LANGUAGE LEARNING AND DEVELOPMENT
https://doi.org/10.1080/15475441.2023.2277472

More Than Looks: Exploring Methods to Test Phonological Discrimination in the Sign Language Kata Kolok

Hannah Lutzenberger[a,b], Marisa Casillas[c,d], Paula Fikkert[e], Onno Crasborn[e], and Connie de Vos[b]

[a]Tilburg Center for Cognition and Communication, Tilburg University, Tilburg, Netherlands; [b]Department of English Language and Linguistics, University of Birmingham, Birmingham, England; [c]Language Development Department, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands; [d]Department of Comparative Human Development, University of Chicago, Chicago, USA; [e]Centre for Language Studies, Radboud University, Nijmegen, Netherlands

ABSTRACT

The lack of diversity in the language sciences has increasingly been criticized as it holds the potential for producing flawed theories. Research on (i) geographically diverse language communities and (ii) on sign languages is necessary to corroborate, sharpen, and extend existing theories. This study contributes a case study of adapting a well-established paradigm to study the acquisition of sign phonology in Kata Kolok, a sign language of rural Bali, Indonesia. We conducted an experiment modeled after the familiarization paradigm with child signers of Kata Kolok. Traditional analyses of looking time did not yield significant differences between signing and non-signing children. Yet, additional behavioral analyses (attention, eye contact, hand behavior) suggest that children who are signers and those who are non-signers, as well as those who are hearing and those who are deaf, interact differently with the task. This study suggests limitations of the paradigm due to the ecology of sign languages and the sociocultural characteristics of the sample, calling for a mixed-methods approach. Ultimately, this paper aims to elucidate the diversity of adaptations necessary for experimental design, procedure, and analysis, and to offer a critical reflection on the contribution of similar efforts and the diversification of the field.

Introduction

Language acquisition research is fundamentally biased toward a few European languages (Blasi et al., 2022; Kidd & Garcia, 2022; Lieven & Stoll, 2009; Pye, 2021), and the bias for spoken languages is even stronger (Lillo-Martin & Henner, 2021). With an increasing interest in the acquisition of rural languages and more and more sign languages being documented, we can get a peek into the diversity of language acquisition settings which may lead us to question some fundamental assumptions about how children are socialized with language. For example, while caregiver child-directed speech has long been considered fundamental for language learning, research has shown that in communities where caregiver child-directed speech is rare, children regardless learn the adult language on similar time-scales as children in high-density child-directed speech communities (e.g., Casillas et al., 2020a, 2020b; Cristia et al., 2019). This raises the question of not whether but *how* the bias toward urban languages impacts core theories of language acquisition and questions whether existing theories are reliable (see also Singh et al., 2022 for a discussion of this on perceptual narrowing).

CONTACT Hannah Lutzenberger h.lutzenberger@bham.ac.uk Tilburg Center for Cognition and Communication, Tilburg University, Tilburg, Netherlands

There is another fundamental tension in the acquisition literature that aggravates the language bias: urban languages are not only studied more frequently, they are studied primarily with experiments, while rural languages are often investigated using observational or ethnographic approaches (Cristia et al., 2023). This methodological dissociation is extremely prominent in sign language acquisition studies; studies on the acquisition of rural sign languages are scarce (Edward, 2022) yet using experimental approaches to study acquisition of rural signing varieties is virtually unattempted (Horton, 2018; Lutzenberger, 2023). Cristia et al. (2023) push for combining observational and experimental approaches in rural communities to make the leap toward cross-study comparability and validation of findings from languages under-represented in the scientific literature. While this will certainly work in favor of minimizing the methodological dissociation, it does not address the lurking question of how feasible and worthwhile such endeavors are : experimental studies in rural communities undergo many adaptations in terms of design, procedure, and analysis (Cristia et al., 2023; Frost & Casillas, 2021) and still often result in non-traditional data and/or non-significant results (Cristia et al., 2020). For researchers working in these communities, this raises the question of whether it is worthwhile to spend limited resources and time pursuing these experimental approaches rather than developing other methods better suited to small populations. In a nutshell, it is worth asking whether bringing lab paradigms into the context of small, rural language communities is a truly viable option for addressing the fundamental issues of bias and representation in scientific studies of language development.

This paper is dedicated to evaluating this question with a single case study. We report and reflect on the first experimental study of the acquisition of sign phonology in Kata Kolok (KK), a sign language that originated in a rural enclave in Bali due to sustained hereditary deafness (Marsaja, 2008). We elaborate on challenges and benefits associated with designing, conducting, and analyzing this experiment, focusing on adapting familiar experimental paradigms to child signers growing up in a rural context: we worked with a low-tech familiarization paradigm and aimed to test children's ability to discriminate between minimally different phonological sign stimuli. As the main outcome of this study, we identify challenges and interesting avenues for further developments. We also critically evaluate how such studies can contribute to the body of literature on language acquisition and serve to build, sharpen, and corroborate theories about how children learn language, specifically phonology.

Dual adaptations: Adapting experimental methods to rural contexts and signing child participants

Many of the widespread and foundational measures of language processing assume a hearing participant using a spoken language, and theories of psycholinguistics emerge from a limited set of socioculturally similar languages (see also Singh et al., 2022). As existing paradigms cannot be straightforwardly extended to signing participants or rural communities (Cristia et al., 2023), there is the potential for critical gaps in psycholinguistic theories. In the present case study, we are dealing with a dual adaptation: (i) Adapting an experiment that is well-established with children acquiring languages in urban contexts to a rural context (*field adaptation*), and (ii) adapting a paradigm that is well-established with children learning a spoken language to children learning a signed language (*modality adaptation*). In the following, we review methodological adaptations of both kinds with a special focus on the early acquisition of sign phonology for the modality adaptations.

Field-adaptations

With the increasing interest in carrying out experiments on the acquisition of rural languages, we are continuously learning about how established experimental paradigms need to be adapted to match rural communities. Adaptations include simple translations into the target language, additional instructions for populations that are not familiar with experimental setups, ensuring that participants and collaborators understand the task, practical adaptations to the experimental environment, technical equipment, recruitment strategies, etc. (Cristia et al., 2023).

A variety of methodological, technical, and design adaptations are needed for different communities. Sarvasy et al. (2023) successfully use mobile eye-tracking for investigating clause-planning in Nungon, a small Papuan language in rural Papua New Guinea (PNG). Within the same language community, Sarvasy et al. (2019) also explore the use of mobile electroencephalography (EEG) to measure brain responses (P600 and N400 effects) to different types of syntactic violations and included an exit survey targeting the experience of participants during the experiment. Mulak et al. (2021) conduct two well-established word learning paradigms (fast-mapping and cross-situational word learning) with adult participants from rural PNG; only after major adaptations, fast-mapping but not cross-situational word learning was administered successfully. Similarly, Frost and Casillas (2021) ensured task understanding and comfort by making major changes to stimuli, instructions, and the number of practice items in a statistical learning experiment with speakers of Yélî Dnye used on Rossel Island (PNG), but still did not manage to conceptually replicate the original findings their study was based on.

Adaptations may be community-specific. Cristia et al. (2020) report challenges in using their non-word repetition elicitation task with child speakers of Tsimane' (Bolivia). After their initial design using sound playback failed, they revisited their design to (i) a group game with children and adults in which Tsimane' research assistants act as a model for the other participants, (ii) update stimuli gradually while in the field to increase their naturalness, and (iii) allow group-size to vary flexibly between three and nine participants according to availability. In another community, child speakers of Yélî Dnye (PNG) respond positively to a more traditional design using playback (Cristia & Casillas, 2022). The only necessary adaptations were language-appropriate stimuli and the distribution of stimuli across children to maximize data obtained from participants.

