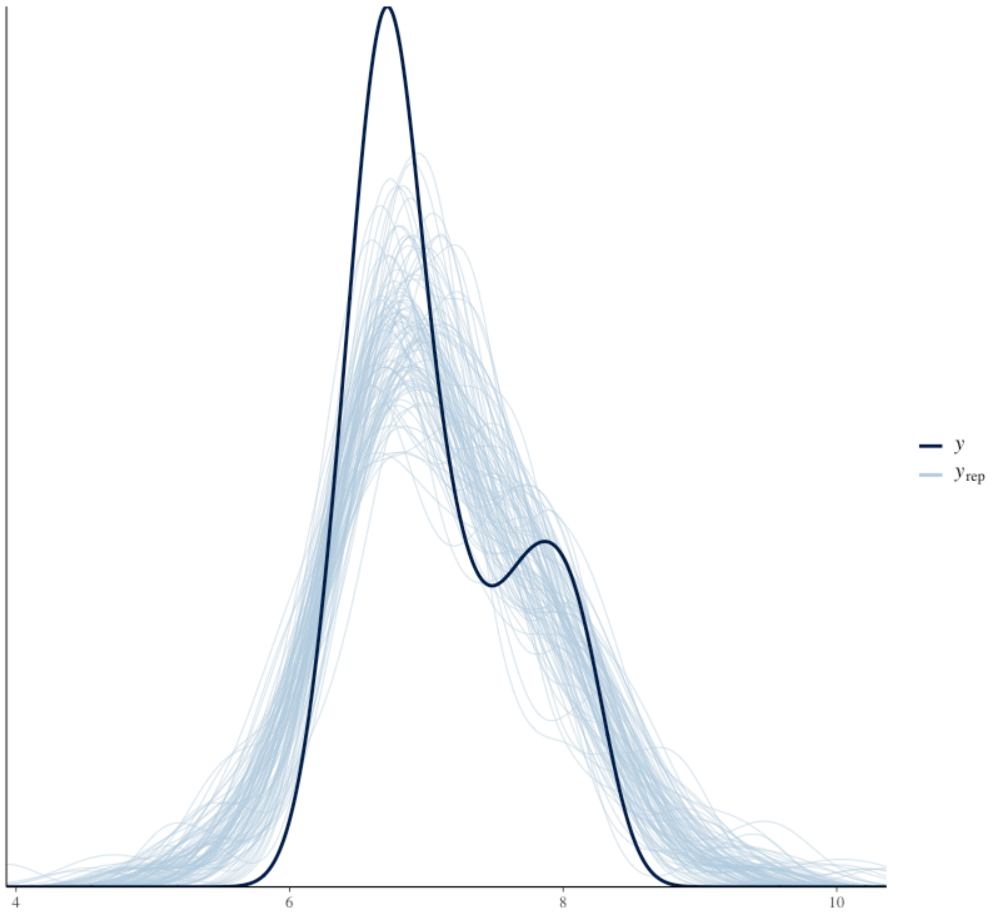


Figure S9.2– Posterior predictive checks for the atypical moderator model. The black line indicates the actual data, each of the blue lines indicates predictions from the model.

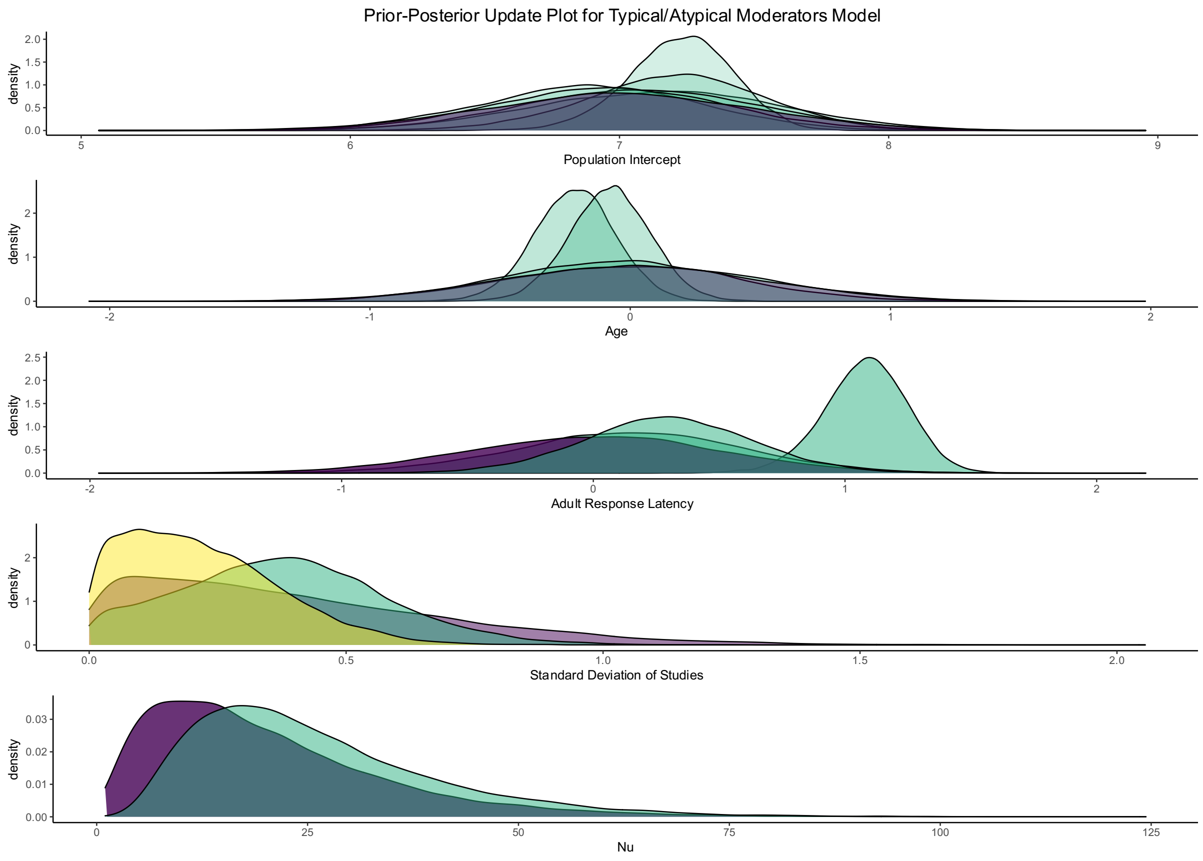


Figure S9.3– Prior posterior update checks for the key parameters in the atypical moderator model. Violet distributions indicate the prior distributions, other colors posterior estimates.

Table S9.2 - Typical/atypical moderators model estimates

	Population level estimate Mean (95% CIs)	Article level variance SD (95% CIs)	Study level variance SD (95% CIs)
Intercept	-	0.36 (0.03 0.77)	0.22 (0.01 0.56)
Age (linear)	-	0.20 (0.01 0.55)	0.14 (0.01 0.41)
Age (squared)	-	0.17 (0.01 0.51)	0.13 (0.00 0.38)

Autism	7.13 (6.21 8.01)	-	-
Cochlear Implant	6.88 (6.06 7.66)	-	-
Typical	7.21 (6.81 7.56)	-	-
Insecure Attached	7.22 (6.53 7.86)	-	-
Preterm	7.08 (6.22 7.94)	-	-
Stutterer	6.96 (6.12 7.79)	-	-
Autism:mean_age_infants_months	-0.00 (-0.93 0.95)	-	-
Cochlear Implant:mean_age_infants_months	0.03 (-0.93 0.98)	-	-
Typical:mean_age_infants_months	-0.08 (-0.40 0.23)	-	-
Insecure Attached:mean_age_infants_months	-0.12 (-1.00 0.77)	-	-
Preterm:mean_age_infants_months	-0.06 (-0.91 0.78)	-	-
Stutterer:mean_age_infants_months	-0.04 (-0.87 0.80)	-	-
Autism:Imean_age_infants_month	0.02 (-0.94 0.97)	-	-

sE2			
Cochlear Implant:Imean_age_infants_monthsE2	-0.01 (-0.98 0.96)	-	-
Typical:Imean_age_infants_monthsE2	-0.22 (-0.54 0.09)	-	-
Insecure Attach:Imean_age_infants_monthsE2	0.03 (-0.94 0.99)	-	-
Preterm:Imean_age_infants_monthsE2	0.04 (-0.79 0.86)	-	-
Stutterer:Imean_age_infants_monthsE2	-0.07 (-0.61 0.45)	-	-
Autism:adult_response_latency_log_centered	0.14 (-0.79 1.00)	-	-
Cochlear Implant:adult_response_latency_log_centered	0.04 (-0.91 0.99)	-	-
Typical:adult_response_latency_log_centered	1.06 (0.70 1.37)	-	-
Insecure	-0.02 (-0.95 0.92)	-	-

Attached:adult_response_latency_log_centered			
Preterm:adult_response_latency_log_centered	0.27 (-0.39 0.92)	-	-
Stutterer:adult_response_latency_log_centered	0.06 (-0.83 0.94)	-	-
Nu	25.76 (6.67 62.17)		

S10. Prior Robustness Checks

In the plots below we assessed whether increasing the uncertainty of our priors (on the x-axis) – and therefore making them matter less in the posterior estimation – affected our results (on the y-axis). As it can be seen in the plots, our priors did not seem to unduly affect the estimations (the red dots are quite stable across all choices of priors). The only possible exception is the prior for adult response latency, which relation to child response latency might be slightly underestimated, that is, having used a broader more uncertain priors would have slightly increased the posterior estimate. We do not consider this an issue since: 1) the shrinkage is small; 2) meta-analytically estimated effects tend to be inflated compared to many labs replications (Kvarven et al., 2020) and therefore regularizing the estimate to a more conservative value might actually be useful; 3) our qualitative inference of a substantial relation between child and adult response latencies does not change.

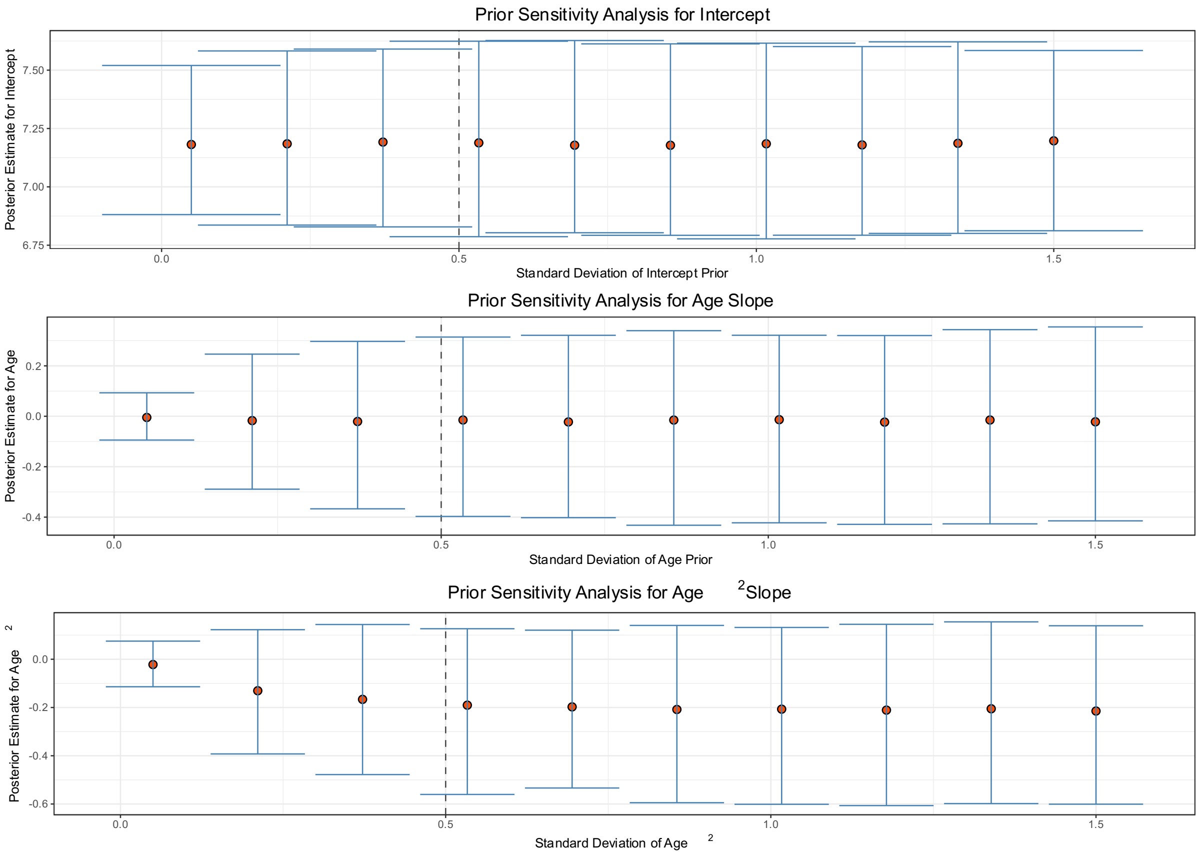
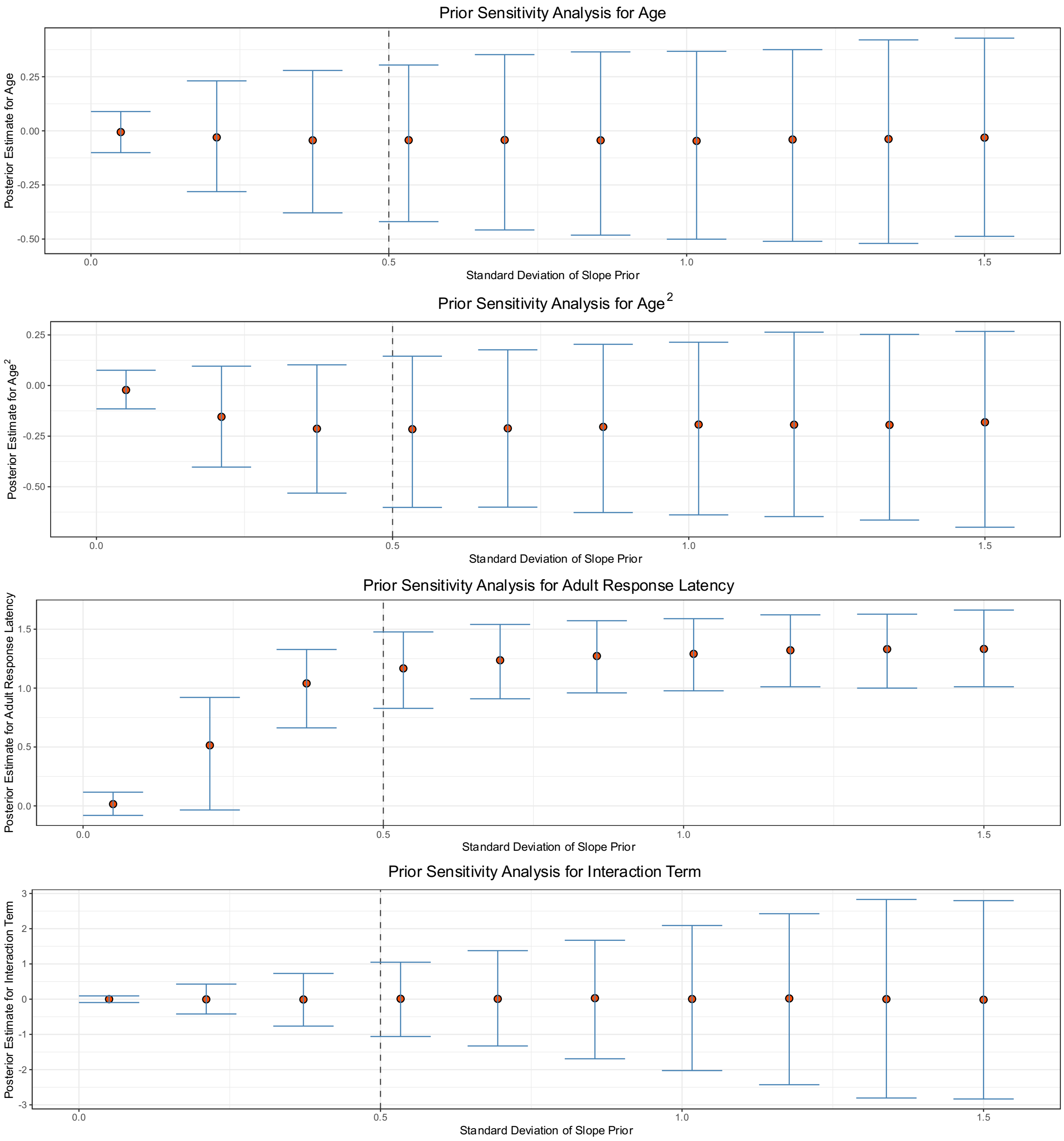


Figure S10.1– Panel of plots showing how the intercept, age and age^2 estimates in the age model change with different standard deviations for the priors. The vertical dashed line indicates the standard deviation of the prior chosen for the model.


Prior Sensitivity Analysis for Age

Prior Sensitivity Analysis for Age^2

Prior Sensitivity Analysis for Adult Response Latency

Prior Sensitivity Analysis for Interaction Term

Figure S10.2– Panel of plots showing how the intercept, age, age^2, log adult response latency, and interaction between familiarity of interlocutors, interactional setting, activity type and language. The vertical dashed line indicates the standard deviation of the prior chosen for the model.