Besides adaptations concerning the experimental design, often the testing environment or experimental protocols are altered as well. In remote communities with unstable or non-existent electricity, equipment needs to be used that is solar-powered or that can be connected to backup batteries (Frost & Casillas, 2021; Sarvasy et al., 2023). Instead of a sound-attenuated lab environment, data collection may vary in location and setup, e.g., using portable devices such as tablets or laptops to visit participants' houses (Cristia & Casillas, 2022; Frost & Casillas, 2021). Informed consent may be obtained verbally (Cristia et al., 2020), participant compensation may vary (Cristia et al., 2020; Frost & Casillas, 2021), and personal details of participants such as age, years of schooling, and literacy skills may be rough estimates (Cristia et al., 2020). In the case of remote signing communities, researchers who have worked with homesigners in Nicaragua have emphasized the importance of designing tasks that need little or no language and minimal instructions due to limited experience with formal education and (picture-)literacy (Brentari et al., 2017; Coppola, 2002; Gagne & Coppola, 2017). Most importantly, the comfort of the participants (and caregivers) with participation and with the presence of the researcher, often a foreigner, may lead to additional adjustments, e.g.,, non-ideal use or placement of equipment (Cristia & Casillas, 2022).

While the number of studies on sign languages used in rural contexts is steadily growing (de Vos & Pfau, 2015; Zeshan & de Vos, 2012), most studies are based on spontaneous data or picture elicitation tasks rather than experimental data and studies on the acquisition of these languages are rare (de Vos, 2012a; Horton, 2018; Hou, 2016; Lutzenberger et al., 2023). This holds also more generally for studies on the acquisition of sign phonology; with the exception of Mann et al. (2010) adapted non-word repetition task with older children, all studies focusing on how signing children acquire phonology are based on spontaneous data, to our knowledge (e.g., Bonvillian & Siedlecki, 1996; Cheek et al., 2001; Conlin et al., 2000; Karnopp, (2008); Lutzenberger et al., 2023; Meier, 2006; Morgan, 2006; Siedlecki & Bonvillian, 1993). Thus, there is a great deal yet to learn about how to adapt experimental measures of developing phonology to rural signing communities.[1]

[1]Note that studies adapting established methods to the field setting exist but either target adults or older children and/or do not focus on sign phonology, which is why they are not being included in this review. For examples, consult work on Nicaraguan Sign Language (e.g., Gleitman et al., 2019; Pyers & Senghas, 2009; Senghas & Coppola, 2001).

While experiment-based measures of language development are still in their infancy for remote communities, the pattern of work so far teaches us that the adaptations required are often intricate, and multifaceted, and vary from community to community and task to task. More and more studies of rural communities are needed to allow us to explore and corroborate the generalizability of existing paradigms and the claims and theories made about language (acquisition).

Modality-adaptations: Experiments on the acquisition of sign phonology

Instead of focusing on diversity, sign language research has often concentrated on proving their status as fully fledged languages or on testing linguistic universals. Like spoken languages, sign languages show phonology (Stokoe, 1960). Signs are organized sub-lexically in features (e.g., hand configuration, location, movement type, movement shape, etc.) (for various theoretical accounts of sign phonology, see, e.g.,, Brentari, 1998, 2019; Eccarius & Brentari, 2010; Kooij, 2002; Sandler, 1989, 2012) and feature contrasts may lead to minimal pairs of signs where two signs share all but one feature (Figure 1).[2] For example, the KK signs PRAY and SHY differ only in location (forehead in PRAY; chin in SHY; Figure 1a) and THINK and PRAY only in handshape (extended index in THINK; flat handshape in PRAY; Figure 1b). In the following, we discuss the literature on modality adaptations relevant to this study, i.e., targeting the early acquisition of sign phonology.

Experiments studying the acquisition of sign phonology by young children are rare. Most studies have adapted two major paradigms used with speaking children targeting the ability to discriminate between linguistic contrasts: the habituation paradigm (Werker et al., 1998; Stager & Werker, 1997; for a recent review see; Fennell et al., 2017) and the preferential looking paradigm (for an overview see Golinkoff et al., 2017). The habituation paradigm generally exposes participants (usually infants) repeatedly to the same stimulus until their interest, measured in looking time, sucking time, or heart rate, decreases. Then, they are exposed to a novel stimulus which, upon discrimination, is expected to elicit higher interest than the familiar one (novelty effect or dishabituation). The preferential looking paradigm generally capitalizes on looking time differences between two visual stimuli that are presented simultaneously; increased interest in one of them is interpreted as a preference. In both cases, adaptations to the visual spatial modality of sign languages are necessary; for example, eye gaze control and eye contact are central to signed communication (Lieberman et al., 2014) and therefore experiments involving signed stimuli or signing participants require close monitoring of eye gaze behavior.

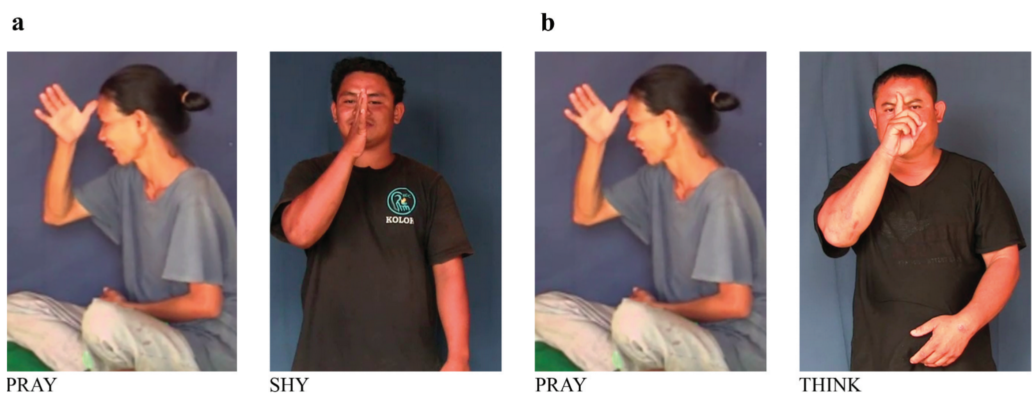

Figure 1. Minimal pairs of signs in KK; minimal pairs of location (a) and of handshape (b).

[2]For an excellent and detailed discussion about sign phonology in Kenyan Sign Language, see Morgan (2017).

Using a habituation paradigm, Carroll and Gibson (1986; habituation-dishabituation task) and Schley (1991; habituation-recovery task) show that infants who are naïve to American Sign Language (ASL) are sensitive to movement contrasts of ASL signs. Carroll and Gibson (1986) expose 4-month-olds to minimal pairs of movement in ASL and find that infants differentiate global and single movement contrasts based on differences in looking time, but this is not the case for all movement dimensions. Schley (1991) disentangle movement type and semantic context by testing hearing infants (aged 3.5 months) without ASL exposure on aspectual (in Schley, 1991 "inflectional") movement manipulations within the same ASL sign (LOOK-AT). After three tokens of a movement type, infants look longer to a novel movement type than to a new member from the same category. Wilbourn and Casasola (2007) habituate ASL-naïve infants (aged 0;4 and 0;10) on the ASL sign FINISH and then test whether they detect manipulations on single parameters, based on their looking behavior. Infants identify manipulations of location and nonmanual markers but not handshape or movement.

Using a preferential looking paradigm, Krentz and Corina (2008) present hearing infants without ASL exposure (aged 0;6 and 0;10) with short clips of ASL narratives and non-linguistic pantomime on two different screens. Their younger but not older participants look longer at ASL stimuli, i.e., preferred language over pantomime. Using both a habituation and a preferential looking paradigm, Blau's (2019) preliminary results seem to suggest successful discrimination in U.S. infants (deaf signers, hearing signers, hearing non-signers) when exposed to one of the two unknown sign languages (German Sign Language and Russian Sign Language) or ASL and SEE (Signed Exact English).

Others combine both paradigms (Baker et al., 2006; Palmer Baker et al., 2012). Baker et al. (2006) use pictures of ASL handshapes from a continuum that adult native ASL signers perceive categorically (Baker et al., 2005; but see also Best et al., 2010; Emmorey et al., 2003 for inconclusive evidence for categorical perception among adult signers), and present them in blocks of four handshapes in several conditions of either identical or different tokens from the same or different categories. They determine whether children discriminate between ASL handshapes and non-ASL handshapes based on looking times to handshapes that change across linguistic category boundaries, mirroring the pattern of adult signers. ASL-naïve infants show a preference for ASL handshapes only before age 1;0 (Baker et al., 2006) but this preference persists after 1;0 for hearing infants with ASL exposure (Palmer Baker et al., 2012). In this case, preference is measured in terms of looking duration and location.

Mann et al. (2010) adapt a non-word repetition task of 40 nonsense signs, with 91 deaf children acquiring British Sign Language (BSL; aged 3;0–11;0) and 46 hearing non-signing children (aged 6;0–11;0). Children are presented with stimuli and asked to repeat them. They find that phonological complexity (handshape and movement) impacts sign production in all children; deaf children increasingly use their knowledge of BSL, reflecting in more accurate repetitions in older children.

Besides a range of different paradigms, the types of stimuli vary across studies. Krentz and Corina (2008) and Blau (2019) use naturalistic narratives; Mann et al. (2010) and Wilbourn and Casasola (2007) use video recordings of a signer producing isolated (nonsense) signs; Carroll and Gibson (1986) use cropped videos that only show the hand of the signer in front of the torso; Palmer and colleagues (Baker et al., 2006; Palmer Baker et al., 2012) show pictures of only the hand.

Overall, experiments studying the acquisition of sign phonology are somewhat scarce, are largely based on ASL-stimuli, and are mostly administered to ASL-naïve children or hearing bimodal bilinguals of English and ASL (exceptions Blau, 2019; Mann et al., 2010). Thus, most studies target the loss of sensitivity to signed phonemic contrasts (also known as *perceptual narrowing*) rather than the acquisition of sign phonology in signing children.

Kata Kolok

The sign language Kata Kolok (KK) arose due to high levels of congenital deafness in a Balinese village of ~3,000 people (Friedman et al., 1995; Winata et al., 1995). Today, 33 permanent residents are deaf

(Lutzenberger, 2019). Since its emergence, KK has become entrenched in the village's cultural and linguistic landscape (de Vos, 2012b; Marsaja, 2008).

Like other traditional Balinese villages, this community has strong kinship ties. Clan membership, regulated through a patrilineal system, organizes the village geographically and socially (Covarrubias, 1937; Marsaja, 2008). Within clans, villagers live in family compounds, that is arrays of multigenerational households with a shared courtyard. This social structure distributes childcare among members of the nuclear family and relatives within the same family compound.

The tight-knit community structure also leads to a high percentage of hearing signers. This creates a rich and diverse linguistic environment for deaf (and hearing) children who learn KK from birth. After a decade without the birth of deaf babies, eight babies were born into deaf households between 2014 and 2018 – two deaf and six hearing children have at least one deaf caregiver and are learning KK as (one of) their primary language(s).

An in-depth study of KK phonology is ongoing (Lutzenberger, 2022). However, the lexicon shows typological peculiarities. KK signs use locations that are unusual in other sign languages, e.g., the hip, the teeth, or the tongue (de Vos, 2012b; Marsaja, 2008). In addition, the handshape inventory is small and includes many basic handshapes (de Vos, 2012b; Lutzenberger, 2018; Lutzenberger et al., 2019; Marsaja, 2008). Variation in sign form is attested within the community (Lutzenberger et al., 2021), and appears to be governed by factors such as hearing status and gender (Mudd et al., 2020).

In sum, the sign language KK is integral to the community's social and linguistic environment. Alongside deaf signers, many hearing signers provide a language model for children who acquire KK from birth. The KK lexicon shows cross-signer variation, a tendency for unusual locations, and a handful of frequently used handshapes, whose phonetic or phonemic status is yet to be determined.

This study

The methodological dissociation between studies of phonological acquisition in urban and in rural communities is paired with a disproportional scientific interest in spoken over signed languages. Therefore, there exist no experimental studies investigating the acquisition of sign phonology by children who acquire a rural sign language. The present study sets out to address this inequality by presenting the first experiment conducted with child signers of KK, an experiment adapted to this specific rural fieldwork context; we applied a dual adaptation to the visual familiarization paradigm to be suitable for young child signers of KK, with data being collected during a field trip in 2018. Testing sign language acquisition in the KK community using experiments is unique and time-sensitive, which is why we opted for a low-tech experiment; after no new birth into deaf families between 2005 and 2013, eight children were born to deaf parents since 2014. Given that experiments to investigate phonological acquisition typically target young children, it was critical to implement an experiment rather quickly. While in the following we present our experiment in the traditional style of an empirical research article on child language development, the ultimate goal of this paper is to elucidate the various kinds of adaptations that were necessary throughout the design, procedure, and analysis of the experiment and to offer a critical reflection of these efforts to evaluate the contribution of such endeavors.

Low-tech visual familiarization paradigm in the field

We conducted a low-tech visual familiarization experiment in the field. Visual familiarization differs from visual habituation in that children are exposed to the same stimulus for a fixed number of showings and then presented with a novel stimulus. We predicted different behaviors between signing and non-signing children; due to KK exposure, we expected that signing children would be sensitive to the handshape contrasts and, thus, look longer to the novel stimuli than non-signing children. This prediction is based on evidence that mature learners, in our case the signing children given their experience with KK, more often show a preference for novelty than non-experienced learners, i.e., our non-signing children (Hunter & Ames, 1988). Note that it has been reported that some children exhibit a preference for the familiar stimulus after familiarization (familiarity effect) (Fennell et al., 2017; Oakes, 2010).

Methods

Participants

Since we are primarily interested in the acquisition of KK phonology, we exhaustively sampled the community for young children who had at least one deaf primary caregiver and could thus be considered to learn KK natively. The resulting *signing* group consisted of eight native KK signers between the ages of 0;4–4;4; two deaf monolinguals and six hearing bimodal bilinguals (spoken Balinese and KK). In addition, we tested age-matched *non-signing* children who were recruited by two main criteria: (i) the closest match in registered birth date to one of the signing children and (ii) belonging to families with very limited or no signing skills. Note that we did not find an age-match for one signing child (CSR) and in cases where the non-signing participant was very fussy, we double-matched signing children, resulting in more non-signing than signing children. Signing children were recruited directly by HL who is familiar with deaf villagers and has been conducting mid- to long-term data collection on these sign acquiring children since 2014 (*Kata Kolok Child Signing Corpus (KKCSC)* (2007-2023). Non-signing children were recruited with the help of a local hearing research assistant through the use of the local nurse's registry. Find an overview of the sample in Table 1.