S11. Exploratory Analyses of Gender & Country/Language

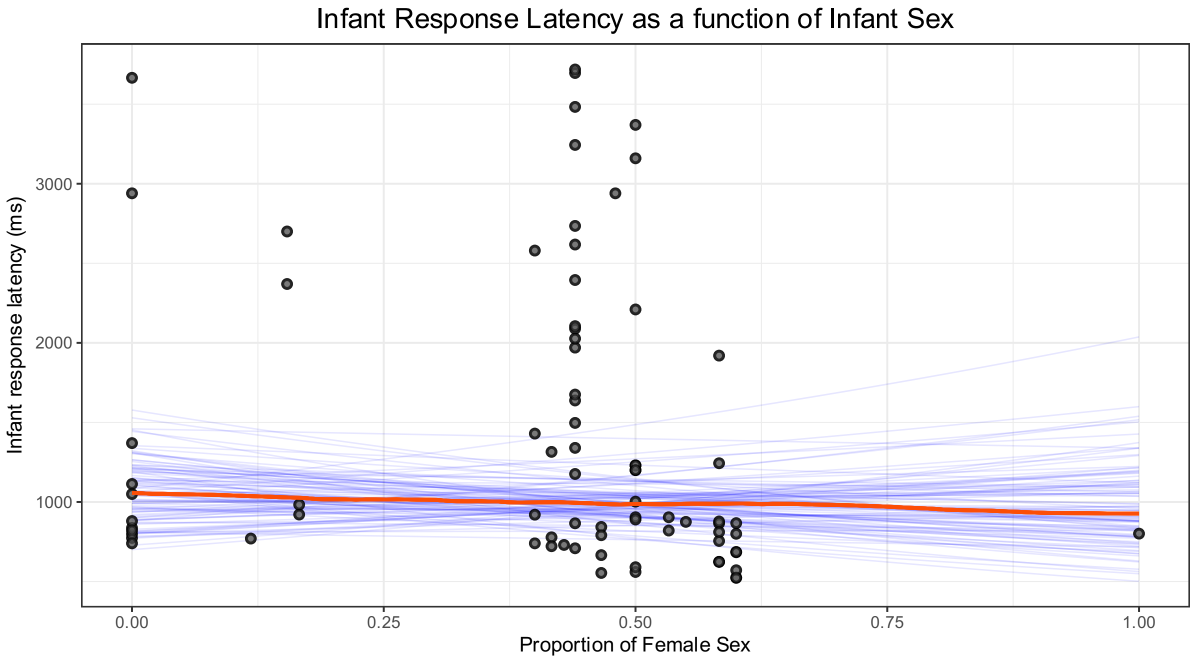


Figure S11.1 – Plot showing infant response latency as a function of infant sex. The data indicate a widely uncertain null effect of the proportion of infants who are of female sex on infant response latency: -0.12 with 95% CI [-0.74 0.50].

Language	Number of Estimates	Infant Response Latency Estimate
English (US)	51	930.07 (793.66 1151.27)
German	19	1514.15 (938.81 2281.81)
English (UK)	2	2189.91 (1097.99 3328.44)
French	2	1317.29 (610.97 2985.14)
Portuguese	1	1126.11 (424.36 2845.58)
Swedish	3	1118.87 (484.15 2283.08)

Table S11.1 – Overview of how response latency estimates differ according to individual languages. The reader should note that in many cases, these estimates are based on a small amount of data extracted from a limited number of studies (cf. discussion in section 2.4).

List of References for Supplementary Materials

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28.

Cristia, A., Tsuji, S., & Bergmann, C. (2020). *A meta-analytic approach to evaluating the explanatory adequacy of theories*. OSF Preprints. https://doi.org/10.31219/osf.io/83kg2

Fusaroli, R., Lambrechts, A., Bang, D., Bowler, D. M., & Gaigg, S. B. (2017). Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis. *Autism Research: Official Journal of the International Society for Autism Research*, *10*(3), 384–407. https://doi.org/10.1002/aur.1678

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization in Bayesian workflow. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *182*(2), 389–402.

Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., Kennedy, L., Gabry, J., Bürkner, P.-C., & Modrák, M. (2020). Bayesian Workflow. *ArXiv:2011.01808 [Stat]*. http://arxiv.org/abs/2011.01808

Kvarven, A., Strømland, E., & Johannesson, M. (2020). "Comparing meta-analyses and preregistered multiplelaboratory replication projects": Author Correction. *Nature Human Behaviour*, *4*(6), 659–663. https://doi.org/10.1038/s41562-020-0864-3

Parola, A., Simonsen, A., Bliksted, V., & Fusaroli, R. (2020). Voice patterns in schizophrenia: A systematic review and Bayesian meta-analysis. *Schizophrenia Research*, *216*, 24–40. https://doi.org/10.1016/j.schres.2019.11.031

R Core Team. (2018). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing.

RStudio Team. (2020). *RStudio: Integrated Development for R.* RStudio.

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 104038. https://doi.org/10.1016/j.jml.2019.104038

Stan Development Team. (2021). *Stan Modeling Language Users Guide and Reference Manual, 2.27*.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Weed, E., & Fusaroli, R. (2020). Acoustic Measures of Prosody in Right-Hemisphere Damage: A Systematic Review and Meta-Analysis. *Journal of Speech, Language, and Hearing Research*, *63*(6), 1762–1775. https://doi.org/10.1044/2020_JSLHR-19-00241

3 General Discussion

The findings from these three studies highlight that behavioural adaptations in early child-caregiver interactions are shaped by social contingency, and that the nature of these adaptations depends on many different factors, from ontogeny and culture to biology. In this section, I return to the following overarching questions that were posed in the introduction (repeated below):

I. *To what extent do caregivers speaking different languages exhibit variability in the acoustic modifications they make in their IDS? Do caregivers change the acoustics of their IDS as the infant grows older?*

II. *How can we characterise the timing structures of early child-caregiver vocal interactions and their developmental trajectories during ontogeny? To what extent can we gain information about the mechanisms underlying the timing of turn-taking?*

The meta-analysis of IDS acoustics in ***Study I*** provided support to the claim that IDS exhibits systematic acoustic differences compared to ADS across many different languages and cultures, notably a higher f_0, a higher degree of f_0 variability, an expanded vowel space area, a slower articulation rate and a longer vowel duration. There were also dynamic changes in certain properties of IDS acoustics during early development; f_0, articulation rate and vowel duration became more similar to ADS during early ontogeny, whereas vowel space area and f_0 variability remained stable up until two and three years of age, respectively. The models also provided evidence for a high degree of between-study and between-language heterogeneity, which may stem from including studies that differed from each other in substantial ways.

The meta-analytic results were partially supported by the acoustic analysis of Danish IDS and ADS in ***Study II***. Here, the findings indicated that Danish caregivers produced IDS with similar prosodic properties to those found in cross-linguistic patterns, with a higher median f_o, a greater degree of f_o variability, and a slower rate of articulation than ADS. In contrast, the vocalic measures of Danish IDS went against the cross-linguistic tendencies in the meta-analysis, with caregivers' IDS showing vowel space reduction or no changes compared to the vowel space of ADS. Danish IDS was also characterised by formant raising, lower discriminability between vowels, more within-vowel variability when compared with ADS. Only articulation rate changed with child age and became more similar to ADS in IDS the older the child.

The meta-analysis on the development of infants' turn-taking skills in ***Study III*** showed some tentative evidence that children slow down the response latencies of their vocalisations before speeding up again at around 20 months of age. In addition to this developmental trajectory, the results indicated that infant response latencies exhibited i) a strong correlation with adult response latencies, ii) similar ranges of latencies across languages, iii) slowdowns in contexts of question-answer pairs, and iv) no differences across familiarity of environment or interactant. The sparsity of the meta-analytic turn-taking data, however, meant that these developmental patterns only received tentative support. The data available to investigate these conditions remained sparse, were often confounded across moderators, and aggregated data with substantial differences within the same condition.

In the following sections, I contextualise these findings and discuss how this work adds to our knowledge about how social contingency shapes the speech content and temporal structure of early child-caregiver interactions. **Section 3.1** starts off the discussion by providing a framework to consider behavioural adaptations in child-caregiver interactions as emergent products of reciprocal feedback among ontogenetic, sociocultural and biological constraints.

With this framework in mind, **Section 3.2** explores how the findings from these three individual studies add to our knowledge about dynamic changes in behavioural adaptations during ontogeny (**Section 3.2.1**), differences across cultures (**Section 3.2.2**) as well as the balance between cross-cultural and biological constraints (**Section 3.2.3**). **Section 3.3** discusses the key directions that this work carves for future research and argues for a need to focus on turn-by-turn behavioural adaptations carried out in the 'here-and-now' (**Section 3.3.1**), incorporation of individual differences (**Section 3.3.2**), theory-driven cross-cultural and -linguistic comparisons (**Section 3.3.3**), comparative work using computational and theory-driven models from the literature on non-human animal vocalisations (**Section 3.3.4**), and explicit theory development (**Section 3.3.5**).

3.1 Interactions All the Way Down: Behavioural Adaptations, Cultural Variability & Biophysical Constraints

Early child-caregiver communicative exchanges unfold in the split-second world of social interaction and provide rich affordances that scaffold cognitive and linguistic development through processes of reciprocal alignment of behaviours (Fusaroli et al., 2019; Fusaroli & Tylén, 2012; Goodwin, 2000, 2007; Tylén et al., 2013). These pragmatic social contexts allow interactants to pursue and coordinate communicative goals, such as organising emotion, physiology, and behaviours (Feldman, 2012; Tamis-LeMonda et al., 2012; Tamis-LeMonda et al., 2014; Tarullo et al., 2017). While there are certain properties of these interactive situation spaces that have an ad hoc character (e.g., child crying or fussiness prompting caregiver responses), there are other more stable properties that play structuring roles on slower timescales (Boyd & Richerson, 2009; Henrich & McElreath, 2003; Heyes, 2016; Mesoudi & Whiten, 2004; Tylén et al., 2013). The findings in the three studies suggest that behavioural adaptations emerge as a product of complex interactions among the following three main factors that operate on different timescales, as explained below and shown in **Figure 1**:

- ***Ontogeny***: The developing capacities and knowledge of the child during early ontogeny change the social affordances in child-caregiver interactions and demand that caregivers produce behavioural adaptations appropriate to the context while allowing for dynamic changes through processes of reciprocal alignment (e.g., gradually increasing speech rate or reducing response latencies in response to children's developing speech processing capabilities);
- ***Sociocultural Norms***: The immersion of child-caregiver interactions in a specific sociocultural context places profound normative pressures on behavioural adaptations (e.g., there may be cultural variability in the modalities that caregivers use (Keller et al., 2009; Little et al., 2016) and variability in contexts of interaction (Clegg et al., 2021);
- ***Biophysical Constraints***: Our sensorimotor engagements with the world and each other are constrained by fundamental bounds on our perceptual acuity (e.g., our visual perception is limited in range and resolution), production capacities (e.g., size and movement of vocal organs), as well as inherent physical properties of objects (e.g., smaller objects tend to have higher resonant frequencies).

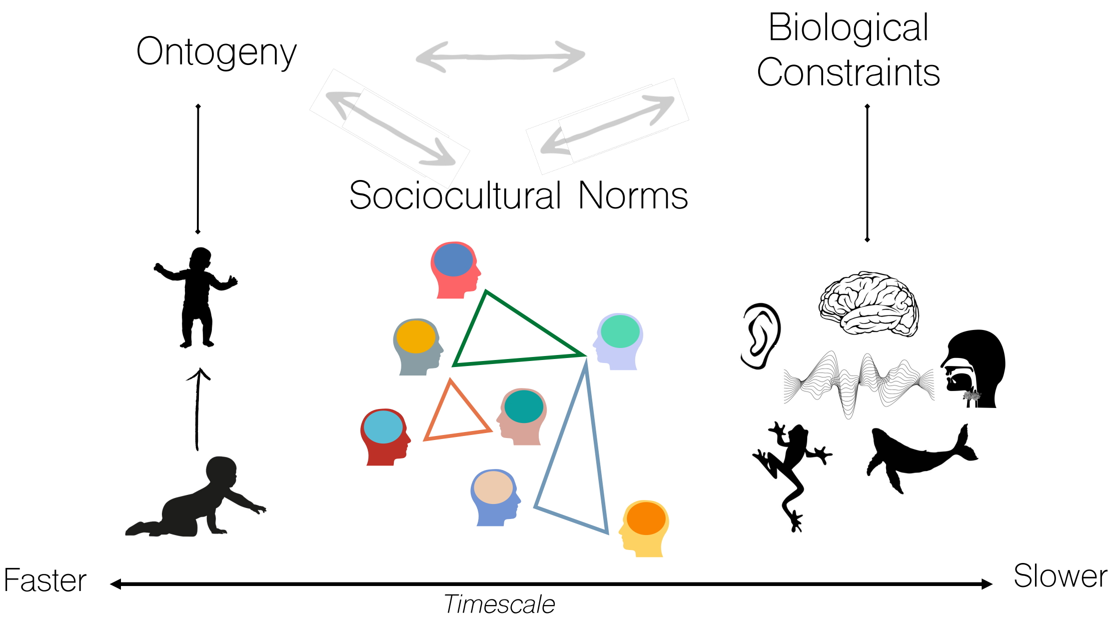


Figure 1: Visualisation of how child-caregiver interactions emerge from the reciprocal interplay among ontogeny, sociocultural norms, and biological constraints that guides behavioural adaptations. Each component operates on a different timescale. Developments in children's knowledge, abilities and skills change the relevance of different behavioural adaptations on the fast timescale of the child's first years of life (Ontogeny); Behavioural pressures from cross-cultural norms operate on a slower timescales in that they reflect and are established by a long history of interactions in a given community of speakers (Sociocultural Norms); Biological constraints operate on even longer timescales and include affordances introduced by the human anatomy and by physical objects in the interactive environmen (Biophysical Constraints).

Communicative interactants engage in a social dance not just with each other, but also with the local history of interactions, the developing capacities of the individual child, sociocultural norms and their biological grounding (Bolis & Schilbach, 2018, 2020; Fusaroli et al., 2019; Fusaroli & Tylén, 2012; Tylén et al., 2013). This complex interplay of contributions scaffold and constrain online child-caregiver interactions and forms an indivisible configuration not composed of separately functioning variables (Bolis & Schilbach, 2018, 2020; Heyes, 2016; Griffin & Gonzalez, 2003). At the same time, these moment-to-moment transactions between child and caregiver shape and reciprocally feed into longer timescales not just to reflect but also to establish cross-cultural norms and change our biology over longer timescales (Boyd & Richerson, 2009; Creanza et al., 2017; Feldman et al., 1985; Rendell et al., 2011).

What would this multifaceted construal of early child-caregiver interactions imply for a simple game of peek-a-boo? On the short timescale of ontogeny, peek-a-boo would require caregivers to adapt the timing of the face reveal in accordance with the infant's developing conception of object permanence, visual perception, and communicative skills. These contingent adaptations suggest that the skills, knowledge and characteristics of each interactant change the nature of child-caregiver interactions and provide new affordances for social interactions (Mareschal, 2007; Oakes & Rakison, 2019; Thelen & Smith, 1994). There may also be relevant sociocultural norms that provide structural constraints on a game of peek-a-boo. Cultures may differ in whether caregivers obscure the viewpoint of themselves or the child, what they say at the reveal (e.g., '*peek-a-boo*', '*there you are*', '*bubu-settete*'), or whether briefly hiding behind a tree counts as an instance of peek-a-boo. Lastly, there are also fundamental biological constraints. Human perceptual limits for perceiving events place a lower bound on the time interval of caregivers covering their face (Dykiert et al., 2012; Radvansky & Zacks, 2011), the size of and individual variability in vocal organs place constraints on sound production (Belyk & McGettigan, 2022; Simpson, 2001), and the acoustic medium influences the transmission of sounds (Ladefoged, 1996; Medwin, 1975). These observations show how caregivers and children co-construct and co-regulate ecosocial situation spaces that emerge from complex interactions among variables that operate on varying timescales (Eliasoph & Lichterman, 2003; Fusaroli & Tylén, 2012; Steffensen & Pedersen, 2014; Tylén et al., 2010, 2013). The concrete realisations of the structural potentialities of child-caregiver interactions represent a dynamic compromise between the local history of child-caregiver interactions and considerations rooted in ontogeny, sociocultural norms and biological constraints (Bolis & Schilbach, 2018, 2020; Veissière et al., 2020).

3.2 How Does This Work Add to Our Knowledge?

Social contingency plays a role in all of these facets of the situation spaces formed by child-caregiver interactions (Bolis & Schilbach, 2020; Fusaroli & Tylén, 2012; Tylén et al., 2010, 2013). Building a scientific understanding of how real-time behavioural adaptations unfold in situated structural affordances requires consideration of this interplay of complex interactions. In the following sections, I discuss how the findings from our three studies add to our knowledge about how social affordances that emerge from ontogeny, sociocultural norms and biological factors constrain behavioural adaptations in early child-caregiver interactions.

3.2.1 Changes in Behavioural Adaptations During Ontogeny

How children interact with the world is in constant flux in early development, and children's developing skills and exploration of their environment determine, in part, the environment that they interact with (Clark & Estigarribia, 2011; Olson & Masur, 2011, 2012; Thelen & Smith, 1994; Twomey & Westermann, 2018). As children develop new capacities to move their bodies and engage in interactions with others (Adolph et al., 2012; Karasik et al., 2011, 2014), for example, their visual and verbal input changes (Needham et al., 2002; Soska et al., 2010; Soska & Adolph, 2014). When children start to manipulate and explore objects, caregivers exhibit a greater propensity to label those objects (Suarez-Rivera et al., 2022). Similarly, infants' transition from crawling to walking frees up their hands to carry objects to bid for caregiver's attention (Karasik et al., 2011, 2014) and offers new affordances for exploring their environment, engaging in joint attention and interacting with their caregivers (Smith et al., 2018; Yu & Smith, 2012). The environment is thus negotiated by the developing capacities of the child and their changing interactions with the environment (Mareschal, 2007; Oakes & Rakison, 2019).

In the context of child-caregiver vocal interactions, the developing knowledge, skills and behaviours of children likewise create new interactional affordances and require different types of behavioural adaptations. For example, caregivers change children's language environments when they align their verbal responses to children's developing interests and skills (Braarud & Stormark, 2008; Elmlinger et al., 2019, 2023; Goldstein & Schwade, 2008; Gros-Louis et al., 2006, 2014; Smith & Trainor, 2008). Developing behaviours thus shape the feedback children receive, but also change what they attend to; for example, children's individual production experience exerts an influence on their attention to speech sounds in the speech stream (DePaolis et al., 2011, 2013; Majorano et al., 2014) and creates individual trajectories of downstream development in babble and word production (Vihman, 1993, 2009; 2017). Children play an active role in shaping their own development by creating social affordances for caregivers to adapt their behaviour in ways tailored to a specific point in ontogeny (Hilbrink, 2015; Laing & Bergelson, 2020).

This co-constructive process of child development has crucial implications for our conception of changes in caregivers' behavioural adaptations during early ontogeny. In the meta-analysis of the acoustic features of IDS in ***Study I***, we argued that dynamic change in certain features may be reflective of caregivers' adaptations to the developmental states of their children. The decrease in vowel duration and increase in articulation rate during development – the latter feature also found in the cross-sectional sample for Danish IDS in ***Study II*** – may be a function of children's developing speech processing skills, as a slower articulation rate in IDS eases the cognitive load (Christiansen & Chater, 2016; Werker & Tees, 1999). Just like sitting independently allows children to manipulate objects (Needham et al., 2002; Perone et al., 2008; Soska et al., 2010; Soska & Adolph, 2014), then, children's developing speech processing skills may provide different social affordances for caregivers to adapt their speech. The role of social contingency in vocal exchanges may also explain the finding of a decrease in f_0 in IDS in the meta-analysis in ***Study I***. Caregivers may be reacting to children's changing preferences to attend to different acoustic features during early development (Panneton et al., 2006) and it may reflect the changing functions of IDS, from being more about conveying positive affect to being concerned with highlighting linguistic information (Fernald, 1992; Kitamura & Notley, 2009; McRoberts et al., 2009). Interpretation of the stability of f_0 variability and vowel space area during early development requires caution but may indicate ongoing developmental relevance, potentially in their roles of grabbing and directing attention and clarifying the input (Peter et al., 2016; Song et al., 2010).

Study III allows us to gain insights from both sides of the interaction. The highly tentative evidence for the quadratic developmental trajectory of child response latencies may reflect how children's developing linguistic skills has cascading effects on their interactional abilities and in turn influence the temporal structure of early child-caregiver interactions (Levinson, 2006, 2016). Importantly, strong correlations between child and adult response latencies in the findings indicate that the developing skills of children provide varying affordances for caregivers to adapt the timing of their vocal exchanges in ways that ensure mutual convergence and co-construction of a shared tempo (Hilbrink, 2015; Warlaumont et al., 2014).

From these findings, it is clear that ontogeny exerts important influences on behavioural adaptations in early child-caregiver interactions. Caregivers' adaptation of their behaviour to suit the developing skills of children means that children play an active role in shaping the input they receive (Elmlinger et al., 2019, 2023; Laing & Bergelson, 2020; Ritwika et al., 2020; Warlaumont et al., 2014). As a product of this co-construction of behavioural adaptations according to children's ontogenetic abilities, children exhibit a high degree of individual variability in behavioural, cognitive and linguistic skills (Bates et al., 2017; Bjorklund, 2022; Donnelly & Kidd, 2021; Bottema-Beutel & Kim, 2021). In **Section 3.3.2**, I argue that we need to take these individual differences seriously in developmental theory to obtain a higher resolution of the causal factors involved in these behavioural adaptations.

3.2.2 Undersampling, Over-interpreting: Behavioural Adaptations Across Cultures & Contexts

To better understand the impact of cross-cultural social dynamics on real-time behavioural adaptations, developmental research needs to incorporate speaker data from diverse languages and cultures (Christiansen et al., 2022; Kidd et al., 2023; Kidd & Garcia, 2022a; Trecca et al., 2021). Cultural differences emerge from cyclic processes of accumulating, changing and transmitting behaviours from earlier generations to establish bodies of knowledge (Birch & Heyes, 2021; Henrich, 2015; Heyes, 2018). Cultural learning is implicit and rapid (Boyd et al., 2011; Tomasello, 2009; Tomasello et al., 1993), and caregivers can thus quickly learn to emulate cultural norms and expectations for child-caregiver interactions by observing other caregivers (Bolis & Schilbach, 2020; Heyes, 2016; Calder, 2019).

As noted in the introduction, there is a striking diversity in caregiving practices around the world (Clegg et al., 2021; Keller et al., 2009; Keller et al., 1988; Konner, 2016; Little et al., 2016; Loukatou et al., 2022; Shneidman, 2010). These different interactive contexts provide different affordances for behavioural adaptation; for example, caregivers provide contingent responses to more vocalisations during book reading and puppet play (Gros-Louis et al., 2016), and toy play often involves more directives and fewer labelling and conversation-eliciting utterances (Blake et al., 2006; Gros-Louis et al., 2016; Hoff-Ginsberg, 1991). These dimensions of cultural variation are consequential because many theoretical formulations of communicative exchanges in early child-caregiver interactions rely primarily on face-to-face communication and visually mediated joint attention – the child-centred style of social interaction most characteristic of industrialised cultures (Harkness & Super, 2002; Keller et al., 1988; Lancy et al., 2010; Little et al., 2016).

The systematic reviews and meta-analyses in ***Study I*** and ***Study III*** in this dissertation showed a widespread sampling bias in the literature on IDS and turn-taking. It should be noted that neither the turn-taking meta-analysis nor the IDS meta-analysis disentangled culture from language, despite each variable likely having separate influences on the dynamics of child-caregiver interactions. For example, it may be the case that pitch variability would be less available for caregivers to use in tonal languages, or that the cultural practice of rarely replying to children's emotional displays changes the acoustic expression of IDS in that culture. The same is true in ***Study II***; the unique vocalic properties of IDS could in principle be caused both by the sound structure of IDS or stigma against using baby language in Danish society (cf., Johnston & Wong, 2002). These examples highlight the multifaceted nature of culture and its use as an umbrella term for different constellations of community practices, languages, family structures, economic practices, modes of subsistence, and so on (Crocetti & Salmela-Aro, 2018; Rogoff & Angelillo, 2002). Research also needs a holistic construal of culture for convenience of analysis and to become aware of the dimensions of variability in human cultural practices (Christiansen et al., 2022). However, to improve our understanding of the causal structures of child-caregiver IDS and turn-taking patterns, we also need narrower hypotheses regarding which parts of cultural variation may contribute to a pattern, as argued further in **Section 3.3.3**.

By only being able to include turn-taking data on six distinct languages, the meta-analysis in ***Study III*** highlighted the considerable limitations in our knowledge of how the temporal structure of early child-caregiver interactions differs across cultures. This is especially the case when considering the substantial confounding of contextual variables; for example, all studies of non-English child interactions with unfamiliar adults also use free play as an experimental task, making it impossible to attribute any differences to their relevant conditions (Jager et al., 2008). If we look to the adult literature on turn-taking, there is also little systematic comparison of fine-grained turn-taking structures across languages and cultures (Dingemanse & Enfield, 2023; Stivers, 2009). And yet, based on this small set of mainly Indo-European languages, the early development of turn-taking in ontogeny has been argued to be the product of a universal interactional system underlying the human capacity for language (Levinson, 2006, 2016). Because these studies focus on a single sliver of the different constellations of ways that children and caregivers interact with each other across cultures, however, only highly tentative inferences can be made based on the available evidence at the current point in time. The findings in ***Study III*** argue for a greater caution in interpreting the results as reflecting universal human capacities until we have a base of evidence that is more representative of different cultural norms and practices, as discussed further in **Section 3.3.3**.

The aggregation of the scarce evidence that was available from studies on turn-taking showed a robust influence of interactive context on child response latency; for example, child response latencies in question-answer pairs showed consistently longer response times than those in spontaneous conversations or free play situations. The finding that context plays a crucial role in the temporal structures of interaction carries important implications for the diverse range of interactive contexts afforded by different cultural norms (Loakes et al., 2013; Loukatou et al., 2022; Shneidman & Goldin-Meadow, 2012). The context dependence of interactions matters for how we plan future cross-cultural studies of how child turn-taking develops, as discussed further in **Section 3.3.3**.

The meta-analysis on IDS in ***Study I*** included a greater number of languages but still exhibited a high degree of oversampling from WEIRD cultures and languages, which influences our conception of IDS acoustics and their importance during early development. The diverse range of interactive contexts would also provide different affordances for caregivers to use IDS (Broesch et al., 2016; Kärtner et al., 2016). Despite the variation in cultural practices and sound structures across the languages included in the meta-analysis, the findings indicated robust tendencies in certain acoustic features; prosodic modulation, for example, showed robustness, with caregivers speaking with a higher f_0 and f_0 variability as well as a slowed articulation rate across the vast majority of languages and cultures. In **Section 3.2.3**, I discuss potential underlying biophysical and cultural reasons for these systematic patterns of behavioural adaptations.

The meta-analysis in ***Study I*** also provided some evidence indicating that interactive context (i.e., experimental task and recording environment) influences the acoustics of IDS; the f_0 and f_0 variability in spontaneous speech was higher than read speech, and f_0 and f_0 variability were lower in home recording environments than in laboratory environments. These systematic differences in the estimates highlight the inherent limitations of relying on the results of a single study, as discussed in **Section 1.3**. As a concrete example of this, the analysis of Danish IDS in ***Study II*** only investigated the acoustic features of spontaneous speech that was recorded in a home environment and was elicited in a free play context. These specific contextual constraints may provide certain affordances for caregivers to change only certain aspects of their behaviour, and thus the results may not generalise to other play contexts (e.g., read speech or for recordings elicited in a laboratory context). These study-specific sources of variability matter for our theoretical conceptions and speak in favour of the value of meta-analyses in their ability to aggregate results across different methodological intersections of studies to evaluate generalisability (Devezer & Buzbas, 2023; Zettersten, Cox et al., 2023).

3.2.3 Biology as a Constraint on Behavioural Adaptations and Culture

There are few aspects of human culture that have not been shaped by our biology, and in turn, human culture has also shaped our biology (Smith, 2020; Creanza et al., 2017). Cultural and biological environments of interactions are thus fundamentally intertwined and constitute a multifaceted constellation of structural affordances that can guide real-time human interactions (Bolis & Schilbach, 2020; Creanza et al., 2017; Fusaroli & Tylén, 2012; Tylén et al., 2010, 2013).

In the context of child-caregiver interactions, biophysical constraints encompass the affordances imposed on our sensorimotor engagements with the world through fundamental limits to our perceptual sensitivity and production capacities (Bolis & Schilbach, 2020; Creanza et al., 2017). These general principles of perceptual processing and memory limitations shape the temporal structure of language and interaction (Christiansen & Chater, 2016; Levinson, 2006). For adult caregivers, this entails that there are upper limits to the speed of speech production (Lammert et al., 2018) and lower bounds on response times (Dykiert et al., 2012; Radvansky & Zacks, 2011). The change in the cognitive abilities of children during early development can to a certain extent similarly be construed as being dependent on biocognitive constraints that change both their interactions with the world and the world itself (Adolph et al., 2012; Karasik et al., 2011, 2014); for example, cognitive developments in speech processing influence how children listen to the speech stream, but also how fast caregivers speak with them (Lee et al., 2014; Narayan & McDermott, 2016; Warlaumont et al., 2014).

There are also dimensions of the interactive environment that are determined by their physical properties alone; for example, the f_0 of a vibrating object is inversely related to the mass of the vibrating membrane and higher-intensity sounds tend to come from larger objects compared to sounds of lower intensity (Johnson, 2004; Ladefoged, 1996; Stevens, 2000). These fundamental associations between sounds and objects have been invoked to explain the high prevalence of sound symbolism across human languages (Emmorey, 2014; Sidhu & Pexman, 2019), as well as the higher than average extent of iconicity and onomatopoeia in child-directed input, which have been argued to facilitate early language development (Imai & Kita, 2014; Laing, 2019; Massaro & Perlman, 2017).

The finding of clear commonalities in the pitch properties of IDS across languages and cultures in ***Study I*** may derive from biophysical constraints and form-function relationships in child-caregiver communicative signals (Bryant, 2022; Bryant & Barrett, 2007; Pisanski et al., 2022). Because the rate of vibration of a membrane relates to its mass, and because the mass of vocal cords in mammals correlates with overall body mass, a higher rate of vibration may serve to convey the impression of a small, non-threatening physical size (Morton, 1977; Ohala, 2008). Similarly, the lilting rhythms and fluctuating melodies of IDS and infant-directed songs (Bryant & Barrett, 2007; Fernald, 1992; Hilton et al., 2022) convey gentle, tone-like sounds that contrast with the irregular and rough voice quality that has a higher likelihood of arising due to secondary vibrations in higher-mass membranes (Johnson, 2004; Morton, 1977; Ohala, 2008). In this sense, the acoustics of IDS relies on the opposite mechanism to that of cats arching their backs or birds extending their wings and fanning out their tail feathers to appear as large as possible in their receivers' visual field to convey a large, physically intimidating size. These fundamental biophysical constraints on behavioural adaptations may underlie the strong tendency for caregivers speaking different languages to modify the acoustics of their voices in similar directions, especially in terms of pitch properties to appear non-threatening (Fernald, 1989; Hilton et al., 2022).

However, ***Study I*** also showed that other acoustic features of caregivers' IDS exhibited a greater degree of cross-cultural variability and context dependence. Behavioural adaptations that exhibit a high degree of cross-cultural variation may to a greater extent represent changes that have emerged as adaptations to specific socio-historical contexts on a shorter timescale (Boyd et al., 2011; Eliasoph & Lichterman, 2003; Steffensen & Pedersen, 2014). For example, caregivers in a culture that places high value on verbal skills and educational achievement may to a greater extent employ IDS with the purpose of clarifying the input and therefore require a different acoustic expression of IDS. The cultural evolution literature shows that normative constraints can emerge variably for different cultures on a rapid timescale (Henrich & McElreath, 2003; Heyes, 2018; Mesoudi & Whiten, 2004) and may produce a greater degree of variability in behavioural adaptations across cultures.

Cross-cultural variation in the acoustic features of IDS may also stem from the specific sound structure of different languages; that is, certain properties of languages may provide different affordances for interactants to carry out behavioural adaptations (Christiansen et al., 2022; Trecca et al., 2021). As proposed in ***Study II***, a greater degree of vowel variability in Danish IDS compared to ADS may aid children in the long term by increasing the robustness of their category generalisations (Raviv et al., 2022). These types of constraints on behavioural adaptations that come from the sound structure of languages can be considered shaped by both culture and biology. Languages are cultural objects in the sense of being passed down generations through cultural transmission, but are also fundamentally biological objects that capitalise on physiological and perceptuo-motor affordances from the human body (Smith, 2020; Stevens, 1989). Cultural and biological environments of interactions thus create a set of structural affordances that can guide and scaffold real-time human interactions, which in turn feed back into the dynamic formation of these constraints (Bolis & Schilbach, 2020; Tylén et al., 2013).

The early onset of sophisticated turn-taking structures in infancy in ***Study III*** has also been argued to reflect biological constraints in our cognitive architecture and corresponding abilities to produce and process vocal material in interaction (Levinson, 2006, 2016). The transience of the speech signal and incrementality of our speech processing abilities creates fundamental pressures that shape the structures of vocal exchanges across cultures (Christiansen & Chater, 2016; Levinson, 2016; Stivers, 2009). Structural stabilities in interaction are thus bound to our biological bodies and physical environments, but this does not diminish the crucial role of culture and sound structure in determining behavioural adaptations as well; for example, the peculiar sound structure of Danish may require different processing strategies to other languages, and there is some evidence that these strategies leave a trace on adult speech perception systems (Ishkhanyan et al., 2020; Trecca et al., 2021). The opaque sound structure of Danish may thus influence the turn-taking structures of child-caregiver interactions, requiring a greater period of time for children to respond (Stivers, 2009), or require specific acoustic adaptations not required by other languages (Cox et al., 2022).

Interim Conclusion

The findings from the studies included in the present dissertation indicate that the concrete realisations of potential structures of child-caregiver interactions emerge from compromises among the unique history of interactions between the child and caregiver and considerations rooted in ontogeny, sociocultural norms and biological constraints (Fusaroli, Gangopadhyay, et al., 2014; Fusaroli, Rączaszek-Leonardi, et al., 2014; Tylén et al., 2013; Veissière et al., 2020). Where do we go from here? How do we most efficiently dedicate resources to improve our theories about the mechanisms underlying child-caregiver behavioural adaptations? I turn to these questions in the next section and map out future directions of research.

3.3 Where Do We Go From Here?

Perhaps unsurprisingly, the findings in these three individual studies pose more questions than they answer. This is a natural part of the scientific process; as our knowledge grows, so does the breadth of questions we can pose about that knowledge. As argued in the introduction, cumulative science practices can help us move our knowledge forward by mapping out the sampling space of previous studies, evaluating their generalisability, identifying their blind spots, limitations, and remaining open questions, and proposing potential mechanisms and causal pathways to investigate further (Fusaroli et al., 2022; Zettersten, Cox et al., 2023). Here, I argue that the three studies in this dissertation point to five distinct trajectories of future research:

I. Improved resolution of behavioural adaptations on a turn-by-turn basis;
II. Detailed measures of individual differences to capture variability in trajectories of child development;
III. Theory-driven studies of behavioural adaptations across cultures and languages;
IV. Methodological cross-pollination from models of interactions in non-human animals;
V. Leveraging cumulative science and theory to explore new questions in a systematic and iterative manner.

I will discuss each of these strands of research in turn and argue for how they can contribute to our knowledge about early child-caregiver behavioural adaptations in future investigations.

3.3.1 Improved Resolution of Behavioural Adaptations on a Turn-By-Turn Basis

Making progress in our investigations of behavioural adaptations in early child-caregiver interactions requires understanding both how the knowledge of child and caregiver is brought to bear in a task in a moment in time and how these everyday second-by-second activities accumulate to influence changes over longer timescales (Oakes & Rakison, 2019; Thelen & Smith, 1994). The method of aggregating results in meta-analyses in ***Study I*** and ***Study III*** in this dissertation meant that we could mainly paint developmental patterns with broad paint strokes. Dynamic changes in the acoustic features of IDS and child turn-taking skills were based on results from individual studies, each often investigating a single point in developmental time. Similarly, in ***Study II***, we relied on a cross-sectional sample where developmental patterns may be confounded by individual differences between participants (Maxwell & Cole, 2007; Robinson et al., 2005). In ***Study III***, we found a correlation between adult and infant response latencies in the meta-analysis of turn-taking suggesting that children and caregivers co-construed shared rhythms of interaction by adapting the timing of their vocal exchanges. However, crucially, these results cannot comment on the dynamics of co-adaptation on a turn-by-turn basis and how the split-second world of local interactions creates change on longer timescales of development. Incorporating these multiple timescales into our statistical models of child-caregiver adaptations (e.g., models such as dyadic coupling models which implement theory-driven causal mechanisms of intra- and interpersonal adjustment) would provide a more nuanced perspective on the dynamic ebb-and-flow waves of mutual convergence and scaffolding that takes place during local transactions in early child development (Ritwika et al., 2020; Warlaumont et al., 2014).

Focusing on a turn-by-turn, local timescale in different contexts would also permit greater insight into the functions of behavioural adaptations; for example, child response latencies may scale with the varying informational complexity of the caregiver's utterance (Casillas et al., 2016), or caregivers may consistently produce utterances with an expanded vowel space area in contexts where speech intelligibility aids the communicative goal (e.g., a labelling scenario in a book reading activity (Lovcevic et al., 2022)), or the pitch variability of individual utterances may serve to grab and direct infant attention on a turn-by-turn basis to guide learning of new words (Nencheva, Piazza & Lew-Williams, 2021; Nencheva & Lew-Williams, 2022). Future research should thus to a greater extent examine how local communicative exchanges create a set of social affordances that scaffold, on a turn-by-turn basis, the co-creation of behavioural coordination (Elmlinger et al., 2019, 2023; Fusaroli et al., 2019).