Stimuli

The stimuli consist of videos of an animated monkey producing KK signs and mispronunciations (non-word signs).[3] Previous elicitations with adult KK signers revealed a strong focus on identifying people, places, and events rather than focusing on the content presented to provide targeted responses (such as describing a picture). Given that the current study aims at differentiating between pairs of minimally different signs, we opted for animated stimuli rather than using a human signer to minimize distractions from the task, e.g., through trying to identify the signer. Monkeys are ideal for two reasons: (i) they are native to Bali and thus culturally appropriate and (ii) their anatomy is

Table 1. Overview of participants. Naming convention follows the codes used for signing children who are included in the KKCSC. Generally, the composition is decoded as follows: CS for signing children vs. CNS for non-signing children, followed by one or two letters as an acronym. Age-matched children can be identified by sharing the same last letter in the string, e.g. CSA and CNSA. CSA and CSC are the only deaf children in the sample.

signing group					*non-signing* group		
participant	age (months)	hearing status	gender	deaf caregiver	participant	age (months)	gender
CSA	49	deaf	female	two deaf parents	CNSA	49	female
					CSNAII	52	female
CSR	45	hearing	female	deaf grandfather	-	-	-
CSHM	32	hearing	male	deaf grandparents	CNSM	31	female
CSD	23	hearing	female	one deaf parent	CNSD	22	male
CSS	18.5	hearing	male	two deaf parents	CNSS	19	male
					CNSSII	19	male
CSC	18	deaf	female	one deaf parent	CNSC	18	female
CSW	18	hearing	male	two deaf parents	CNSW	17.5	male
CST	4	hearing	female	two deaf parents	CNST	4	female

[3]We assume that our mispronunciations are non-word equivalents. In order to rule out that, we unintentionally created nonce signs, i.e. signs that are impossible in KK, more research into KK phonology would be necessary as well as further testing of the stimuli on different populations.

similar to human anatomy, allowing the animated character to sign naturally. The skeleton of the monkey was designed by Jeroen Derks (Max Planck Institute for Psycholinguistics, Nijmegen) and the animations of each stimulus video by HL using the animation software MAYA Autodesk[4] (version 2018).[5]

Signs consist of a preparatory phase, the stroke, and a retraction phase (Kita et al., 1998) with peak informativeness at the stroke (Figure 2). We animated stimuli in three equally long phases where the stroke was initiated at the 24th frame and released at the 48th frame. Videos were rendered in high quality at 24 frames per second and edited in Adobe Premiere to video clips of 02.26 seconds, showing one complete sign cycle (Figure 2). All stimulus videos show the monkey's full body from a frontal perspective.

Stimuli were selected based on familiarity and form. We judged these by (i) the KK dataset in Global Signbank (Lutzenberger, 2020) that is based on elicited and spontaneous data from deaf adults, (ii) field-observations, and (iii) language fluency in KK of two of the authors (HL and CdV). Specifically, familiarity was assessed based on frequency in the input and early production by children; that is, we chose signs that deaf caregivers reported known by the (signing) children and that have been observed frequently in the input and early production during personal fieldwork HL and CdV and in naturalistic recordings of signing children within the KKCSC. All chosen signs COW were one-handed signs with frequent handshapes where handshape frequency was based on the lexical database. We selected the following signs: COW, FATHER, GRANDPARENT, MOTHER, PRAY, STAY, THIRSTY, BATHE, CRY, DOG, RAIN, SHY.

Stimuli were animated in pairs: a target production and a mispronunciation. We manipulated signs for handshape since research suggests categorical perception of this parameter in ASL (Baker et al., 2005; Emmorey et al., 2003; but see; Best et al., 2010), and it is often considered the most contrastive property of the sign (Johnston & Schembri, 1999). We manipulated signs in two ways: all form aspects of the sign were kept identical, but we modified whether fingers were extended or bent at the finger base joint (Type 1; COW, FATHER, GRANDPARENT, MOTHER, PRAY) and whether fingers were straight or curved (Type 2; BATHE, CRY, DOG, SHY) (see Figure 3). For example, the target sign PRAY is produced with all fingers extended (B handshape) and thumb contacting the forehead. For the Type 1 manipulation, we created an exact copy except that all fingers are bent into a fist (A handshape). Similarly, for the target sign SHY all fingers are extended and spread, thumb contacting the chin. For the Type 2 manipulation, we changed the extended fingers to be curved, which might make changes in these signs perceptually more subtle than Type 1 manipulations. Given this and the large age range among our participants, signs were arranged in two blocks according to the type of manipulation: Type 1 and Type 2 (Figure 3; full stimuli list in Appendix A; videos available under https://osf.io/w62tm/).

Prior to the experiment, all stimuli were piloted with three deaf adult native KK signers in order to assess their recognizability, clarity, and whether they were known to young children. During this pilot

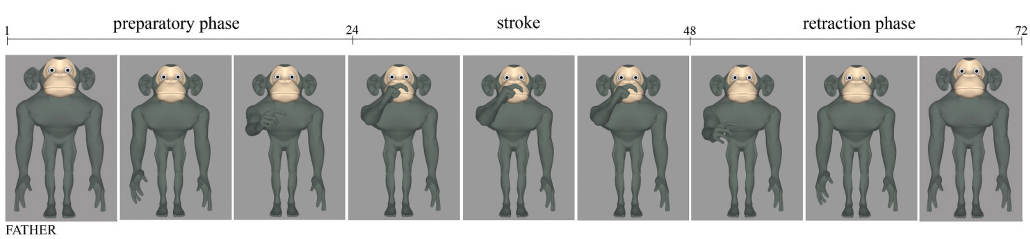

Figure 2. Example of animated stimulus sign father.

[4]https://www.autodesk.com/products/maya/overview?term=1-YEAR&support=null.
[5]Source code, materials, and stimuli can be found on the Open Science Framework (https://osf.io/w62tm/).

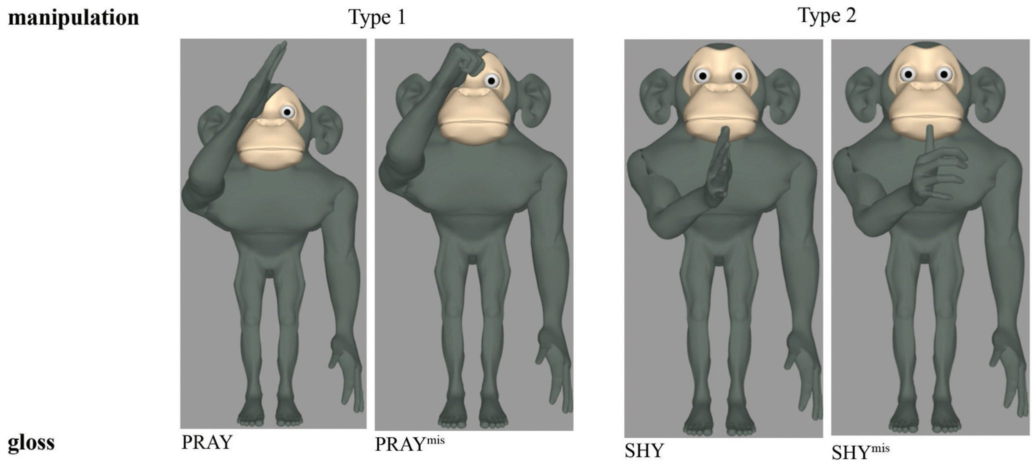

Figure 3. Example of pairs of stimuli.

we asked the deaf adults, all of them relatives of the signing children, to identify and reproduce the sign, to judge whether this is a valid sign in KK or not, and whether their children would be familiar with the sign.[6] Whenever participants recognized and judged the sign correctly, identified the manipulation as a variant that does not exist in KK, and stated that the children would know the sign, we kept the sign, else signs were eliminated. We eliminated three signs (THIRSTY, STAY, RAIN) since they were not identified correctly by all signers.

Lastly, we created four filler animations using Apple Keynote. We designed and animated prototypes of balloons, confetti, fish, and monkeys, crossing the screen vertically or horizontally at different speeds. Filler animations lasted 30 seconds.

Procedure

Parental consent was obtained before the experiment and after explaining the goal and the procedure. For deaf parents, HL provided explanations in KK, and for hearing parents, a hearing research assistant communicated the information in Balinese and translated questions. After the experiment, children were gifted a soap bubble tube or a stamp. Accompanying caregivers were compensated with the equivalent of a day's salary.

The experiment was presented on a MacBook Pro (13.3 inches). The laptop was placed on a 40-cm-high stool in a quiet room at a house under construction at the edge of the village. As recreating a completely isolated lab setting would be impossible in this community, we tried to minimize distractions as much as possible. Children sat on the floor or their caregiver's lap, facing the laptop, and the experimenter sat next to the laptop, facing the child and caregiver (see Figures 4 and 10 for examples). Caregivers were instructed not to respond to the stimulus or disturb the child, however, they were allowed to encourage the child to continue watching the screen and to interact with the child if required. HL conducted all the experiments, interacting directly with the children as an experimenter. During the experiment, the experimenter's task was primarily to observe the child and advance trials; when a child (or caregiver) insisted on interaction, the experimenter would interact minimally with the child, e.g. by nodding or pointing to the screen. The experimenter interacted freely with the child and caregiver during breaks. No other people were in the room. Each session was videotaped with three cameras: (i) the front laptop camera recorded the child's face and eye movements, (ii) one Canon Legria HF G26 camera behind the laptop recorded a frontal view of the child

[6]For details about a slightly larger perception based study using the same methodology, see Lutzenberger (2022).

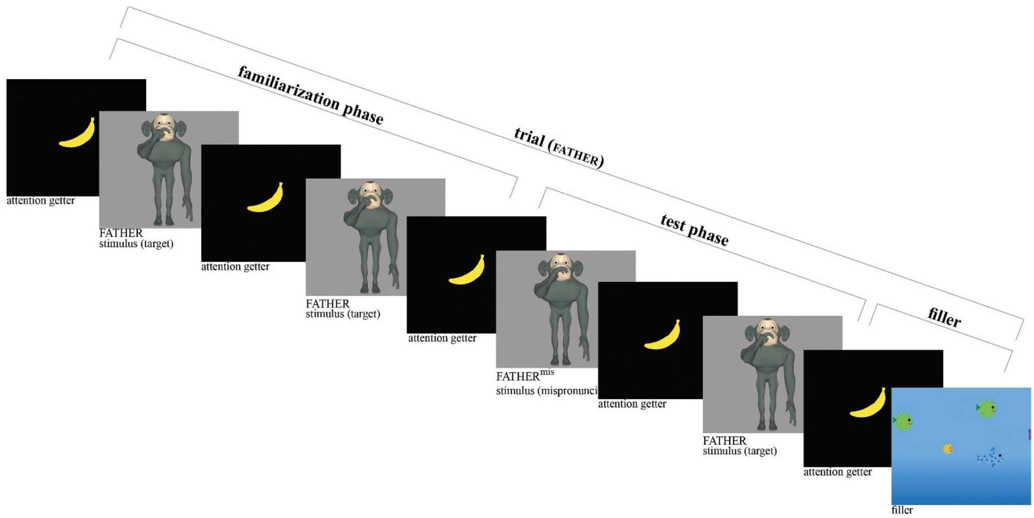

Figure 4. Example of the procedure.

and (iii) another Canon Legria HF G26 camera behind the child recorded the laptop screen and child–caregiver interactions.

Children were tested in a visual familiarization paradigm. All children started with a practice trial, showing COW and COW^{mis} repeatedly in semi-random order, to familiarize the child with the experiment. Experimental trials consisted of a familiarization phase showing the same stimulus five times and a test phase consisting of a novel stimulus and the familiar stimulus (Figure 4). The experimenter directed a slideshow using the forwards key on the laptop. We used three types of screens: (i) black screen with a blinking attention-getter in the center, (ii) stimulus video, and (iii) filler animation. Each stimulus was preceded by an attention getter to direct the child's gaze to where the stimulus would appear. Once the child's gaze fixated on the attention getter, a stimulus was shown on loop for as long as the child looked at the screen. When the child averted their gaze, the experimenter proceeded to the next attention getter. Each completed test phase was followed by a filler animation.

Each child watched as many trials as possible before they lost interest in the videos. We counterbalanced the order of stimuli, and whether children were familiarized with the target or the mispronunciation. Given the age range of our participants, the order of stimuli was randomized within blocks; each child started with Type 1 manipulations and only afterward proceeded to Type 2 manipulations.

If the child got too fussy to continue, the experiment was paused and the child played with soap bubbles until they were ready to continue. In order to obtain the maximal amount of data from each participant, we allowed the data collection to be distributed over two visits.

Annotation

HL annotated all data in ELAN [Computer software] (5.9) (2020) (Figure 5). We annotated each trial (e.g., *father*) for the phase (familiarization vs. test) and the presented stimulus (target vs. mispronunciation). We also annotated sign cycles and strokes. Looking time was coded (i) tolerantly where looks with off-screen glances of less than 1 s are allowed between fixation periods, and (ii) conservatively where each continuous fixation was annotated. Despite the two approaches to coding, we only report the first approach given the high amount of variation in the data and limited data points. We excluded looks where the child saw less than half of the sign's stroke. We report the proportion of time the child spent looking at each trial.

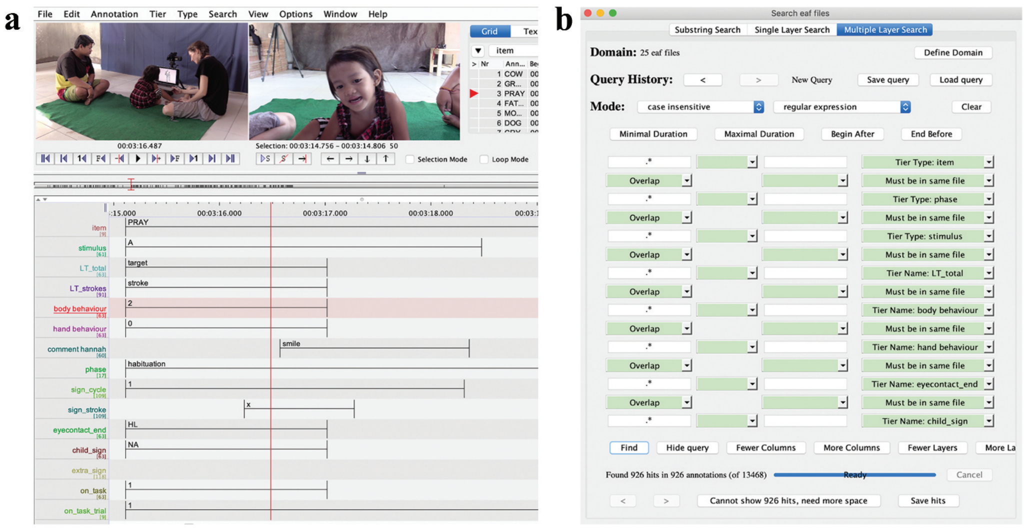

Figure 5. Example of ELAN file and multiple layer search query used for data export.

In addition, we coded for three behavioral measures: attention, eye contact, and hand activity.

(1) Attention. Attention is measured and reported in two ways: (i) attention span, and (ii) attention level. First, attention span is calculated based on how many items the child completed. For example, if child1 watched COW, FATHER, and PRAY and child2 watched COW, GRANDPARENT, FATHER, MOTHER, PRAY, BATHE, and DOG before getting too distracted to continue the experiment, child2's attention span is larger than that of child1. Second, attention level is based on rating how distracted a child was from the task. We assigned scores from 1 (fully attentive) to 5 (fully distracted) for each trial, reflecting the baseline and the progression of focus/distraction from the task by a child over the entire experiment. For example, if child1 was very attentive and sat still during the first two items but then started being very distracted with running around and having to be summoned back to the task, their attention level might have been rated with 1 for the first two items and with 4 for the third item.