3.3.2 Detailed Measures of Individual Differences to Capture Variability in Trajectories of Child Development

Children's individual developmental trajectories exhibit a high degree of variability, and development branches into an infinite number of directionalities both within and beyond the child (Oakes & Rakison, 2019; Thelen & Smith, 1994). The majority of studies in child development rely on comparing children at different ages and using results to infer patterns about how behavioural adaptations change as a result of developing skills. Relying on age as a descriptive proxy for development *in toto* likely conceals many of the substantial individual differences in social, linguistic and motor development among children. Child age was used as a measure of development in the meta-analyses in ***Study I & II*** because there was not sufficient uniformity in social measures in an adequate amount of studies for us to meta-analyse them. Though age as a variable allows us to capture that change of some kind is happening, it cannot provide insights on the mechanisms of change. The findings of a longitudinal study, for example, may indicate that the response latency of children's vocalisations slows down between 18 to 22 months of age; however, relying on the measure of age in this context in and of itself does not allow us to probe underlying causes of this change, and interpretations of age effects are notoriously difficult and often problematic (cf., Bergmann et al., 2018). Tracing when skills emerge based on interpreting both significant and null findings is challenging due to the impossibility of disentangling whether a null effect is a product of the absence of a given skill, low experimental power, measurement noise, or any of the other dimensions of hidden variability (Bergmann et al., 2018; Cruz Blandón et al., 2023; Koile & Cristia, 2021; Yarkoni, 2022). As argued further in **Section 3.3.5**, meta-analytic methods partially ameliorate this situation by offering a more systematic approach to statistically evaluating whether age explains large amounts of variance, and if not, to delineate the limitations of our knowledge (Zettersten, Cox et al., 2023).

Incorporating rich measures of individual differences into our models in future research would provide us with more opportunities to theorise about the cognitive mechanisms underlying these complex skills. The importance of individual differences applies to both child and caregiver. For example, individual differences in the social, motor and linguistic skills of children may alter how they interact with caregivers as well as the input that they receive from caregivers, as shown by studies of caregiver interactions with cochlear-implanted children (Dilley et al., 2020; Kondaurova et al., 2013, 2020; Kondaurova & Bergeson, 2011; Wieland et al., 2015) and autistic children (Leezenbaum et al., 2014; Neimy et al., 2017; Warlaumont et al., 2014). Although based on a scarce base of data, the turn-taking meta-analysis in ***Study***

III also showed evidence that atypically developing infants exhibited slower response latencies compared to typically developing children. On the other side of the interaction, caregivers with postpartum depression exhibit different prosodic properties in IDS (Kaplan et al., 2001; Lam-Cassettari & Kohlhoff, 2020; Porritt et al., 2014), and some subtle qualitative differences in behavioural adaptations depend on the interaction styles of caregivers (Amano et al., 2006; Benders et al., 2021; Ferjan Ramírez, 2022). Incorporating these individual differences among participants into our models of development would provide a fuller picture of how the characteristics of both children and caregivers contribute to the nature of the interaction and change as a result of the interaction (Gómez & Strasser, 2021; Oakes & Rakison, 2019).

The inclusion of individual differences would also provide crucial theoretical insights into directionalities of the behavioural adaptations: specifically, whether the adaptation should be considered speaker-centric or listener-centric. For example, caregivers' adaptations may revolve around accomplishing a speaker-centric goal (e.g., expression of affect or friendliness) or a listener-centric goal (e.g., helping their children to process speech by slowing down their production). This tension between speaker- and listener-centric adaptations underlies many arguments about the functions of IDS. For example, the heightened f_0 variability of early IDS may relate both to its speaker-centric role in enabling caregivers to express emotions and convey affect (Fernald, 1992; Trainor et al., 2000) and to its listener-centric role in grabbing and directing infant attention (Nencheva, Piazza & Lew-Williams, 2021; Nencheva & Lew-Williams, 2022; Räsänen et al., 2018). Both speaker- and listener-centric behavioural adaptations still rely crucially on the social contingency of the interaction, but this distinction has crucial implications for any functional explanations that can be proposed for individual features. For example, vowel space expansion can be argued to reflect both directionalities; it may serve the speaker-centric role of conveying non-threatening behaviour (Kalashnikova et al., 2017) as well as the listener-centric role of improving speech intelligibility or modelling the infant vocal tract (Hartman et al., 2017; Kuhl et al., 1997; Liu et al., 2003).

There is no reason for these distinct directionalities not to co-exist and be variably relevant during early ontogeny; for example, the core function of vowel space expansion for young children who are initially unable to decode linguistic messages may to a greater extent be speaker-centric to convey a non-threatening attitude and express positive emotions (Fernald, 1989). In contrast, caregivers interacting with older children who are becoming more advanced in language development may be more concerned with the listener-centric function of ensuring the intelligibility and clarity of vowels in the speech stream. Other behavioural adaptations to a greater extent reflect listener-centric accommodation; for example, a slower articulation rate likely accommodates children's initial limitations in speech processing, and likewise for turn-taking, both children and adults adapt the timing of their vocal exchanges in ways that resemble mutual convergence (Hilbrink, 2015; Ritwika et al., 2020; Warlaumont et al., 2014).

The directionalities of behavioural adaptations likely also depend on the individual social abilities of the interactants. There is little investigation into how the subtle individual differences of both the child and caregiver influence behavioural adaptations during early ontogeny, but given that listener-centric adaptations require social abilities such as theory-of-mind, there is good reason to suspect that these properties matter (Bolis & Schilbach, 2018; Fusaroli et al., 2019; Kennedy & Adolphs, 2012). One way to investigate these directionalities and the extent to which these change during development, then, may be to identify constructs for individual differences and model their interactions in statistical models. For example, caregivers with high scores on theory-of-mind measures may be better at responding to limitations in child skills and produce stronger adaptations as a result (e.g., a slower speech rate to aid speech processing, expanded vowel space, and a slower response latency to allow the child to take part in the interaction). By examining the variability introduced by the social, linguistic and motor skills of each of the interactants in early child-caregiver communicative exchanges, then, we can provide a fuller picture of the causal mechanisms and directionalities underlying behavioural adaptations.

3.3.3 Theory-driven Studies of Behavioural Adaptations Across Cultures and Languages

As shown in our meta-analyses of IDS and turn-taking, the insights we have about these developmental phenomena rely on a small number of languages and very similar sample populations with WEIRD characteristics (Evans & Levinson, 2009; Kidd & Garcia, 2022b; Christiansen et al., 2022; Trecca et al., 2021). An expansion in the diversity of cultures and languages would offer a richer foundation to base our understanding on; however, an attempt to simply expand diversity in and of itself is likely to be very slow and very dispendious with potentially relatively little return on investment in the short term (Christiansen et al., 2022). How can we allocate resources in the most efficient way to maximise our scientific knowledge about behavioural adaptations in early child-caregiver interactions? Although we should leave the door open to serendipity and unexpected findings – which also play a crucial role in investigating regions of the methodological space in an iterative and exploratory manner (Devezer et al., 2019) – contributions to our understanding of child-caregiver behavioural adaptations requires theory-driven, targeted comparisons of a set of cultural contexts that differ along useful dimensions (Legare & Nielsen, 2015; Legare & Harris, 2016). This could include breaking down the multi-faceted nature of culture into some of its component variables. Causal inference about specific cross-linguistic variables that moderate child-caregiver behavioural adaptations could be facilitated by keeping unobserved cultural variables similar, such as child-rearing practices or method of subsistence. Or vice versa, we can isolate the effect of cultural variables (e.g., child-rearing practice or subsistence method) by examining child-caregiver dyads in distinct cultures that speak the same language. Targeted, theory-driven comparisons within and between multiple groups can thus isolate particular cultural or linguistic variables of interest as causal factors (Deffner et al., 2021).

3.3.4 Methodological cross-pollination from models of interactions in non-human animals

Communicative exchanges play a crucial role in the lives of many biological organisms and exhibit a high degree of species-specific adaptations (Verga et al., 2023; Pika et al., 2018). A cross-species angle on behavioural adaptations can help us specify our theoretical predictions more explicitly and contextualise results from child-caregiver vocal exchanges in humans. Our investigations of human interactions often carry implicit assumptions that derive from our own experience as social interactants. As discussed in ***Study III***, for example, overlaps are often considered failures of turn-taking, and studies therefore often exclude response latencies above and below certain thresholds (cf., Rifai et al., 2022), despite research showing that successful communication and rapport depends on interactants' frequent use of backchanneling and synchrony (Cho et al., 2023; Cummins, 2019; Dideriksen et al., 2023). Analyses of turn-taking in non-human animals often use methods like dataset rotation to ensure non-randomness in temporal structures (Deruiter & Solow, 2008; Kershenbaum et al., 2014) and computational formulations that allow for explicit encoding and refinement of theoretical predictions (Greenfield, 1994; Ravignani & de Reus, 2019; Takahashi, 2013). This explicit encoding of prior theoretical expectations could also be useful for investigating how turn-taking develops in children. For example, the Interaction Engine Hypothesis (Levinson, 2006, 2016) predicts that a slowdown will happen together with the parallel development of language. Because the link between this conceptual idea and quantitative predictions is underspecified, there can be many observations that constitute evidence for or against the claim (Scheel, 2022). By explicitly encoding our prior theoretical expectations of developmental trajectories in computational models, we would have a better basis on which to find supporting or contradicting evidence – and in turn, we can refine and further develop the computational model of the underlying patterns in light of new evidence (cf., Guest & Martin, 2021).

Investigating the cross-species generalisation of patterns would also allow for systematic exploration of behavioural adaptations through a shared computational framework (Cox, Templeton & Fusaroli, 2023). For example, we may see qualitative similarities or differences in the form and functions of turn-taking structures in humans, bush crickets (Greenfield, 1994) and harbour seals (Ravignani, 2019; Ravignani et al., 2023). In turn, this can facilitate inferences about the fundamental social capacities required for this complex feat of cognition (Bolis & Schilbach, 2020; Levinson, 2006; Verga et al., 2023). Models of how conspecifics in the animal world engage in juvenile-directed vocalisations may also advance our theories about how different degrees of sociality and cooperative behaviours can shape the dynamic interplay between interactants in child-caregiver interactions (Dumas, 2021). These insights can also go in the other direction, from the human to the non-human world of interaction; exploring human language permits deeper insights into how various situational contexts and levels of informational complexity can influence behavioural adaptations (Casillas et al., 2016; Dideriksen et al, 2023; Fusaroli et al 2013, 2014; Templeton et al., 2022, 2023). Cross-species comparisons and methodological cross-pollination, then, can improve our understanding of how human interactive behaviour fits into the space of potential interactive structures in the animal kingdom (Cox, Templeton & Fusaroli, 2023).

3.3.5 Cumulative Perspectives & Theory Development

Science is most effective when approached as a collaborative and cumulative effort of theory development that strives to establish consistent patterns across various populations, conditions, and contexts, all while acknowledging the limitations inherent in making such generalisations (Zettersten, Cox et al., 2023; ManyBabies Consortium, 2020). Theories create new research questions and build expectations, and new observations in turn lead to theory development (Devezer & Buzbas, 2023). As discussed in **Section 1.3**, results from studies serve an instrumental role in arriving at better theories (Buzbas et al., 2023; Devezer et al., 2019; Devezer & Buzbas, 2023). For example, the IDS preference effect does not explain why children prefer to attend to this speech style (e.g., Segal & Newman, 2015; Fernald & Kuhl, 1987). This is simply the effect, and the job of developmental psychologists is to ask *why* children prefer to attend to IDS over ADS. In other words, science is about formulating unifying explanations that allow us to generate testable hypotheses and predictions that distinguish between competing theories (Guest & Martin, 2021; Devezer et al., 2019).

One way to assess where the evidence stands, discover current knowledge lacunae, and determine the most informative next steps for future research is to engage in critical self-reflection via cumulative science practices (ManyBabies Consortium, 2020; Zettersten, Cox et al., 2023). However, these methods should be viewed in light of their service to theory development. Without a theoretical grounding to direct empirical research, empirical results are unlikely to cumulatively contribute to existing bodies of knowledge (Guest & Martin, 2021, 2023; Chalmers, 2013). Crafting a scientific theory is like composing a symphony; each observation is a musical note, but amassing random notes alone does not create a symphony – rather, the essence of science lies in organising and structuring empirical results into a comprehensive framework that allows us to derive expectations and produce knowledge. Scientific theory development requires us to engage in critical reflection on how to incorporate emerging insights into our existing knowledge framework and to identify the most informative next steps. As discussed in **Section 1.3.2** and explored in ***Study II***, by using Bayesian methods of posterior passing, we can foster a critical and productive conversation between prior and new results that can help us contextualise our findings and cumulatively improve our scientific knowledge (Fusaroli et al., 2022; Brand et al., 2019).

Interim Conclusion

These research strands propose that child-caregiver interactions can be better understood by i) investigating behavioural adaptations on multiple timescales; ii) incorporating individual differences; iii) conducting theory-driven cross-cultural studies; iv) contextualising results through cross-species comparisons; and v) developing explicit scientific theories. Cumulative science practices serve as useful tools for all of these distinct trajectories of research by mapping out the sampling space, and proposing potential mechanisms and causal pathways to investigate further (Fusaroli et al., 2022; Zettersten, Cox et al., 2023).

4 Concluding Remarks

The findings presented in the three studies here provide compelling evidence that both the speech content and temporal structure of early child-caregiver interactions originate in the mutual feedback loops between interactants (Dumas, 2021; Fusaroli et al., 2019; Hilbrink, 2015). These behavioural adaptations are shaped not only by individual interactants but also by social affordances from the developing capacities of the child, sociocultural norms and biophysical constraints (Bolis & Schilbach, 2018, 2020; Fusaroli et al., 2019; Fusaroli & Tylén, 2012; Tylén et al., 2013). In early child-caregiver interactions, each interactant changes with every local interaction and enters the next round of interaction as a new person with new interactional affordances (Gómez & Strasser, 2021; Oakes & Rakison, 2019; Thelen & Smith, 1994; Vygotsky, 1978). These studies pose more questions than they answer. What I hope for this dissertation is that it can serve as a roadmap for future research directions and can contribute towards an integrated empirical-theoretical paradigm of behavioural adaptations in early child-caregiver interactions.

List of References

Abu-Zhaya, R., Seidl, A., & Cristia, A. (2017). Multimodal infant-directed communication: How caregivers combine tactile and linguistic cues. *Journal of Child Language*, *44*(5), 1088–1116.

Adolph, K. E., Cole, W. G., Komati, M., Garciaguirre, J. S., Badaly, D., Lingeman, J. M., Chan, G. L., & Sotsky, R. B. (2012). How do you learn to walk? Thousands of steps and dozens of falls per day. *Psychological Science*, *23*(11), 1387–1394.

Adriaans, F., & Swingley, D. (2012). Distributional learning of vowel categories is supported by prosody in infant-directed speech. *Proceedings of the Annual Meeting of the Cognitive Science Society 34*(34). https://escholarship.org/uc/item/7pt904fz

Albin, D. D., & Echols, C. H. (1996). Stressed and word-final syllables in infant-directed speech. *Infant Behavior and Development*, *19*(4), 401–418.

Amano, S., Nakatani, T., & Kondo, T. (2006). Fundamental frequency of infants' and parents' utterances in longitudinal recordings. *The Journal of the Acoustical Society of America*, *119*(3), 1636–1647. https://doi.org/10.1121/1.2161443

Aron, A. R., & Poldrack, R. A. (2006). Cortical and subcortical contributions to stop signal response inhibition: Role of the subthalamic nucleus. *Journal of Neuroscience*, *26*(9), 2424–2433.

Auspurg, K., & Brüderl, J. (2021). Has the credibility of the social sciences been credibly destroyed? Reanalyzing the "Many Analysts, One Data Set" Project. *Socius*, *7*, 23780231211024421.

Barthel, H., & Quené, H. (2015). *Acoustic-phonetic properties of smiling revised: Measurements on a natural video corpus*. Proceedings of the 18th international congress of phonetic sciences.

Basbøll, H. (2005). *The phonology of Danish*. Oxford University Press.

Bates, E., Dale, P. S., & Thal, D. (2017). Individual differences and their implications for theories of language development. *The Handbook of Child Language*, 95–151.

Baumgartner, H. A., Alessandroni, N., Byers-Heinlein, K., Frank, M. C., Hamlin, J. K., Soderstrom, M., Voelkel, J. G., Willer, R., Yuen, F., & Coles, N. A. (2023). How to build up big team science: A practical guide for large-scale collaborations. *Royal Society Open Science*, *10*(6), 230235.

Belyk, M., & McGettigan, C. (2022). Real-time magnetic resonance imaging reveals distinct vocal tract configurations during spontaneous and volitional laughter. *Philosophical Transactions of the Royal Society B*, *377*(1863), 20210511.

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., 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.

Bennett, C. M., Miller, M. B., & Wolford, G. L. (2009). Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction. *Neuroimage*, *47*(Suppl 1), S125.

Beppu, A., & Griffiths, T. (2009). Iterated learning and the cultural ratchet. *Proceedings of the Annual Meeting of the Cognitive Science Society, 31*(31).

Bergelson, E., Casillas, M., Soderstrom, M., Seidl, A., Warlaumont, A. S., & Amatuni, A. (2019). What do North American babies hear? A large-scale cross-corpus analysis. *Developmental Science*, *22*(1), e12724.

Bergelson, E., & Swingley, D. (2018). Young infants' word comprehension given an unfamiliar talker or altered pronunciations. *Child Development*, *89*(5), 1567–1576.

Bergmann, C., Tsuji, S., Piccinini, P. E., Lewis, M. L., Braginsky, M., Frank, M. C., & Cristia, A. (2018). Promoting replicability in developmental research through meta-analyses: Insights from language acquisition research. *Child Development*, *89*(6), 1996–2009.

Birch, J., & Heyes, C. (2021). The cultural evolution of cultural evolution. *Philosophical Transactions of the Royal Society B*, *376*(1828), 20200051.

Bjorklund, D. F. (2022). *Children' s thinking: Cognitive development and individual differences*. Sage publications.

Blake, J., Macdonald, S., Bayrami, L., Agosta, V., & Milian, A. (2006). Book reading styles in dual-parent and single-mother families. *British Journal of Educational Psychology*, *76*(3), 501–515.

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

Blumstein, D. T., & Récapet, C. (2009). The sound of arousal: The addition of novel non-linearities increases responsiveness in marmot alarm calls. *Ethology*, *115*(11), 1074–1081.

Bohn, O.-S. (2013). Acoustic characteristics of Danish infant directed speech. *Proceedings of Meetings on Acoustics, 19*(1), 060055.

Bolis, D., & Schilbach, L. (2018). Observing and participating in social interactions: Action perception and action control across the autistic spectrum. *Developmental Cognitive Neuroscience*, *29*, 168–175.

Bolis, D., & Schilbach, L. (2020). 'I interact therefore I am': The self as a historical product of dialectical attunement. *Topoi*, *39*, 521–534.