(2) Eye contact. Given the central role of eye contact in signed communication, we coded whether the child made eye contact with an interlocutor at the end of a stimulus presentation (categorical variable). For example, if child1 looked to the experimenter or to the accompanying caregiver after having seen a stimulus, it would be coded as eye contact; it would be coded as no eye contact if child1 just looked to the floor or elsewhere in the room.

(3) Hand activity. We rated how actively children engaged their hands during each stimulus showing using a mixed scoring of Likert scale (1-5, with 1 being still and 5 being fully in motion) and categorical for gestures or signs (coded as S for both). A mixed scoring was necessary to disentangle communicative hand movements, i.e., signs and gestures, from unrelated ones, e.g., nose-picking or beat gestures. For the sake of simplicity, this measure includes communicative responses that are non-manual such as expressing negation through a headshake. For example, if child1 pointed to the screen while observing a stimulus, this would be coded as sign/gesture (S), whereas any handwringing or rubbing hands on the thighs would qualify for numeric scoring of activity level based on how big the movement was.

We used multiple systematic searches in ELAN via the function Multiple Layer Search to export data (Figure 5) and performed analyses in R (R Core Team, 2019). A second coder recoded roughly 20% of the data for all three behavioral measures. The two coders reached high levels of inter-coder agreement for eye contact (Cohen's Kappa = 0.933), and for hand activity in terms of identifying communicative behavior (Cohen's Kappa = 1). For attention levels, the agreement was low (Cohen's Kappa = 0.326) due to different baseline scorings; in most cases the given score differed by a single digit (e.g. coder1 rated the attention level as 2 and coder2 rated it as 1) but the difference in scoring was higher than a single digit in 12.8% of the data.

Analysis

This study takes a mixed-methods approach. First, we test the hypothesis that only signing children discriminate between the target and the mispronunciation with a mixed-effects model, using R (R Core Team, 2019) and the lme4 package's lmer() function (Bates et al., 2015). We chose mixed-effects modeling over ANOVA as it deals better with few(er) data and (more) variation. We log-transformed looking time to reduce the positive skewness of the data (Winter, 2019). The model tests whether group (signing vs. non-signing), stimulus (novel vs. familiarized), and block (Type1 vs. Type2) predict looking time during the test phase, with random intercepts for participant and item (lmer($LookingTimeAtTest_{log}$ ~ group + test_condition + block + (1|participant) + (1|item))). We coded the contrasts using sum coding, following Alday (n.d.). Second, the heterogeneity and small size of the sample warrant additional qualitative analyses. We hypothesize that language skills in KK affect behavior so that signing children behave more attentively and more communicatively than non-signing children. We analyze three behavioral measures: attention, eye contact, and hand activity.

Findings

Data annotation yielded 989 looks, of which 25.7% were excluded due to the child seeing less than half the stroke ($N = 210$) or an incomplete test phase ($N = 44$). The remaining 751 looks, 370 looks from non-signing and 395 looks from signing children, were used in the analyses. Raw data can be found on OSF.

Looking time

For the analyses of looking time, we included only those trials that had at least five observations in the familiarization phase (habituation criterion) and two observations in the test phase (test phase: $N_{signing}$ = 68; $_{non\text{-}signing}$ = 64). The two groups do not show great differences in their looking behavior during the test phase ($Mean_{signing} = 8.34$; $SD_{signing} = 0.74$; $Mean_{non\text{-}signing} = 8.40$; $SD_{non\text{-}signing} = 0.86$) (Figure 6). In the Type 1 block, both groups looked longer during the familiarization than during the test phase. In the Type 2 block, the signing group looked marginally longer for stimuli in the test phase. However, neither of the two groups looked significantly longer to the novel stimulus than to the familiarized stimulus. Indeed, the model (lmer($LookingTimeAtTest_{log}$ ~ group + test_condition + block + (1 | participant) + (1 | item))) did not reveal any significant differences. Find the model output in Table 2.[7]

Behavioral measures

In the following section, we report the results from our three behavioral measures: attention (attention span and attention level), eye contact, and hand activity. Since this study has been designed and conducted with a focus on signing children, we provide results for both groups but elaborate more on the signing children; we also provide more qualitative insights into the data from signing children as

[7]Note that any statistical analysis should be interpreted with caution, given the large variation and small size of our sample. Inferential statistics are generally used to extrapolate from samples to the population. While we statistically treat the signing group as a sample in our analyses, it actually encompasses the entire population of relevant participants, in which case the means represent true population means for this task.

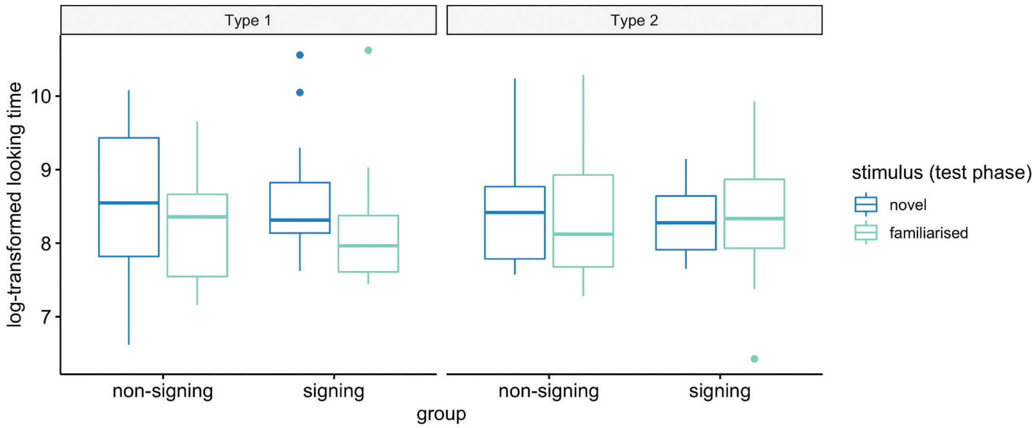

Figure 6. No effect of looking time (log-transformed). The graph shows the different blocks in two panels (type 1 and type 2) and plots the looking time to the novel and the familiarised stimulus in the test phase by group.

Table 2. Model output table.

	Estimate	Std. Error	t value
(Intercept)	8.232426699	0.156634361	52.55824224
groupsigning	−0.029277351	0.164746273	−0.177711768
phase_stimtest_novel	0.281506256	0.157909587	1.782705288
blockType2	−0.053721252	0.145998966	−0.367956384

there are important differences between hearing and deaf participants such as the fact that only hearing children have access to auditory input.

Attention

Our attention measure yields two main observations: signing children have (i) a longer attention span (number of completed trials), and (ii) a higher attention level (rating distraction from the task).

The attention span of signing children is longer than that of non-signing children, i.e., signing children completed more trials than non-signing children (Figure 7). With two exceptions (CST: 0;4; hearing; CSW: 1;6; hearing), all signing children completed all eight trials. CST, a four-month-old infant, completed six trials, and for CSW, an 18-month-old hearing child, we had to interrupt the final trial at the final stimulus due to fussiness. Fewer non-signing children completed all eight trials; only three completed eight (CNSAII: 1;7, CNSC: 1;6, CNSM: 2;7), one child (CNSA: 4;1) completed seven, four children (CNSD: 1;10, CNSS: 1;7, CNST: 0;4, CNSW: 1;5) completed six, and one child (CNSSII: 1;7) completed four trials. In short, the signing children upheld their interest in the task longer than the non-signing children.

The signing children were more attentive than the non-signing children, i.e., signing children were less distracted from the task than non-signing children (Figure 7). Although most children (except CNSS: 1;7; hearing; non-signing) got more distracted over time, the level of distraction of signing children was generally rated lower (mean = 2.03; range = 1–4; SD = 0.89) than of non-signing children (mean = 2.54; range = 1–5; SD = 1.27). The two deaf children had the highest attention levels: CSA (4;1; deaf; signing) was extremely focused during all eight trials and CSC (1;6; deaf; signing) started to show some signs of mild distraction half-way through the experiment.