Bombar, M. L., & Littig Jr, L. W. (1996). Babytalk as a communication of intimate attachment: An initial study in adult romances and friendships. *Personal Relationships*, *3*(2), 137–158.

Bornstein, M. H., & Putnick, D. L. (2012). Stability of language in childhood: A multiage, multidomain, multimeasure, and multisource study. *Developmental Psychology*, *48*(2), 477-491.

Bortfeld, H., & Morgan, J. L. (2010). Is early word-form processing stress-full? How natural variability supports recognition. *Cognitive Psychology*, *60*(4), 241–266.

Bottema-Beutel, K., & Kim, S. Y. (2021). A systematic literature review of autism research on caregiver talk. *Autism Research*, *14*(3), 432–449

Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J. A., & Adcock, R. A. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, *582*(7810), 84–88.

Boyd, R., & Richerson, P. J. (2009). Culture and the evolution of human cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1533), 3281–3288.

Boyd, R., Richerson, P. J., & Henrich, J. (2011). Rapid cultural adaptation can facilitate the evolution of large-scale cooperation. *Behavioral Ecology and Sociobiology*, *65*, 431–444.

Braarud, H. C., & Stormark, K. M. (2008). Prosodic modification and vocal adjustments in mothers' speech during face-to-face interaction with their two-to four-month-old infants: A double video study. *Social Development*, *17*(4), 1074–1084.

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.

Brand, C. O., Ounsley, J. P., van der Post, D. J., & Morgan, T. J. H. (2019). Cumulative Science via Bayesian Posterior Passing: An introduction. *Meta-Psychology*, *3*. https://doi.org/10.15626/MP.2017.840

Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for 'motionese': Modifications in mothers' infant-directed action. *Developmental Science*, *5*(1), 72–83.

Brischoux, F., & Angelier, F. (2015). Academia's never-ending selection for productivity. *Scientometrics*, *103*, 333–336.

Broesch, T., & Bryant, G. (2018). Fathers' infant-directed speech in a small-scale society. *Child Development*, *89*(2), e29–e41.

Broesch, T., Rochat, P., Olah, K., Broesch, J., & Henrich, J. (2016). Similarities and differences in maternal responsiveness in three societies: Evidence from Fiji, Kenya, and the United States. *Child Development*, *87*(3), 700–711.

Broesch, T. L., & Bryant, G. A. (2015). Prosody in infant-directed speech is similar across western and traditional cultures. *Journal of Cognition and Development*, *16*(1), 31–43.

Bruner, J. S. (1974). From Communication to Language—A Psychological Perspective. *Cognition*, *3*(3), 255–287. https://doi.org/10.1016/0010-0277(74)90012-2

Bryant, G. A. (2020). Evolution, structure, and functions of human laughter. *The Handbook of Communication Science and Biology*, *1*, 63–77.

Bryant, G. A. (2022). Vocal communication across cultures: Theoretical and methodological issues. *Philosophical Transactions of the Royal Society B*, *377*(1841), 20200387. https://doi.org/10.1098/rstb.2020.0387

Bryant, G. A., & Barrett, H. C. (2007). Recognizing intentions in infant-directed speech: Evidence for universals. *Psychological Science*, *18*(8), 746–751. https://doi.org/10.1111/j.1467-9280.2007.01970.x

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science*, *296*(5572), 1435–1435.

Buzbas, E. O., Devezer, B., & Baumgaertner, B. (2023). The logical structure of experiments lays the foundation for a theory of reproducibility. *Royal Society Open Science*, *10*(3), 221042.

Byers-Heinlein, K., Tsui, A. S. M., Bergmann, C., Black, A. K., Brown, A., Carbajal, M. J., Durrant, S., Fennell, C. T., Fiévet, A.-C., & Frank, M. C. (2021). A multilab study of bilingual infants: Exploring the preference for infant-directed speech. *Advances in Methods and Practices in Psychological Science*, *4*(1), 2515245920974622.

Calder, J. (2019). The fierceness of fronted/s: Linguistic rhematization through visual transformation. *Language in Society*, *48*(1), 31–64.

Carozza, S., & Leong, V. (2021). The role of affectionate caregiver touch in early neurodevelopment and parent–infant interactional synchrony. *Frontiers in Neuroscience*, *14*, 613378.

Cartmill, E. A., Armstrong III, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences*, *110*(28), 11278–11283.

Casillas, M., Bobb, S. C., & Clark, E. V. (2016). Turn-taking, timing, and planning in early language acquisition. *Journal of Child Language*, *43*(6), 1310–1337.

Casillas, M., Brown, P., & Levinson, S. C. (2020). Early language experience in a Tseltal Mayan village. *Child Development*, *91*(5), 1819–1835. https://doi.org/10.1111/cdev.13349

Chaiken, M. (1990). The ontogeny of antiphonal calling in European starlings. *Developmental Psychobiology*, *23*(3), 233–246. https://doi.org/10.1002/dev.420230304

Chalmers, A. F. (2013). *What is this thing called science?*. Hackett Publishing.

Chambers, C. D., & Tzavella, L. (2022). The past, present and future of registered reports. *Nature Human Behaviour*, *6*(1), 29–42.

Chen, Y., Matheson, L. E., & Sakata, J. T. (2016). Mechanisms underlying the social enhancement of vocal learning in songbirds. *Proceedings of the National Academy of Sciences*, *113*(24), 6641–6646.

Cho, S., Cola, M., Knox, A., Pelella, M. R., Russell, A., Hauptmann, A., Covello, M., Cieri, C., Liberman, M., Schultz, R. T., & Parish-Morris, J. (2023). Sex differences in the temporal dynamics of autistic children's natural conversations. *Molecular Autism*, *14*(1), 1-14.

Chong, S., Werker, J. F., Russell, J. A., & Carroll, J. M. (2003). Three facial expressions mothers direct to their infants. *Infant and Child Development: An International Journal of Research and Practice*, *12*(3), 211-232.

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*, E62. https://doi.org/10.1017/S0140525X1500031X

Christiansen, M. H., Kallens, P. C., & Trecca, F. (2022). Towards A Comparative Approach to Language Acquisition. *Current Directions in Psychological Science*, *31*, 131–138. https://doi.org/10.1177/09637214211049229

Claesen, A., Gomes, S., Tuerlinckx, F., & Vanpaemel, W. (2021). Comparing dream to reality: An assessment of adherence of the first generation of preregistered studies. *Royal Society Open Science*, *8*(10), 211037.

Clark, E. V., & Estigarribia, B. (2011). Using speech and gesture to introduce new objects to young children. *Gesture*, *11*(1), 1–23.

Clegg, J. M., Wen, N. J., DeBaylo, P. H., Alcott, A., Keltner, E. C., & Legare, C. H. (2021). Teaching through collaboration: Flexibility and diversity in caregiver–child interaction across cultures. *Child Development*, *92*(1), e56–e75.

Clyburne-Sherin, A., Fei, X., & Green, S. A. (2019). Computational reproducibility via containers in psychology. *Meta-Psychology*, *3*.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61*(5), 1584–1595. https://doi.org/10.1111/j.1467-8624.1990.tb02885.x

Corbeil, M., Trehub, S. E., & Peretz, I. (2013). Speech vs. Singing: Infants choose happier sounds. *Frontiers in Psychology*, *4*, 372.

Coretta, S., Casillas, J. V., Roessig, S., Franke, M., Ahn, B., Al-Hoorie, A. H., Al-Tamimi, J., Alotaibi, N. E., AlShakhori, M. K., & Altmiller, R. M. (2023). Multidimensional Signals and Analytic Flexibility: Estimating Degrees of Freedom in Human-Speech Analyses. *Advances in Methods and Practices in Psychological Science*, *6*(3). doi:10.1177/25152459231162567

Cox, C., 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.

Cox, C., 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, 1672–1696. https://doi.org/10.1111/cdev.13950

Cox, C., Templeton, E., & Fusaroli, R. (2023). Fine-Tuning Social Timing: From Non-Human to Human Animals and Back. A Commentary on The Evolution of Social Timing by Verga, Kotz and Ravignani (2023). *Physics of Life Reviews*, *47*, 79–81.

Crago, M. B., & Allen, S. E. (1997). *Linguistic and cultural aspects of simplicity and complexity in Inuktitut child directed speech*. 91–102.

Craver, C. F. (2009). Mechanisms and natural kinds. *Philosophical Psychology*, *22*(5), 575–594.

Craver, C. F. (2014). The ontic account of scientific explanation. *Explanation in the Special Sciences: The Case of Biology and History*, 27–52.

Creanza, N., Kolodny, O., & Feldman, M. W. (2017). Cultural evolutionary theory: How culture evolves and why it matters. *Proceedings of the National Academy of Sciences*, *114*(30), 7782–7789.

Cristia, A. (2010). Phonetic enhancement of sibilants in infant-directed speech. *The Journal of the Acoustical Society of America*, *128*(1), 424–434. https://doi.org/10.1121/1.3436529

Cristia, A. (2023). A systematic review suggests marked differences in the prevalence of infant-directed vocalization across groups of populations. *Developmental Science*, *26*(1), e13265.

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

Cristia, A., Seidl, A., Singh, L., & Houston, D. (2016). Test–retest reliability in infant speech perception tasks. *Infancy*, *21*(5), 648–667.

Cristia, A., Tsuji, S., & Bergmann, C. (2022). A meta-analytic approach to evaluating the explanatory adequacy of theories. *Meta-Psychology*, *6*.

Crocetti, E., & Salmela-Aro, K. (2018). The multifaceted nature of identity. *European Psychologist, 23*(4), 273-277. https://doi.org/10.1027/1016-9040/a000340

Cruz Blandón, M. A., Cristia, A., & Räsänen, O. (2023). Introducing Meta-analysis in the Evaluation of Computational Models of Infant Language Development. *Cognitive Science*, *47*(7), e13307.

Cummins, F. (2019). *The ground from which we speak: Joint speech and the collective subject*. Cambridge Scholars Publishing.

Cychosz, M., Cristia, A., Bergelson, E., Casillas, M., Baudet, G., Warlaumont, A. S., Scaff, C., Yankowitz, L., & Seidl, A. (2021). Vocal development in a large-scale crosslinguistic corpus. *Developmental Science*, e13090. https://doi.org/10.1111/desc.13090

Dailey, S., & Bergelson, E. (2022). Language input to infants of different socioeconomic statuses: A quantitative meta-analysis. *Developmental Science*, *25*(3), e13192.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, *208*(4448), 1174–1176.

Deffner, D., Rohrer, J. M., & McElreath, R. (2021). A Causal Framework for Cross-Cultural Generalizability. *Advances in Methods and Practices in Psychological Science*, *5*(3). https://doi.org/10.1177/25152459221106366

DePaolis, R. A., Vihman, M. M., & Keren-Portnoy, T. (2011). Do production patterns influence the processing of speech in prelinguistic infants? *Infant Behavior and Development*, *34*(4), 590–601.

DePaolis, R. A., Vihman, M. M., & Nakai, S. (2013). The influence of babbling patterns on the processing of speech. *Infant Behavior and Development*, *36*(4), 642–649.

Deruiter, S. L., & Solow, A. R. (2008). A rotation test for behavioural point-process data. *Animal Behaviour*, *76*(4), 1429–1434. https://doi.org/10.1016/j.anbehav.2008.06.016

Devezer, B., & Buzbas, E. O. (2023). *Rigorous exploration in a model-centric science via epistemic iteration*. MetaArXiv: https://osf.io/qe46u/

De Vos, C., Torreira, F., & Levinson, S. C. (2016). Turn-timing in signed conversations: coordinating stroke-to-stroke turn boundaries. *Frontiers in Psychology, 6(268), 183-196.*

Devezer, B., Nardin, L. G., Baumgaertner, B., & Buzbas, E. O. (2019). Scientific discovery in a model-centric framework: Reproducibility, innovation, and epistemic diversity. *PloS One*, *14*(5), e0216125.

Dideriksen, C., Christiansen, M. H., Tylén, K., Dingemanse, M., & Fusaroli, R. (2023). Quantifying the interplay of conversational devices in building mutual understanding. *Journal of Experimental Psychology: General*, *152*(3), 864-889.

Dideriksen, C., & Fusaroli, R. (2018). *CDS is not what you think-Hypoarticulation in Danish Child Directed Speech*.

Dilley, L., 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.

Dimitrova, N., & Moro, C. (2013). Common ground on object use associates with caregivers' gesturese. *Infant Behavior and Development*, *36*(4), 618–626.

Dingemanse, M., & Liesenfeld, A. (2022). *From text to talk: Harnessing conversational corpora for humane and diversity-aware language technology*. 5614–5633.

Dingemanse, M., & Enfield, N. (2023). Interactive repair and the foundations of language. *Trends in Cognitive Sciences, 1-25*. https://doi.org/10.1016/j.tics.2023.09.003

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)*.

Donnelly, S., & Kidd, E. (2021). The longitudinal relationship between conversational turn-taking and vocabulary growth in early language development. *Child Development*, *92*(2), 609–625.

Dumas, G. F., Merle T. (2021). Reciprocity and alignment: Quantifying coupling in dynamic interactions. *Royal Society Open Science*, *8*(5), 210138–210138. https://doi.org/0.1098/rsos.210138

Dunn, J., & Shatz, M. (1989). Becoming a conversationalist despite (or because of) having an older sibling. *Child Development*, 399–410.

Dykiert, D., Der, G., Starr, J. M., & Deary, I. J. (2012). *Age differences in intra-individual variability in simple and choice reaction time: Systematic review and meta-analysis*. PLoS ONE 7(10): e45759. https://doi.org/10.1371/journal.pone.0045759

Eaves, B. S., 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

Eckert, P., & Labov, W. (2017). Phonetics, phonology and social meaning. *Journal of Sociolinguistics*, *21*(4), 467–496.

Eliasoph, N., & Lichterman, P. (2003). Culture in interaction. *American Journal of Sociology*, *108*(4), 735–794.

Elmlinger, S. L., Schwade, J. A., & Goldstein, M. H. (2019). The ecology of prelinguistic vocal learning: Parents simplify the structure of their speech in response to babbling. *Journal of Child Language*, *46*(5), 998–1011.

Elmlinger, S. L., Schwade, J. A., Vollmer, L., & Goldstein, M. H. (2023). Learning how to learn from social feedback: The origins of early vocal development. *Developmental Science*, *26*(2), e13296.

Emmorey, K. (2014). Iconicity as structure mapping. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130301.

Englund, K. T. (2005). Voice onset time in infant directed speech over the first six months. *First Language*, *25*(2), 219–234.

Englund, K. T. (2018). Hypoarticulation in infant-directed speech. *Applied Psycholinguistics*, *39*(1), 67–87.

Englund, K. T., & Behne, D. M. (2005). Infant directed speech in natural interaction—Norwegian vowel quantity and quality. *Journal of Psycholinguistic Research*, *34*(3), 259–280. https://doi.org/10.1007/s10936-005-3640-7

Ervin-Tripp, S. (1982). Ask and it shall be given unto you: Children’s requests. *Georgetown University Roundtable on Languages and Linguistics. Contemporary Perceptions of Language: Interdisciplinary Dimensions*, 235–245.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*(5), 429–448.

Feldman, M. W., Cavalli-Sforza, L. L., & Peck, J. R. (1985). Gene-culture coevolution: Models for the evolution of altruism with cultural transmission. *Proceedings of the National Academy of Sciences*, *82*(17), 5814–5818.

Feldman, R. (2012). Bio-behavioral synchrony: A model for integrating biological and microsocial behavioral processes in the study of parenting. *Parenting*, *12*(2–3), 154–164.

Ferguson, C. A. (1964). Baby talk in six languages. *American Anthropologist*, *66*(6_PART2), 103–114.

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.

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.

Ferjan Ramírez, N. (2022). Fathers' infant-directed speech and its effects on child language development. *Language and Linguistics Compass*, *16*(1), e12448.

Ferjan Ramírez, N., Lytle, S. R., & Kuhl, P. K. (2020). Parent coaching increases conversational turns and advances infant language development. *Proceedings of the National Academy of Sciences*, *117*(7), 3484–3491.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, 1497–1510. https://doi.org/10.2307/1130938

Fernald, A. (1992). Meaningful melodies in mothers' speech to infants. In *Nonverbal vocal communication: Comparative and developmental approaches* (pp. 262–282). Cambridge University Press.

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, *10*(3), 279–293.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology*, *27*(2), 209-221.

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

Fernandez, A. A., & Knörnschild, M. (2020). Pup directed vocalizations of adult females and males in a vocal learning bat. *Frontiers in Ecology and Evolution*, *8*, 265.

Ferrari, P. F. P., Annika; Ionica, Consuel; Suomi, Stephen J. (2009). Reciprocal Face-to-Face Communication between Rhesus Macaque Mothers and Their Newborn Infants. *Current Biology : CB*, *19*(20), 1768–1772. https://doi.org/10.1016/j.cub.2009.08.055

Floccia, C., Keren-Portnoy, T., DePaolis, R., Duffy, H., Delle Luche, C., Durrant, S., White, L., Goslin, J., & Vihman, M. (2016). British English infants segment words only with exaggerated infant-directed speech stimuli. *Cognition*, *148*, 1–9. https://doi.org/10.1016/j.cognition.2015.12.004

Ford, C. E., Fox, B. A., & Thompson, S. A. (1996). Practices in the construction of turns: The "TCU" revisited. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)*, *6*(3), 427–454.

Francis, G. (2012). The psychology of replication and replication in psychology. *Perspectives on Psychological Science*, *7*(6), 585–594.

Fusaroli, R., Cox, C., Cho, S., Pelella, M. R., Liberman, M., Cieri, C., Hauptmann, A., Russell, A., Hulick, A. T., & Tena, K. (2023). Autistic Children Are Faster at Turn-Taking, More Strongly Driven By Their Own Tempo, and Less Coordinated with Their Interlocutor Than Neurotypical Controls. *INSAR 2023*.

Fusaroli, R., Grossman, R., Bilenberg, N., Cantio, C., Jepsen, J. R. M., & Weed, E. (2022). Towards a cumulative science of vocal markers of autism: A cross-linguistic meta-analysis-based investigation of acoustic markers in American and Danish autistic children. *Autism Res.15*(4), 653-664. doi: 10.1002/aur.2661

Fusaroli, R., Grossman, R., Bilenberg, N., Cantio, C., Jepsen, J. R. M., & Weed, E. (2022). Toward a cumulative science of vocal markers of autism: A cross-linguistic meta-analysis-based investigation of acoustic markers in American and Danish autistic children. *Autism Research*, *15*(4), 653–664.

Fusaroli, R., Rączaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, *32*, 147–157.