In sum, the two groups of children differed in both attention measures. Signing children were more attentive to the task and kept attention levels higher for longer; non-signing children became distracted more easily and more quickly. This observation seems in line with the prediction that signing children recognize the stimuli as linguistic input that catches and holds their attention, while non-signing children do not and are therefore fatigued more rapidly than signing children.

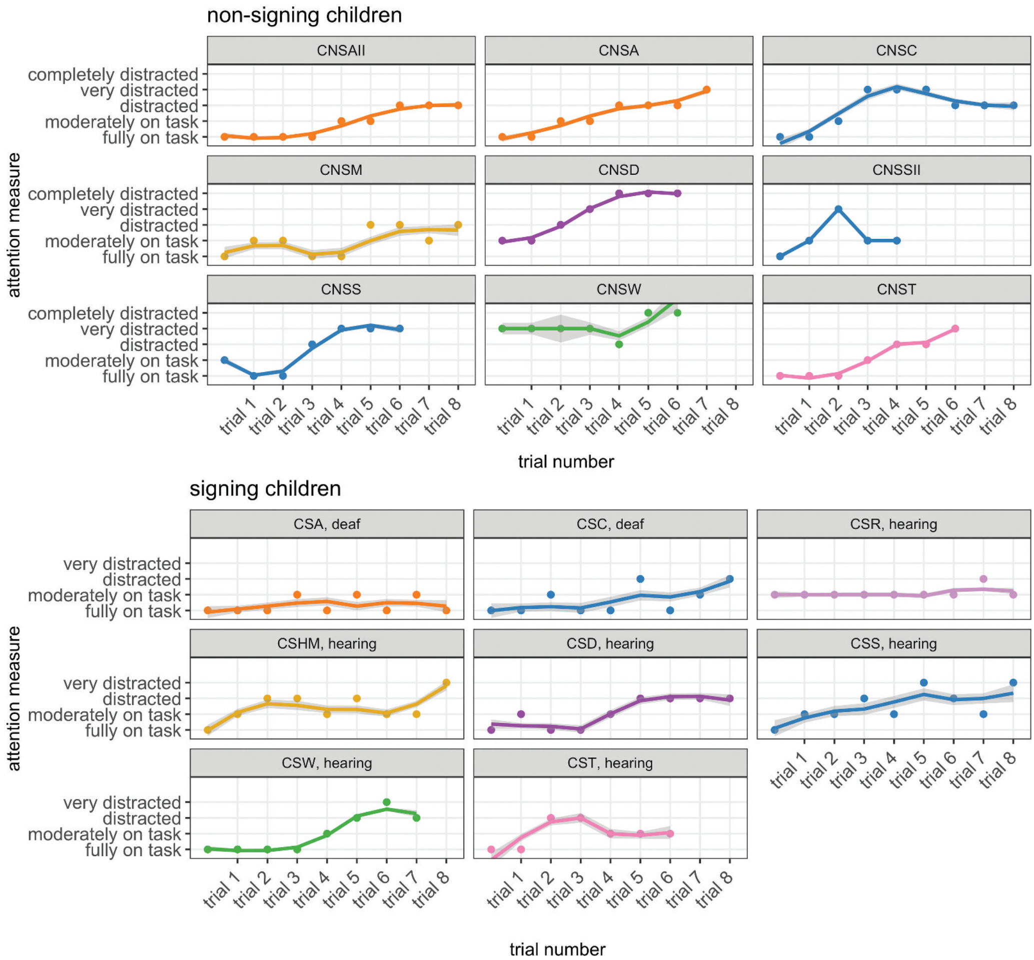

Figure 7. Attention measure per trial across the full experiment (eight trials). Plots are ordered by deafness and age in descending order, starting with the oldest deaf participant.

Alternatively, it is also possible that signing children are simply more interested in the signed stimuli given their familiarity with signed input.

Eye contact

Eye contact was coded as a categorical variable with three levels (no eye contact, eye contact with experimenter and eye contact with caregiver). Here, we report the rate and timing of eye contact. Experimenter and caregiver were collapsed to "interlocutor."

The signing children established eye contact with an interlocutor more often than the non-signing children (Figure 8a). The signing children looked to an interlocutor in 37.1% (143/385) of the cases, whereas the non-signing children did so only in 18.3% (65/355). Nevertheless, individual variation is high: two non-signing children (CNSSII: 1;7; hearing; non-signing; CNSD: 1;10; hearing; non-signing) established eye contact more often than other non-signing participants and two signing children (CSHM: 2;8; hearing, signing; and CSD: 1;11; hearing; signing) did so less often than the other signing children who are hearing. Note that the high rate of eye contact among deaf signing children is driven by CSA, who has the highest rate of eye contact overall (57.8%; 67/116; Figure 8b).

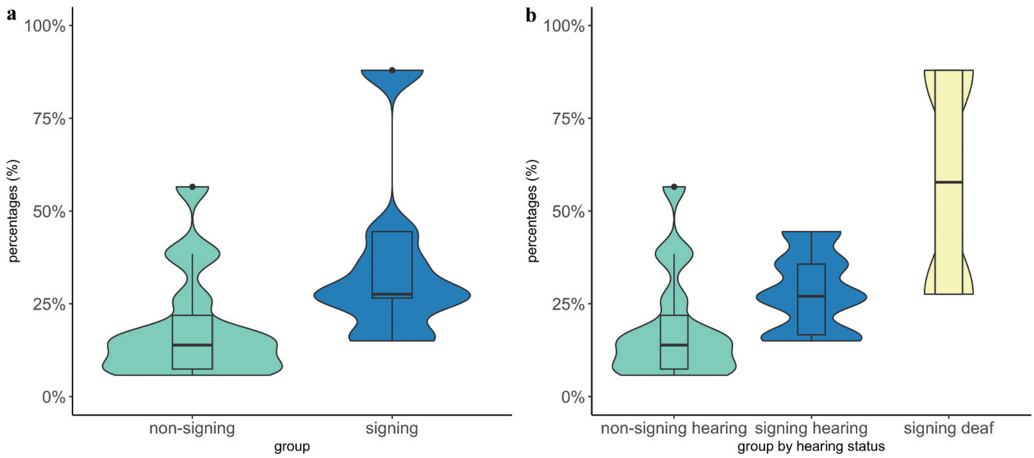

Figure 8. Prominence of eye contact with an interlocutor per group.

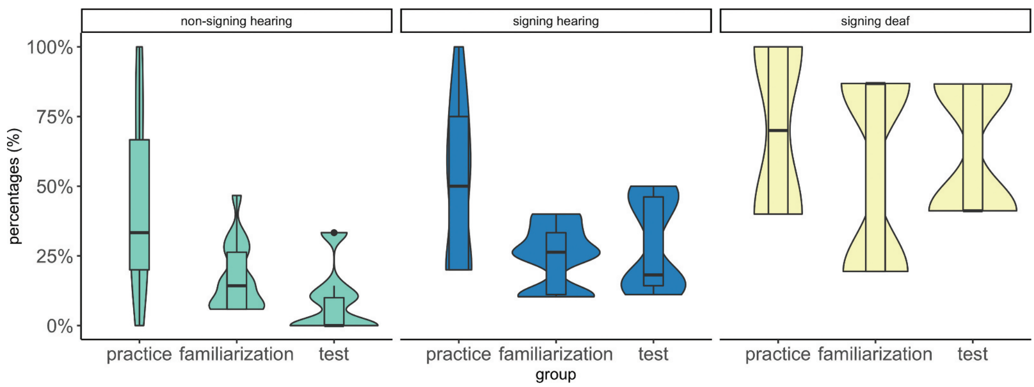

Figure 9. Eye contact by phase and group.