Fusaroli, R., Gangopadhyay, N., & Tylén, K. (2014). The dialogically extended mind: Language as skilful intersubjective engagement. *Cognitive Systems Research*, *29*, 31–39.

Fusaroli, R., & Tylén, K. (2012). Carving language for social coordination: A dynamical approach. *Interaction Studies*, *13*(1), 103–124.

Fusaroli, R., & Tylén, K. (2016). Investigating conversational dynamics: Interactive alignment, Interpersonal synergy, and collective task performance. *Cognitive Science*, *40*(1), 145–171.

Fusaroli, R., Weed, E., Fein, D., & Naigles, L. (2019). Hearing me hearing you: Reciprocal effects between child and parent language in autism and typical development. *Cognition*, *183*, 1–18. https://doi.org/10.1016/j.cognition.2018.10.022

García-Sierra, A., Ramírez-Esparza, N., Wig, N., & Robertson, D. (2021). Language learning as a function of infant directed speech (IDS) in Spanish: Testing neural commitment using the positive-MMR. *Brain and Language*, *212*, 104890.

Garvey, C., & Berninger, G. (1981). Timing and turn taking in children's conversations. *Discourse Processes*, *4*(1), 27–57.

Gaskins, S. (2006). 13 The Cultural Organization of Yucatec Mayan Children's Social Interactions. *Peer Relationships in Cultural Context*, 283-309.

Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. *Department of Statistics, Columbia University*, *348*, 1–17.

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), 1–10. https://doi.org/10.1038/s41598-017-13883-2

Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*, *33*(5), 587–606.

Gilkerson, J., Richards, J. A., Warren, S. F., Oller, D. K., Russo, R., & Vohr, B. (2018). Language experience in the second year of life and language outcomes in late childhood. *Pediatrics*, *142*(4), e20174276. https://doi.org/10.1542/peds.2017-4276

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development*, *71*(4), 878–894.

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19*(5), 515–523. https://doi.org/10.1111/j.1467-9280.2008.02117.x

Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby) talk to me: The social context of infant-directed speech and its effects on early language acquisition. *Current Directions in Psychological Science*, *24*(5), 339–344. https://doi.org/10.1177/0963721415595345

Gómez, E., & Strasser, K. (2021). Language and socioemotional development in early childhood: The role of conversational turns. *Developmental Science*, *24*(5), e13109.

Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, *32*(10),

1489–1522.

Goodwin, C. (2007). Participation, stance and affect in the organization of activities. *Discourse & Society*, *18*(1), 53–73.

Graf Estes, K., & Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meanings. *Infancy*, *18*(5), 797–824.

Gratier, M., Devouche, E. (2017). The Development of Infant Participation in Communication. In Filippa, M., Kuhn, P., Westrup, B. (eds) *Early Vocal Contact and Preterm Infant Brain Development* (55–69). Springer, Cham. https://doi.org/10.1007/978-3-319-65077-7_4

Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., & Parlato-Oliveira, E. (2015). Early development of turn-taking in vocal interaction between mothers and infants. *Frontiers in Psychology*, *6*(1167), 236–245.

Greenfield, M. D. (1994). Synchronous and Alternating Choruses in Insects and Anurans: Common Mechanisms and Diverse Functions. *American Zoologist*, *34*(6), 605–615. https://doi.org/10.1093/icb/34.6.605

Griffin, D., & Gonzalez, R. (2003). Models of dyadic social interaction. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1431), 573–581.

Grønnum, N. (1998). Danish. *Journal of the International Phonetic Association*, *28*(1–2), 99–105.

Gros-Louis, J. W., Meredith J. ;. Goldstein, Michael H. ;. King, Andrew P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development*, *30*(6), 509–516. https://doi.org/10.1177/0165025406071914

Gros-Louis, J. W., Meredith J. ;. King, Andrew P. (2014). Maternal Responsiveness and the Development of Directed Vocalizing in Social Interactions. *Infancy*, *19*(4), 385–408. https://doi.org/10.1111/infa.12054

Gros-Louis, J., West, M. J., & King, A. P. (2016). The influence of interactive context on prelinguistic vocalizations and maternal responses. *Language Learning and Development*, *12*(3), 280–294.

Guest, O., & Martin, A. E. (2021). How computational modeling can force theory building in psychological science. *Perspectives on Psychological Science*, *16*(4), 789–802.

Guest, O., & Martin, A. E. (2023). On logical inference over brains, behaviour, and artificial neural networks. *Computational Brain & Behavior*, 1–15.

Háden, G. P., Mády, K., Török, M., & Winkler, I. (2020). Newborn infants differently process adult directed and infant directed speech. *International Journal of Psychophysiology*, *147*, 107–112.

Han, M., De Jong, N., & Kager, R. (2020). Pitch properties of infant-directed speech specific to word-learning contexts: A cross-linguistic investigation of Mandarin Chinese and Dutch. *Journal of Child Language*, *47*(1), 85–111. https://doi.org/10.1017/S0305000919000813

Hardwicke, T. E., Mathur, M. B., MacDonald, K., Nilsonne, G., Banks, G. C., Kidwell, M. C., Hofelich Mohr, A., Clayton, E., Yoon, E. J., & Henry Tessler, M. (2018). Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal Cognition. *Royal Society Open Science*, *5*(8), 180448.

Harkness, S., & Super, C. M. (2002). Culture and parenting. *Handbook of Parenting*, *2*(2), 253–280.

Harrist, A. W., & Waugh, R. M. (2002). Dyadic synchrony: Its structure and function in children’s development. *Developmental Review*, *22*(4), 555–592.

Hart, B., & Risley, T. R. (1992). American parenting of language-learning children: Persisting differences in family-child interactions observed in natural home environments. *Developmental Psychology*, *28*(6), 1096-1105.

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

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*(7302), 29–29.

Henrich, J. (2015). Culture and social behavior. *Current Opinion in Behavioral Sciences*, *3*, 84–89.

Henrich, J., & McElreath, R. (2003). The evolution of cultural evolution. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews*, *12*(3), 123–135.

Heyes, C. (2016). Born pupils? Natural pedagogy and cultural pedagogy. *Perspectives on Psychological Science*, *11*(2), 280–295.

Heyes, C. (2018). *Cognitive gadgets: The cultural evolution of thinking*. Harvard University Press.

Hilbrink, E. G., Merideth; Levinson, Stephen C. (2015). Early developmental changes in the timing of turn-taking: A longitudinal study of mother–infant interaction. *Frontiers in Psychology*, *6*, 1492–1492. https://doi.org/10.3389/fpsyg.2015.01492

Hilton, C. B., Moser, C. J., Bertolo, M., Lee-Rubin, H., Amir, D., Bainbridge, C. M., Simson, J., Knox, D., Glowacki, L., Galbarczyk, A., Jasienska, G., Ross, C. T., Neff, M. B., Martin, A., Cirelli, L. K., Trehub, S. E., Song, J., Kim, M., Schachner, A., … Mehr, S. A. (2022). Acoustic regularities in infant-directed speech and song across cultures. *Nature Human Behaviour*, 1–12. https://doi.org/10.1038/s41562-022-01410-x

Hirai, M., Kanakogi, Y., & Ikeda, A. (2022). Observing inefficient action can induce infant preference and learning. *Developmental Science*, *25*(1), e13152.

Hoff-Ginsberg, E. (1991). Mother-child conversation in different social classes and communicative settings. *Child Development*, *62*(4), 782–796

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development*, *73*(2), 418–433.

Horton, L., & Singleton, J. (2022). Acquisition of turn-taking in sign language conversations: An overview of language modality and turn structure. *Frontiers in Psychology*, 4413.

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130298.

Ingram, D. (1995). The cultural basis of prosodic modifications to infants and children: A response to Fernald's universalist theory. *Journal of Child Language*, *22*(1), 223–233.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124.

Ishkhanyan, B., Fusaroli, R., Højen, A., Tylén, K., Johansson, C., & Christiansen, M. H. (2020). *Cross-linguistic differences in categorical perception: Comparison of Danish and Norwegian*. PsyArxiv: https://osf.io/preprints/psyarxiv/jpbtw

Iverson, J. M., Capirci, O., Longobardi, E., & Caselli, M. C. (1999). Gesturing in mother-child interactions. *Cognitive Development*, *14*(1), 57–75.

Jager, K., Zoccali, C., Macleod, A., & Dekker, F. (2008). Confounding: What it is and how to deal with it. *Kidney International*, *73*(3), 256–260.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524–532.

Johnston, J. R., & Wong, M.-Y. A. (2002). Cultural differences in beliefs and practices concerning talk to children. *Journal of Speech, Language and Hearing, 45*(5), 916-926

Johnson, K. (2004). Acoustic and auditory phonetics. *Phonetica*, *61*(1), 56–58.

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*, 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.

Kaplan, P. S., Bachorowski, J., Smoski, M. J., & Zinser, M. (2001). Role of clinical diagnosis and medication use in effects of maternal depression on infant-directed speech. *Infancy*, *2*(4), 537–548. https://doi.org/10.1207/S15327078IN0204_08

Karasik, L. B., Tamis-LeMonda, C. S., & Adolph, K. E. (2011). Transition from crawling to walking and infants' actions with objects and people. *Child Development*, *82*(4), 1199–1209.

Karasik, L. B., Tamis-LeMonda, C. S., & Adolph, K. E. (2014). Crawling and walking infants elicit different verbal responses from mothers. *Developmental Science*, *17*(3), 388–395.

Karmazyn-Raz, H., & Smith, L. B. (2023). Discourse with few words: Coherence statistics, parent-infant actions on objects, and object names. *Language Acquisition*, *30*(3–4), 211–229.

Kärtner, J., Crafa, D., Chaudhary, N., & Keller, H. (2016). Reactions to receiving a gift—Maternal scaffolding and cultural learning in Berlin and Delhi. *Child Development*, *87*(3), 712–722.

Keller, H., Borke, J., Staufenbiel, T., Yovsi, R. D., Abels, M., Papaligoura, Z., Jensen, H., Lohaus, A., Chaudhary, N., & Lo, W. (2009). Distal and proximal parenting as alternative parenting strategies during infants' early months of life: A cross-cultural study. *International Journal of Behavioral Development*, *33*(5), 412–420.

Keller, H. S., Axel; Eibl-Eibesfeldt, Irenäus. (1988). Communication Patterns in Adult-Infant Interactions in Western and Non-Western Cultures. *Journal of Cross-Cultural Psychology*, *19*(4), 427–445. https://doi.org/10.1177/0022022188194003

Kennedy, D. P., & Adolphs, R. (2012). The social brain in psychiatric and neurological disorders. *Trends in Cognitive Sciences*, *16*(11), 559–572.

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and social psychology review*, *2*(3), 196-217.

Kershenbaum, A., Bowles, A. E., Freeberg, T. M., Jin, D. Z., Lameira, A. R., & Bohn, K. (2014). Animal vocal sequences: not the Markov chains we thought they were. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1792), 20141370.

Kidd, E., Christiansen, M., Majid, A., Bornkessel-Schlesewsky, I., Schlesewsky, M., & Evans, N. (2023). Harnessing Linguistic Diversity for Theories of Language and Mind. *Proceedings of the Annual Meeting of the Cognitive Science Society,* 45(45).

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

Kidd, E., & Garcia, R. (2022b). How diverse is child language acquisition research? *First Language*, *42*(6), 703–735.

Kidd, E., & Garcia, R. (2022c). Where to from here? Increasing language coverage while building a more diverse discipline. *First Language*, *42*(6), 837–851.

Kitamura, C., & Burnham, D. (1998). *Acoustic and affective qualities of IDS in English*. Fifth international Conference on spoken language processing. http://www.isca-speech.org/archive.

Kitamura, C., & Notley, A. (2009). The shift in infant preferences for vowel duration and pitch contour between 6 and 10 months of age. *Developmental Science*, *12*(5), 706–714. https://doi.org/10.1111/j.1467-7687.2009.00818.x

Ko, E.-S., Seidl, A., Cristia, A., Reimchen, M., & Soderstrom, M. (2016). Entrainment of prosody in the interaction of mothers with their young children. *Journal of Child Language*, *43*(2), 284–309. https://doi.org/10.1017/S0305000915000410

Koile, E., & Cristia, A. (2021). Toward cumulative cognitive science: A comparison of meta-analysis, mega-analysis, and hybrid approaches. *Open Mind*, *5*, 154–173.

Kondaurova, M. V., & Bergeson, T. R. (2011). The effects of age and infant hearing status on maternal use of prosodic cues for clause boundaries in speech. *Journal of Speech, Language, and Hearing Research*, *54*(3), 740–754. https://doi.org/10.1044/1092-4388(2010/09-0225)

Kondaurova, M. V., Bergeson, T. R., & Xu, H. (2013). Age-Related Changes in Prosodic Features of Maternal Speech to Prelingually Deaf Infants with Cochlear Implants. *Infancy*, *18*(5), 825–848. https://doi.org/10.1111/infa.12010

Kondaurova, M. V., Smith, N. A., Zheng, Q., Reed, J., & Fagan, M. K. (2020). Vocal turn-taking between mothers and their children with cochlear implants. *Ear and Hearing*, *41*(2), 362–373.

Konner, M. (2016). Hunter-gatherer infancy and childhood in the context of human evolution. *Childhood: Origins, Evolution, and Implications*, 123–154.

Konvalinka, I., Xygalatas, D., Bulbulia, J., Schjødt, U., Jegindø, E.-M., Wallot, S., Van Orden, G., & Roepstorff, A. (2011). Synchronized arousal between performers and related spectators in a fire-walking ritual. *Proceedings of the National Academy of Sciences*, *108*(20), 8514–8519.

Kosie, J., & Lew-Williams, C. (2023). *Infant-directed Communication*. PsyArxiv: https://osf.io/u5avz/

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

Kvarven, A., Strømland, E., & Johannesson, M. (2020). Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nature Human Behaviour*, *4*(4), 423–434.

Labov, W. (1965). On the mechanism of linguistic change. *Georgetown Monographs on Language and Linguistics*, *18*, 91–114.

Ladefoged, P. (1996). *Elements of acoustic phonetics*. University of Chicago Press.

Laing, C. (2019). A role for onomatopoeia in early language: Evidence from phonological development. *Language and Cognition*, *11*(2), 173–187.

Laing, C., & Bergelson, E. (2020). From babble to words: Infants' early productions match words and objects in their environment. *Cognitive Psychology*, *122*, 101308.

Lakens, D., Hilgard, J., & Staaks, J. (2016). On the reproducibility of meta-analyses: Six practical recommendations. *BMC Psychology*, *4*, 1–10.

Lakens, D., Page-Gould, E., Assen, M. A. L. M. V., Spellman, B., Schönbrodt, F. D., Hasselman, F., Corker, K. S., Grange, J. A., Sharples, A., Cavender, C., Augusteijn, H. E. M., Gerger, H., Locher, C., Miller, I. D., Anvari, F., & Scheel, A. M. (2017). Examining the reproducibility of meta-analyses in psychology: A preliminary report. MetaArXiv Preprints. https://doi.org/10.31222/osf.io/xfbjf

Lam-Cassettari, C., & Kohlhoff, J. (2020). Effect of maternal depression on infant-directed speech to prelinguistic infants: Implications for language development. *PLoS One*, *15*(7), e0236787.

Lammert, A. C., Shadle, C. H., Narayanan, S. S., & Quatieri, T. F. (2018). Speed-accuracy tradeoffs in human speech production. *PloS One*, *13*(9), e0202180.

Lam, J., & Tjaden, K. (2013). Intelligibility of clear speech: Effect of instruction. *Journal of Speech, Language, and Hearing Research*, *56*(5), 1429–1440.

Lam, J., Tjaden, K., & Wilding, G. (2012). Acoustics of clear speech: Effect of instruction. *Journal of Speech, Language, and Hearing Research*, *55*(6), 1807–1821.

Lancy, D. F., Bock, J., & Gaskins, S. (2010). *The anthropology of learning in childhood*. Rowman Altamira.

Landy, J. F., Jia, M. L., Ding, I. L., Viganola, D., Tierney, W., Dreber, A., Johannesson, M., Pfeiffer, T., Ebersole, C. R., & Gronau, Q. F. (2020). Crowdsourcing hypothesis tests: Making transparent how design choices shape research results. *Psychological Bulletin*, *146*(5), 451-479.

Ledgerwood, A. (2014). Introduction to the special section on moving toward a cumulative science: Maximizing what our research can tell us. *Perspectives on Psychological Science*, *9*(6), 610-611.

Lee, C. S., Kitamura, C., Burnham, D., & McAngus Todd, N. P. (2014). On the rhythm of infant-versus adult-directed speech in Australian English. *The Journal of the Acoustical Society of America*, *136*(1), 357–365. https://doi.org/10.1121/1.4883479

Lee, G. Y., & Kisilevsky, B. S. (2014). Fetuses respond to father's voice but prefer mother's voice after birth. *Developmental Psychobiology*, *56*(1), 1–11.

Lee, S. A. S., & Davis, B. L. (2010). Segmental distribution patterns of English infant-and adult-directed speech. *Journal of Child Language*, *37*(4), 767–791.

Leezenbaum, N. B., Campbell, S. B., Butler, D., & Iverson, J. M. (2014). Maternal verbal responses to communication of infants at low and heightened risk of autism. *Autism*, *18*(6), 694–703.

Legare, C. H., & Harris, P. L. (2016). The ontogeny of cultural learning. *Child Development*, *87*(3), 633–642.

Legare, C. H., & Nielsen, M. (2015). Imitation and innovation: The dual engines of cultural learning. *Trends in Cognitive Sciences*, *19*(11), 688–699.

Leung, K. (2011). Presenting post hoc hypotheses as a priori: Ethical and theoretical issues. *Management and Organization Review*, *7*(3), 471-479.

LeVine, R. A. (2007). Ethnographic studies of childhood: A historical overview. *American Anthropologist*, *109*(2), 247–260.

Levinson, S. (2006). Cognition at the heart of human interaction. *Discourse Studies*, *8*(1), 85–93.

Levinson, S. C. (2006). On the Human "Interaction Engine". In N. J. Enfield, & S. C. Levinson (Eds.), Roots of Human Sociality: Culture, Cognition and Interaction (pp. 39-69). Oxford: Berg.

Levinson, S. C. (2016). Turn-taking in human communication–origins and implications for language processing. *Trends in Cognitive Sciences*, *20*(1), 6–14.

Levinson, S. C. ; H., Judith. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1651), 20130302–20130302. https://doi.org/10.1098/rstb.2013.0302

Lew-Williams, C., Ferguson, B., Abu-Zhaya, R., & Seidl, A. (2019). Social touch interacts with infants' learning of auditory patterns. *Developmental Cognitive Neuroscience*, *35*, 66–74.

Lindsay, L., Gambi, C., & Rabagliati, H. (2019). Preschoolers optimize the timing of their conversational turns through flexible coordination of language comprehension and production. *Psychological Science*, *30*(4),

504–515.

Lingle, S., & Riede, T. (2014). Deer mothers are sensitive to infant distress vocalizations of diverse mammalian species. *The American Naturalist*, *184*(4), 510–522.

Lingle, S., Wyman, M. T., Kotrba, R., Teichroeb, L. J., & Romanow, C. A. (2012). What makes a cry a cry? A review of infant distress vocalizations. *Current Zoology*, *58*(5), 698–726.

Little, E. E., Carver, L. J., & Legare, C. H. (2016). Cultural variation in triadic infant–caregiver object exploration. *Child Development*, *87*(4), 1130–1145.

Liu, H., Kuhl, P. K., & Tsao, F. (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. (2005). The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *The Journal of the Acoustical Society of America*, *117*(6), 3879–3889.

Loakes, D., Moses, K., Wigglesworth, G., Simpson, J., & Billington, R. (2013). Children's language input: A study of a remote multilingual Indigenous Australian community. *Multilingua*, *32*(5), 683–711.

Lopez, L. D., Walle, E. A., Pretzer, G. M., & Warlaumont, A. S. (2020). Adult responses to infant prelinguistic vocalizations are associated with infant vocabulary: A home observation study. *PloS One*, *15*(11), e0242232.

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

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.

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

Lovcevic, I., Burnham, D., & Kalashnikova, M. (2022). Infants' Lexical Processing: Independent Contributions of Attentional and Clarity Cues. *Language Learning and Development*, 1–18.

Ludusan, B., Cristia, A., Martin, A., Mazuka, R., & Dupoux, E. (2016). Learnability of prosodic boundaries: Is infant-directed speech easier? *The Journal of the Acoustical Society of America*, *140*(2), 1239–1250.

Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word learning in infant-and adult-directed speech. *Language Learning and Development*, *7*(3), 185–201.

Machery, E. (2020). What is a replication? *Philosophy of Science*, *87*(4), 545–567.

Magyari, L., Bastiaansen, M. C., De Ruiter, J. P., & Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. *Journal of Cognitive Neuroscience*, *26*(11), 2530–2539.

Magyari, L., & De Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Frontiers in Psychology*, *3*, 376.

Magyari, L., De Ruiter, J. P., & Levinson, S. C. (2017). Temporal preparation for speaking in question-answer sequences. *Frontiers in Psychology*, *8*, 211.

Majorano, M., Vihman, M. M., & DePaolis, R. A. (2014). The relationship between infants' production experience and their processing of speech. *Language Learning and Development*, *10*(2), 179–204.

ManyBabies Consortium. (2020). Quantifying sources of variability in infancy research using the infant-directed-speech preference. *Advances in Methods and Practices in Psychological Science*, *3*(1), 24–52. https://doi.org/10.1177/2515245919900809

Mareschal, D. (2007). *Neuroconstructivism: Perspectives and prospects*. Oxford University Press

Marklund, E., & Gustavsson, L. (2020). The dynamics of vowel hypo-and hyperarticulation in Swedish infant-directed speech to 12-month-olds. *Frontiers in Communication*, *79*.

Marklund, E., Marklund, U., & Gustavsson, L. (2021). An Association Between Phonetic Complexity of Infant Vocalizations and Parent Vowel Hyperarticulation. *Frontiers in Psychology*, 2873.

Marks, J. (2009). *Why I am not a scientist: Anthropology and modern knowledge*. Univ of California Press.

Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., & Cristia, A. (2015). Mothers speak less clearly to infants than to adults: A comprehensive test of the hyperarticulation hypothesis. *Psychological Science*, *26*(3), 341–347. https://doi.org/10.1177/0956797614562453

Masataka, N. (1992). Motherese in a signed language. *Infant Behavior and Development*, *15*(4), 453–460.

Masataka, N. (1995). The relation between index-finger extension and the acoustic quality of cooing in three-month-old infants. *Journal of Child Language*, *22*(2), 247–257. https://doi.org/10.1017/s0305000900009776

Masataka, N., Koda, H., Urasopon, N., & Watanabe, K. (2009). Free-ranging macaque mothers exaggerate tool-using behavior when observed by offspring. *PLoS One*, *4*(3), e4768.

Masek, L. R., Ramirez, A. G., McMillan, B. T., Hirsh-Pasek, K., & Golinkoff, R. M. (2021). Beyond counting words: A paradigm shift for the study of language acquisition. *Child Development Perspectives*, *15*(4),

274–280.

Masek, L. R., Weiss, S. M., McMillan, B. T., Paterson, S. J., Golinkoff, R. M., & Hirsh-Pasek, K. (2023). Contingent conversations build more than language: How communicative interactions in toddlerhood relate to preschool executive function skills. *Developmental Science*, *26*(3), e13338.

Massaro, D. W., & Perlman, M. (2017). Quantifying iconicity's contribution during language acquisition: Implications for vocabulary learning. *Frontiers in Communication*, *2*, 4. doi: 10.3389/fcomm.2017.00004

Mathur, M. B., & VanderWeele, T. J. (2019). New metrics for meta-analyses of heterogeneous effects. *Statistics in Medicine*, *38*(8), 1336–1342.

Mathur, M. B., & VanderWeele, T. J. (2020). Sensitivity analysis for publication bias in meta-analyses. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *69*(5), 1091–1119.

Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods*, *12*(1), 23.

McElreath, R. (2018). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC.

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

McRoberts, G. W., McDonough, C., & Lakusta, L. (2009). The role of verbal repetition in the development of infant speech preferences from 4 to 14 months of age. *Infancy*, *14*(2), 162–194. https://doi.org/10.1080/15250000802707062

Medwin, H. (1975). Speed of sound in water: A simple equation for realistic parameters. *The Journal of the Acoustical Society of America*, *58*(6), 1318–1319. https://doi.org/10.1121/1.380790

Meyer, M., Hunnius, S., Van Elk, M., Van Ede, F., & Bekkering, H. (2011). Joint action modulates motor system involvement during action observation in 3-year-olds. *Experimental Brain Research*, *211*, 581–592.

Mesoudi, A., & Whiten, A. (2004). The hierarchical transformation of event knowledge in human cultural transmission. *Journal of Cognition and Culture*, *4*(1), 1–24.

Miller, J. L., Hurdish, E., & Gros-Louis, J. (2018). Different patterns of sensitivity differentially affect infant attention span. *Infant Behavior and Development*, *53*, 1–4.

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.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group*. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, *151*(4), 264–269.

Morey, R. D., Chambers, C. D., Etchells, P. J., Harris, C. R., Hoekstra, R., Lakens, D., Lewandowsky, S., Morey, C. C., Newman, D. P., & Schönbrodt, F. D. (2016). The Peer Reviewers' Openness Initiative: Incentivizing open research practices through peer review. *Royal Society Open Science*, *3*(1), 150547.

Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *The American Naturalist*, *111*(981), 855–869.

Moses, L. J., Baldwin, D. A., Rosicky, J. G., & Tidball, G. (2001). Evidence for referential understanding in the emotions domain at twelve and eighteen months. *Child Development*, *72*(3), 718–735.

Nakahara, F., & Miyazaki, N. (2011). Vocal exchanges of signature whistles in bottlenose dolphins (Tursiops truncatus). *Journal of Ethology*, *29*(2), 309–320. https://doi.org/10.1007/s10164-010-0259-4

Naoi, N., Minagawa-Kawai, Y., Kobayashi, A., Takeuchi, K., Nakamura, K., Yamamoto, J., & Shozo, K. (2012). Cerebral responses to infant-directed speech and the effect of talker familiarity. *Neuroimage*, *59*(2), 1735–1744.

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

Needham, A., Barrett, T., & Peterman, K. (2002). A pick-me-up for infants' exploratory skills: Early simulated experiences reaching for objects using 'sticky mittens' enhances young infants' object exploration skills. *Infant Behavior and Development*, *25*(3), 279–295.

Nencheva, M. L., & Lew-Williams, C. (2022). Understanding why infant-directed speech supports learning: A dynamic attention perspective. *Developmental Review*, *66*, 101047.

Neimy, H., Pelaez, M., Carrow, J., Monlux, K., & Tarbox, J. (2017). Infants at risk of autism and developmental disorders: Establishing early social skills. *Behavioral Development Bulletin*, *22*(1), 6.

Nencheva, M. L., Piazza, E. A., & Lew-Williams, C. (2021). The moment-to-moment pitch dynamics of child-directed speech shape toddlers' attention and learning. *Developmental Science*, *24*(1), e12997.

Nencheva, M. L., & Lew-Williams, C. (2022). Understanding why infant-directed speech supports learning: A dynamic attention perspective. *Developmental Review*, *66*, 101047.

Nguyen, V., Versyp, O., Cox, C., & Fusaroli, R. (2022). A systematic review and Bayesian meta-analysis of the development of turn taking in adult–child vocal interactions. *Child Development*, *93*(4), 1181–1200.

Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology*, *162*, 31–38. https://doi.org/10.1016/j.jecp.2017.04.017

Nikolaus, M., & Fourtassi, A. (2023). Communicative feedback in language acquisition. *New Ideas in Psychology*, *68*, 100985.

Niwano, K., & Sugai, K. (2002). Intonation contour of Japanese maternal infant-directed speech and infant vocal response. *The Japanese Journal of Special Education*, *39*(6), 59–68.

Nosek, B. A., & Bar-Anan, Y. (2012). Scientific utopia: I. Opening scientific communication. *Psychological Inquiry*, *23*(3), 217–243.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600–2606.

Novack, M. A., Brentari, D., Goldin-Meadow, S., & Waxman, S. (2021). Sign language, like spoken language, promotes object categorization in young hearing infants. *Cognition*, *215*, 104845.

Nuijten, M. B., Van Assen, M. A., Augusteijn, H. E., Crompvoets, E. A., & Wicherts, J. M. (2020). Effect sizes, power, and biases in intelligence research: A meta-meta-analysis. *Journal of Intelligence*, *8*(4), 36.

Oakes, L. M., & Rakison, D. H. (2019). *Developmental cascades: Building the infant mind*. Oxford University Press.

Ochs, E., & Schieffelin, B. (1984). Language acquisition and socialization. *Culture Theory: Essays on Mind, Self, and Emotion*, 276–320.

O'Dell, M. L. ; N., Tommi; Lennes, Mietta. (2012). Modeling Turn-Taking Rhythms with Oscillators. *Linguistica Uralica*, *48*(3), 218-NA. https://doi.org/10.3176/lu.2012.3.08

Ohala, J. J. (1980). The acoustic origin of the smile. *The Journal of the Acoustical Society of America*, *68*(S1), S33–S33.

Ohala, J. J. (2008). The ethological basis of certain signals of affect and emotion. *UC Berkeley PhonLab Annual Report*, *4*(4).

Oller, D. K., Caskey, M., Yoo, H., Bene, E. R., Jhang, Y., Lee, C. C., Bowman, D. D., Long, H. L., Buder, E., H.

& Vohr, B. (2019). Preterm and full term infant vocalization and the origin of language. *Scientific reports*, *9*(1), 14734.

Oller, D. K., Griebel, U., Iyer, S. N., Jhang, Y., Warlaumont, A. S., Dale, R., & Call, J. (2019). Language origins viewed in spontaneous and interactive vocal rates of human and bonobo infants. *Frontiers in psychology*, *10*, 729.

Olson, J., & Masur, E. F. (2011). Infants' gestures influence mothers' provision of object, action and internal state labels. *Journal of Child Language*, *38*(5), 1028–1054.

Olson, J., & Masur, E. F. (2012). Mothers respond differently to infants' familiar versus non-familiar verbal imitations. *Journal of Child Language*, *39*(4), 731–752.

O'Neill, M., Bard, K. A., Linnell, M., & Fluck, M. (2005). Maternal gestures with 20-month-old infants in two contexts. *Developmental Science*, *8*(4), 352–359.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716.

Owren, M. J., & Rendall, D. (2001). Sound on the rebound: Bringing form and function back to the forefront in understanding nonhuman primate vocal signaling. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews*, *10*(2), 58–71.

Pace, A., Luo, R., Hirsh-Pasek, K., & Golinkoff, R. M. (2017). Identifying pathways between socioeconomic status and language development. *Annual Review of Linguistics*, *3*, 285–308.

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. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*, *88*, 105918. https://doi.org/10.1016/j.ijsu.2021.105906

Panneton, R., Cristia, A., Taylor, C., & Christine, M. (2023). Positive Valence Contributes to Hyperarticulation in Maternal Speech to Infants and Puppies. *Journal of Child Language*, 1–11.

Panneton, R., Kitamura, C., Mattock, K., & Burnham, D. (2006). Slow speech enhances younger but not older infants' perception of vocal emotion. *Research in Human Development*, *3*(1), 7–19. https://doi.org/10.1207/s15427617rhd0301_2

Papoušek, M., Papoušek, H., & Symmes, D. (1991). The meanings of melodies in motherese in tone and stress languages. *Infant Behavior and Development*, *14*(4), 415–440. https://doi.org/10.1016/0163-6383(91)90031-M

Parsons, C. E., Young, K. S., Parsons, E., Stein, A., & Kringelbach, M. L. (2012). Listening to infant distress vocalizations enhances effortful motor performance. *Acta Paediatrica*, *101*(4), e189.

Pegg, J. E., Werker, J. F., & McLeod, P. J. (1992). Preference for infant-directed over adult-directed speech: Evidence from 7-week-old infants. *Infant Behavior and Development*, *15*(3), 325–345. https://doi.org/10.1016/0163-6383(92)80003-D

Perone, S., Madole, K. L., Ross-Sheehy, S., Carey, M., & Oakes, L. M. (2008). The relation between infants' activity with objects and attention to object appearance. *Developmental Psychology*, *44*(5), 1242-1248.

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), 1–14. 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.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–190.

Pickering, M. J., & Garrod, S. (2021). *Understanding dialogue: Language use and social interaction*. Cambridge University Press.

Pika, S. W., Ray; Kendrick, Kobin H. ;. Vernes, Sonja C. (2018). Taking turns: Bridging the gap between human and animal communication. *Proc R Soc B*, *285*(1880), 20180598. https://doi.org/10.1098/rspb.2018.0598

Pintar Breen, A. I., Tamis-LeMonda, C. S., & Kahana-Kalman, R. (2018). Latina mothers' emotion socialization and their children's emotion knowledge. *Infant and Child Development*, *27*(3), e2077.

Pisanski, K., Bryant, G. A., Cornec, C., Anikin, A., & Reby, D. (2022). Form follows function in human nonverbal vocalisations. *Ethology Ecology & Evolution*, *34*(3), 303–321.

Porritt, L. L., Zinser, M. C., Bachorowski, J.-A., & Kaplan, P. S. (2014). Depression diagnoses and fundamental frequency-based acoustic cues in maternal infant-directed speech. *Language Learning and Development*, *10*(1), 51–67.

Potter, C. E., & Lew-Williams, C. (2019). Infants' selective use of reliable cues in multidimensional language input. *Developmental Psychology*, *55*(1), 1.

Provenzi, L., Scotto di Minico, G., Giusti, L., Guida, E., & Müller, M. (2018). Disentangling the dyadic dance:

Theoretical, methodological and outcomes systematic review of mother-infant dyadic processes. *Frontiers in Psychology*, *9*, 348.

Pye, C. (1986). Quiché Mayan speech to children. *Journal of Child Language*, *13*(1), 85–100.

Radvansky, G. A., & Zacks, J. M. (2011). Event perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(6), 608–620.

Raneri, D. P. (2015). Infant-directed speech: Maternal pitch variability, rate of speech, and child language outcomes. *ProQuest Dissertations Publishing*, 1–56.

Räsänen, O., Kakouros, S., & Soderstrom, M. (2018). Connecting stimulus-driven attention to the properties of infant-directed speech-Is exaggerated intonation also more surprising? *Cognition, 178*, 193-206. https://doi.org/10.1016/j.cognition.2018.05.015

Rasenberg, M., Özyürek, A., & Dingemanse, M. (2020). Alignment in multimodal interaction: An integrative framework. *Cognitive Science*, *44*(11), e12911.

Rasenberg, M., Pouw, W., Özyürek, A., & Dingemanse, M. (2022). The multimodal nature of communicative efficiency in social interaction. *Scientific Reports*, *12*(1), 19111.

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

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

Ravignani, A. (2019). Timing of antisynchronous calling: A case study in a harbor seal pup (Phoca vitulina). *Journal of Comparative Psychology*, *133*(2), 272-277.

Ravignani, A., & de Reus, K. (2019). Modelling Animal Interactive Rhythms in Communication. *Evolutionary Bioinformatics Online*, *15*. https://doi.org/10.1177/117693431882355

Raviv, L., Lupyan, G., & Green, S. C. (2022). How variability shapes learning and generalization. *Trends in Cognitive Sciences*, *26*(6), 462–483.

Reilly, J. S., & Bellugi, U. (1996). Competition on the face: Affect and language in ASL motherese. *Journal of Child Language*, *23*(1), 219–239.

Rendell, L., Fogarty, L., & Laland, K. N. (2011). Runaway cultural niche construction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1566), 823–835.

Rifai, O. M., Fletcher-Watson, S., Jiménez-Sánchez, L., & Crompton, C. J. (2022). Investigating markers of rapport in autistic and nonautistic interactions. *Autism in adulthood*, *4*(1), 3-11.

Ritwika, V., Pretzer, G. M., Mendoza, S., Shedd, C., Kello, C. T., Gopinathan, A., & Warlaumont, A. S. (2020). Exploratory dynamics of vocal foraging during infant-caregiver communication. *Scientific Reports*, *10*(1), 10469.

Roberts, S., Fyfield, R., Baibazarova, E., van Goozen, S., Culling, J. F., & Hay, D. F. (2013). Parental speech at 6 months predicts joint attention at 12 months. *Infancy*, *18*, E1–E15.

Robinson, K., Schmidt, T., & Teti, D. M. (2005). Issues in the use of longitudinal and cross-sectional designs. *Handbook of Research Methods in Developmental Science*, 1–20.

Rocca, R., & Yarkoni, T. (2021). Putting psychology to the test: Rethinking model evaluation through benchmarking and prediction. *Advances in Methods and Practices in Psychological Science*, *4*(3), 25152459211026864.

Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, *1*(1), 27–42.

Rogoff, B., & Angelillo, C. (2002). Investigating the coordinated functioning of multifaceted cultural practices in human development. *Human Development*, *45*(4), 211–225.

Romeo, R. R., Segaran, J., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Yendiki, A., Rowe, M. L., & Gabrieli, J. D. (2018). Language exposure relates to structural neural connectivity in childhood. *Journal of Neuroscience*, *38*(36), 7870–7877.

Rossano, F. (2013). Sequence organization and timing of bonobo mother-infant interactions. *Interaction Studies*, *14*(2), 160–189. https://doi.org/10.1075/is.14.2.02ros

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.

Rotello, C. M., Heit, E., & Dubé, C. (2015). When more data steer us wrong: Replications with the wrong dependent measure perpetuate erroneous conclusions. *Psychonomic Bulletin & Review*, *22*, 944–954.

Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language*, *35*(1), 185–205.

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development*, *83*(5), 1762–1774.

Rowe, M. L., & Goldin-Meadow, S. (2009). Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science*, *323*(5916), 951–953.

Rowe, M. L., Leech, K. A., & Cabrera, N. (2017). Going beyond input quantity: Wh-questions matter for toddlers' language and cognitive development. *Cognitive Science*, *41*, 162–179.

Sacks, H. (1974). An analysis of the course of a joke's telling in conversation. *Explorations in the ethnography of speaking*, *8*, 337-353.

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology*, *69*, 181–203. https://doi.org/10.1146/annurev-psych-122216-011805

Saint-Georges, C., Chetouani, M., Cassel, R., Apicella, F., Mahdhaoui, A., Muratori, F., Laznik, M.-C., & Cohen, D. (2013). Motherese in interaction: At the cross-road of emotion and cognition?(A systematic review). *PloS One*, *8*(10), e78103.

Saito, Y., Aoyama, S., Kondo, T., Fukumoto, R., Konishi, N., Nakamura, K., Kobayashi, M., & Toshima, T. (2007). Frontal cerebral blood flow change associated with infant-directed speech. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, *92*(2), F113–F116.

Salomo, D., & Liszkowski, U. (2013). Sociocultural settings influence the emergence of prelinguistic deictic gestures. *Child Development*, *84*(4), 1296–1307.

Santesso, D. L., Schmidt, L. A., & Trainor, L. J. (2007). Frontal brain electrical activity (EEG) and heart rate in response to affective infant-directed (ID) speech in 9-month-old infants. *Brain and Cognition*, *65*(1), 14–21.

Schachner, A., & Hannon, E. E. (2011). Infant-directed speech drives social preferences in 5-month-old infants. *Developmental Psychology*, *47*(1), 19-25.

Schatz, J. L., Suarez-Rivera, C., Kaplan, B. E., & Tamis-LeMonda, C. S. (2022). Infants' object interactions are long and complex during everyday joint engagement. *Developmental Science*, *25*(4), e13239.

Scheel, A. M. (2022). Why most psychological research findings are not even wrong. *Infant and Child Development*, *31*(1), e2295.

Schertz, H. H., Call-Cummings, M., Horn, K., Quest, K., & Law, R. S. (2018). Social and instrumental interaction between parents and their toddlers with autism: A qualitative analysis. *Journal of Early Intervention*, *40*(1), 20–38.

Schertz, K. E., Kotabe, H. P., Meidenbauer, K. L., Layden, E. A., Zhen, J., Bowman, J. E., Lakhtakia, T., Lyu, M., Paraschos, O. A., & Janey, E. A. (2023). Nature's path to thinking about others and the surrounding environment. *Journal of Environmental Psychology*, 102046.

Schreiner, M. S., Zettersten, M., Bergmann, C., Frank, M. C., Fritzsche, T., Gonzalez-Gomez, N., Hamlin, K., Kartushina, N., Kellier, D. J., & Mani, N. (2022). *Limited evidence of test-retest reliability in infant-directed speech preference in a large pre-registered infant sample*. PsyArxiv: https://osf.io/preprints/psyarxiv/uwche

Schwab, J. F., & Lew-Williams, C. (2016). Language learning, socioeconomic status, and child-directed speech. *Wiley Interdisciplinary Reviews: Cognitive Science*, *7*(4), 264–275.

Segal, J., & Newman, R. S. (2015). Infant preferences for structural and prosodic properties of infant-directed speech in the second year of life. Infancy, 20 (3), 339-351.

Searl, J., & Evitts, P. M. (2013). Tongue–palate contact pressure, oral air pressure, and acoustics of clear speech. *Journal of Speech, Language, and Hearing Research*, *56*(3), 826–839.

Seidl, A., Tincoff, R., Baker, C., & Cristia, A. (2015). Why the body comes first: Effects of experimenter touch on infants' word finding. *Developmental Science*, *18*(1), 155–164.

Shneidman, L. A. (2010). *Language input and acquisition in a Mayan village*. The University of Chicago.

Shneidman, L. A., Arroyo, M. E., Levine, S. C., & Goldin-Meadow, S. (2013). What counts as effective input for word learning? *Journal of Child Language*, *40*(3), 672–686.

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

Shute, B., & Wheldall, K. (1999). Fundamental frequency and temporal modifications in the speech of British fathers to their children. *Educational Psychology*, *19*(2), 221–233. https://doi.org/10.1080/0144341990190208

Sidhu, D. M., & Pexman, P. M. (2019). The sound symbolism of names. *Current Directions in Psychological Science*, *28*(4), 398–402.

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., Bahník, Š., Bai, F., Bannard, C., & Bonnier, E. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*, *1*(3), 337–356.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11),

1359–1366.

Simpson, A. P. (2001). Dynamic consequences of differences in male and female vocal tract dimensions. *The Journal of the Acoustical Society of America*, *109*(5), 2153–2164.

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

Slobin, D. I. (2014). Before the beginning: The development of tools of the trade. *Journal of Child Language*, *41*(S1), 1–17.

Slobin, D. I., & Bowerman, M. (2007). *Interfaces between linguistic typology and child language research*. De Gruyter Mouton. https://doi.org/10.1515/LINGTY.2007.015

Smaldino, P. (2019). Better methods can't make up for mediocre theory. *Nature*, *575*(7783), 9–10.

Smaldino, P. (2023). *Modeling social behavior: Mathematical and agent-based models of social dynamics and cultural evolution*. Princeton University Press.

Smaldino, P. E. (2017). Models are stupid, and we need more of them. *Computational Social Psychology*, 311–331.

Smaldino, P. E., Calanchini, J., & Pickett, C. L. (2015). Theory development with agent-based models. *Organizational Psychology Review*, *5*(4), 300–317.

Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, *3*(9), 160384.

Smith, N. A., & Trainor, L. J. (2008). Infant-directed speech is modulated by infant feedback. *Infancy*, *13*(4), 410–420. https://doi.org/10.1080/15250000802188719

Smith, K. (2020). How culture and biology interact to shape language and the language faculty. *Topics in Cognitive Science*, *12*(2), 690–712.

Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in Cognitive Sciences*, *22*(4), 325–336.

Smith, N. A., & Trainor, L. J. (2008). Infant-directed speech is modulated by infant feedback. *Infancy*, *13*(4), 410–420. https://doi.org/10.1080/15250000802188719

Snow, C. E. (1977a). Mothers' speech research: From input to interaction. *Talking to Children: Language Input and Acquisition*, *3149*, 31-56

Snow, C. E. (1977b). The development of conversation between mothers and babies. *Journal of Child Language*, *4*(1), 1–22.

Snow, C. E. (2013). Social perspectives on the emergence of language. In *The emergence of language* (pp. 275–294). Psychology Press.

Snow, C. E., & Ferguson, C. A. (1977). *Talking to children: Language input and acquisition*. Cambridge University Press.

Soderstrom, M., Blossom, M., Foygel, R., & Morgan, J. L. (2008). Acoustical cues and grammatical units in speech to two preverbal infants. *Journal of Child Language*, *35*(4), 869–902.

Soltis, J. (2004). The developmental mechanisms and the signal functions of early infant crying. *Behavioral and Brain Sciences*, *27*(4), 477–490.

Song, 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

Soska, K. C., & Adolph, K. E. (2014). Postural position constrains multimodal object exploration in infants. *Infancy*, *19*(2), 138–161.

Soska, K. C., Adolph, K. E., & Johnson, S. P. (2010). Systems in development: Motor skill acquisition facilitates three-dimensional object completion. *Developmental Psychology*, *46*(1), 129-138.

Spinelli, M., Fasolo, M., & Mesman, J. (2017). Does prosody make the difference? A meta-analysis on relations between prosodic aspects of infant-directed speech and infant outcomes. *Developmental Review*, *44*, 1–18. https://doi.org/10.1016/j.dr.2016.12.001

Starns, J. J., Cataldo, A. M., Rotello, C. M., Annis, J., Aschenbrenner, A., Bröder, A., Cox, G., Criss, A., Curl, R. A., & Dobbins, I. G. (2019). Assessing theoretical conclusions with blinded inference to investigate a potential inference crisis. *Advances in Methods and Practices in Psychological Science*, *2*(4), 335–349.

Steffensen, S., & Pedersen, S. B. (2014). Temporal dynamics in human interaction. *Cybernetics & Human Knowing*, *21*(1–2), 80–97.

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

Stern, D. N., Spieker, S., Barnett, R., & 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

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, *17*(1), 3–45.

Stevens, K. N. (2000). *Acoustic phonetics* (Vol. 30). MIT press.

Stivers, T. E., N. J. ;. Brown, Penelope; Englert, Christina; Hayashi, Makoto; Heinemann, Trine; Hoymann, Gertie; Rossano, Federico; De Ruiter, Jan Peter; Yoon, Kyung-Eun; Levinson, Stephen C. (2009). Universals and cultural variation in turn taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(26), 10587–10592. https://doi.org/10.1073/pnas.0903616106

Stokes, R., & Hewitt, J. P. (1976). Aligning actions. *American Sociological Review*, 838–849.

Suanda, S. H., Smith, L. B., & Yu, C. (2018). The multisensory nature of verbal discourse in parent–toddler interactions. In Gogate, L. (Ed.) *Multisensory Perception and Communication* (pp. 56–73). Routledge.

Suarez-Rivera, C., Schatz, J. L., Herzberg, O., & Tamis-LeMonda, C. S. (2022). Joint engagement in the home environment is frequent, multimodal, timely, and structured. *Infancy*, *27*(2), 232–254.

Sundberg, U., & Lacerda, F. (1999). Voice onset time in speech to infants and adults. *Phonetica*, *56*(3–4), 186–199.

Szollosi, A., Kellen, D., Navarro, D. J., Shiffrin, R., van Rooij, I., Van Zandt, T., & Donkin, C. (2020). Is preregistration worthwhile? *Trends in Cognitive Sciences*, *24*(2), 94–95.

Takahashi, D. Y. ; N., Darshana Z. ;. Ghazanfar, Asif A. (2013). Coupled Oscillator Dynamics of Vocal Turn-Taking in Monkeys. *Current Biology : CB*, *23*(21), 2162–2168. https://doi.org/10.1016/j.cub.2013.09.005

Tamis-LeMonda, C. S., Kuchirko, Y., & Song, L. (2014). Why is infant language learning facilitated by parental responsiveness? *Current Directions in Psychological Science*, *23*(2), 121–126.

Tamis-LeMonda, C. S., Song, L., Leavell, A. S., Kahana-Kalman, R., & Yoshikawa, H. (2012). Ethnic differences in mother–infant language and gestural communications are associated with specific skills in infants. *Developmental Science*, *15*(3), 384–397.

Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development*, *72*(3), 748–767.

Tartter, V. C. (1980). Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception & Psychophysics*, *27*(1), 24–27.

Tarullo, A. R., John, A. M. S., & Meyer, J. S. (2017). Chronic stress in the mother-infant dyad: Maternal hair cortisol, infant salivary cortisol and interactional synchrony. *Infant Behavior and Development*, *47*, 92–102.

Templeton, E. M., Chang, L. J., Reynolds, E. A., Cone LeBeaumont, M. D., & Wheatley, T. (2022). Fast response times signal social connection in conversation. *Proceedings of the National Academy of Sciences*, *119*(4), e2116915119.

Templeton, E. M., Chang, L. J., Reynolds, E. A., Cone LeBeaumont, M. D., & Wheatley, T. (2023). Long gaps between turns are awkward for strangers but not for friends. *Philosophical Transactions of the Royal Society B*, *378*(1875), 20210471.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. MIT press.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*(1), 53–71.

Tomasello, M. (2009). *The cultural origins of human cognition*. Harvard university press.

Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural learning. *Behavioral and Brain Sciences*, *16*(3), 495–511.

Tomasello, M. (1999). Social cognition before the revolution. *Early Social Cognition: Understanding Others in the First Months of Life*, 301–314.

Trainor, L. J., Austin, C. M., & Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychological Science*, *11*(3), 188–195.

Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin & Review*, *9*(2), 335–340.

Trecca, F., Tylén, K., Højen, A., & Christiansen, M. H. (2021). Danish as a Window Onto Language Processing and Learning. *Language Learning*, *71*, 799–833. https://doi.org/10.1111/lang.12450

Trujillo, J. P., & Holler, J. (2023). Interactionally embedded gestalt principles of multimodal human communication. *Perspectives on Psychological Science*, *18*(5), 1136–1159.

Tsuji, S., Bergmann, C., & Cristia, A. (2014). Community-augmented meta-analyses: Toward cumulative data assessment. *Perspectives on Psychological Science*, *9*(6), 661–665.

Tsuji, S., Bergmann, C., Lewis, M., Braginsky, M., Piccinini, P., Frank, M. C., & Cristia, A. (2017). *MetaLab: A Repository for Meta-Analyses on Language Development, and More.* 2038–2039.

Turnbull, R. (2019). Listener-oriented phonetic reduction and theory of mind. *Language, Cognition and Neuroscience*, *34*(6), 747–768.

Twomey, K. E., & Westermann, G. (2018). Curiosity-based learning in infants: A neurocomputational approach.

Developmental Science, *21*(4), e12629.

Tylén, K., Fusaroli, R., Bundgaard, P. F., & Østergaard, S. (2013). Making sense together: A dynamical account of linguistic meaning-making. *Semiotica*, *2013*(194), 39–62.

Tylén, K., Weed, E., Wallentin, M., Roepstorff, A., & Frith, C. D. (2010). Language as a tool for interacting minds. *Mind & Language*, *25*(1), 3–29.

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.

van den Akker, O. R., van Assen, M. A., Enting, M., de Jonge, M., Ong, H. H., Rüffer, F., Schoenmakers, M., Stoevenbelt, A. H., Wicherts, J. M., & Bakker, M. (2023). Selective hypothesis reporting in psychology: Comparing preregistrations and corresponding publications. *Advances in Methods and Practices in Psychological Science*, *6*(3), 25152459231187988.

van Rooij, I., & Baggio, G. (2020). Theory development requires an epistemological sea change. *Psychological Inquiry*, *31*(4), 321–325.

Veissière, S. P., Constant, A., Ramstead, M. J., Friston, K. J., & Kirmayer, L. J. (2020). Thinking through other minds: A variational approach to cognition and culture. *Behavioral and Brain Sciences*, *43*, e90.

Verga, L., Kotz, S. A., & Ravignani, A. (2023). The evolution of social timing. *Physics of Life Reviews, 46*, 131-151.

Vermillet, A., Tølbøll, K., Litsis Mizan, S., C Skewes, J., & Parsons, C. E. (2022). Crying in the first 12 months of life: A systematic review and meta-analysis of cross-country parent-reported data and modeling of the "cry curve." *Child Development*, *93*(4), 1201–1222.

Vihman, M. (2009). Word learning and the origins of phonological systems. In *Language acquisition* (pp. 15–39). Springer.

Vihman, M. M. (1993). Variable paths to early word production. *Journal of Phonetics*, *21*(1–2), 61–82.

Vihman, M. M. (2017). Learning words and learning sounds: Advances in language development. *British Journal of Psychology*, *108*(1), 1–27.

Vogt, P., & Mastin, J. D. (2013). *Rural and urban differences in language socialization and early vocabulary development in Mozambique*. *35*(35).

Vosoughi, S., & Roy, D. K. (2012). A longitudinal study of prosodic exaggeration in child-directed speech. *Proceedings of the Speech Prosody, 6th International Conference*.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological functions. Cambridge: Harvard

University Press.

Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2014). A social feedback loop for speech development and its reduction in autism. *Psychological Science*, *25*(7), 1314–1324. https://doi.org/10.1177/0956797614531023

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143–2152.

Weisner, T. S., Gallimore, R., Bacon, M. K., Barry III, H., Bell, C., Novaes, S. C., Edwards, C. P., Goswami, B., Minturn, L., & Nerlove, S. B. (1977). My brother's keeper: Child and sibling caretaking [and comments and reply]. *Current Anthropology*, *18*(2), 169–190.

Werker, J. F., & McLeod, P. J. (1989). Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology*, *43*(2), 230. https://doi.org/10.1037/h0084224

Werker, J. F., & Tees, R. C. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology*, *50*(1), 509–535. https://doi.org/10.1146/annurev.psych.50.1.509

Wesner, J. S., & Pomeranz, J. P. (2021). Choosing priors in Bayesian ecological models by simulating from the prior predictive distribution. *Ecosphere*, *12*(9), e03739.

Whitfield, J. A., & Goberman, A. M. (2014). Articulatory–acoustic vowel space: Application to clear speech in individuals with Parkinson's disease. *Journal of Communication Disorders*, *51*, 19–28.

Whitfield, J. A., & Goberman, A. M. (2017). Articulatory-acoustic vowel space: Associations between acoustic and perceptual measures of clear speech. *International Journal of Speech-Language Pathology*, *19*(2), 184–194.

Whitfield, J. A., & Mehta, D. D. (2019). Examination of clear speech in Parkinson disease using measures of working vowel space. *Journal of Speech, Language, and Hearing Research*, *62*(7), 2082–2098.

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

Wilson, M. W., Thomas P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review*, *12*(6), 957–968. https://doi.org/10.3758/bf03206432

Xu, N., Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2013). Vowel hyperarticulation in parrot-, dog-and infant-directed speech. *Anthrozoös*, *26*(3), 373–380.

Xu, Y., & Chuenwattanapranithi, S. (2007). Perceiving anger and joy in speech through the size code. *Proceedings of the International Conference on Phonetic Sciences*, *16*, 2105–2108.

Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, *45*, e1.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122.

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, *125*(2), 244–262.

Zangl, R., & Mills, D. L. (2007). Increased brain activity to infant-directed speech in 6-and 13-month-old infants. *Infancy*, *11*(1), 31–62.

Zettersten, M., Cox, C. M. M., Bergmann, C., Tsui, A., Soderstrom, M., Mayor, J., Lundwall, R. A., Lewis, M., Kosie, J., & Kartushina, N. (2023). *Evidence for infant-directed speech preference is consistent across large-scale, multi-site replication and meta-analysis*. PsyArxiv: https://osf.io/preprints/psyarxiv/etqs7