Eye contact seems to be driven by phase (practice vs. familiarization vs. test; Figure 9). Generally, children most often establish eye contact in the practice phase, i.e., at the beginning of the experiment. Nevertheless, patterns of eye contact among the non-signing children suggest natural decreases in attention over time as they show less eye contact during test phases than during familiarization phases. The signing children establish eye contact more frequently in the test phase than in the familiarization phase. This difference in behavior might indicate that signing children may detect differences between stimuli.

In sum, eye contact is more common among the signing children, and especially among the two deaf children. Given that eye contact is fundamental to signed communication, establishing eye contact with an interlocutor among signing children might indicate the children's readiness to communicate, possibly as a result of recognizing differences between stimuli or about their linguistic content.

Hand activity

Hand activity was scored on a mixed scale (scale 1–5 and categorical). Here, we report only communicative (non-manual) signs or gestures.

Groups differed in the type of hand engagement: only non-signing children experienced parental manipulation, and signing children produced some signs/gestures. Overall, there was communicative

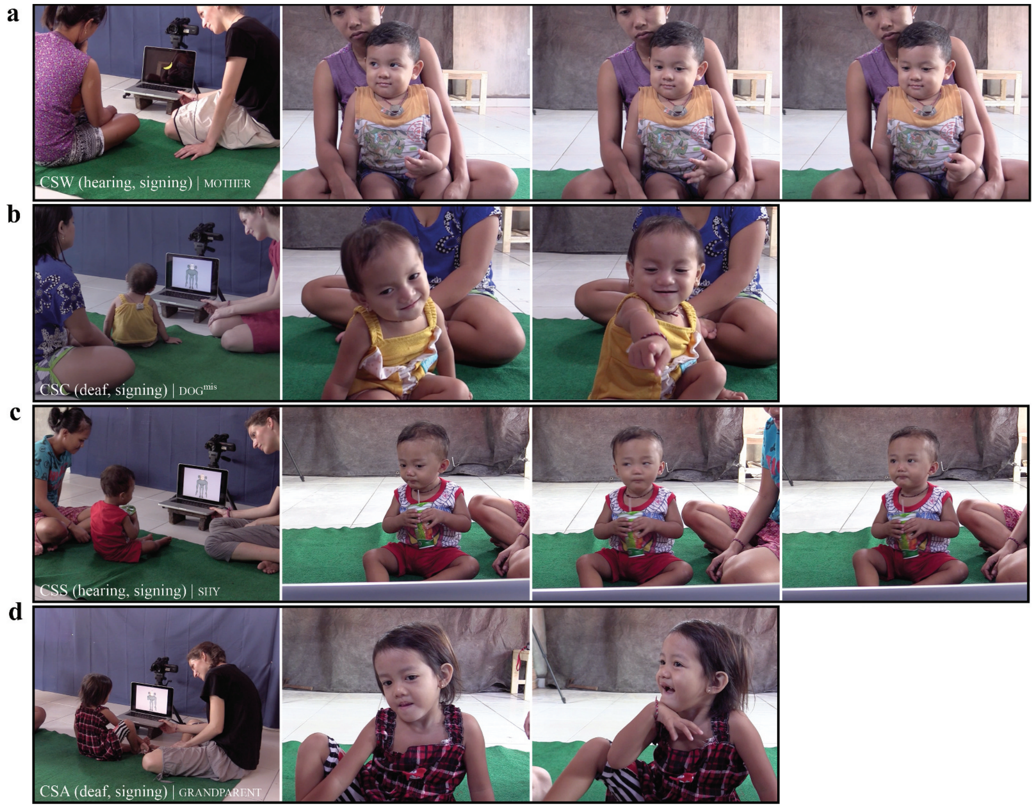

Figure 10. Examples of responses of different signing children.

(non-manual) activity in 9% of the data ($N = 67/740$; non-signing: 7%, $N = 25/355$; signing: 10.9%, $N = 42/385$). One-third of the non-signing children and none of the signing children experienced parental manipulations; although instructed not to interfere, caregivers reached for their children's arms and hands to move them playfully or to imitate the stimulus. Moreover, 4/9 non-signing children and 5/8 signing children used signs/gestures during or immediately following a stimulus presentation. For the non-signing children, gestures occur most often during the practice phase (9.3%, $N = 4/43$) and rarely during the test phase (3.8%, $N = 3/80$), suggesting that hand activity might be a proxy for the child's decreasing interest. That is, they are engaged at first but as the familiarization phase continues, interest faded. For the signing children, hand engagement never occured during the practice phase and occured most often during the test phase (17.5%; $N = 17/97$). This may suggest an increase of interest in the test phase which could be related to identifying differences in the presented stimuli. One further, interesting observation is that both deaf children but only three out of the six hearing signing children produced signs/gestures.

Furthermore, the type of hand activity differed across groups. While pointing to the screen was most common among non-signing children, signing children produced a range of different responses including child productions such as incomplete imitations (Figure 10a), pointing (Figure 10b), direct responses (Figure 10c), and full repetitions (Figure 10d). We elaborate on each of these examples below.

Incomplete imitations of the stimulus are cases where the child repeats (some) of the target sign, possibly diverging from the adult target due to characteristics of child signing (Lutzenberger et al., 2023). For example, upon seeing the stimulus MOTHER, produced by moving the flat handshape repeatedly up and down with constant contact at the ipsilateral chest, CSW (1;6; hearing, signing) produced a sign that resembles the form of the target sign (Figure 10a): CSW repeatedly flexed his wrist and fingers, while the hand rests on the lap. This could be a child production of MOTHER.

Pointing often occured directed toward the screen and may be produced alongside other (non-manual) cues. For example, during the third showing of DOG^{mis} in the familiarization phase, CSC (1;5; deaf, signing) first tilted her head, then smiled, and finally pointed to the screen (Figure 10b). With this, CSC showed clear modality-appropriate communicative behavior of establishing joint attention by incrementally increasing her behavior from two subsequent non-manual cues to also adding a point to the screen.

The communicative intent becomes even more clear in direct responses, such as in the example in Figure 10c. CSS (1;6; hearing, signing) has been familiarized with SHY^{mis} and upon encountering the target sign SHY, he promptly responded to the stimulus by shaking his head as he would in a conversation with a human signer; he responded negatively to the sign SHY produced by the animated monkey while keeping eye contact with the screen rather than establishing eye contact with a human interlocutor in the room. This suggests that he treated the animated monkey as a conversation partner and interpreted the stimulus as a direct question in a conversation, namely whether he feels shy.

Lastly, the clearest cases of communicative behavior are full repetitions as observed in most of CSA's responses (4;1; deaf, signing). Throughout the entire experiment, CSA repeated most stimuli, regardless of whether it was a target sign or a manipulated sign. For example, she observed the stimulus for one (or few) sign cycles, established eye contact with the experimenter, smiled, and then repeated the sign GRANDPARENT (Figure 10d). This suggests that she recognized the stimuli as signs and can easily reproduce them. Clearly, CSA showed complex communicative behavior, combining apt perceptual and productive skills.

Overall, while there is no measurable group-difference between signing and non-signing children in looking times during the test phase, they exhibit different hand activities. For non-signing children, hand behaviors are limited to few instances of pointing and the timing suggests that they are incidental. For signing children, hand behavior is diverse and often holds communicative intent, suggesting that they interact with the experiment on a more linguistic level, at least to a certain extent.

General discussion and conclusion

This paper contributes to the ongoing discussion about how to most effectively and inclusively achieve comparable research on child language development in urban and rural contexts. We present a case study from KK to illustrate a dual adaptation of an established paradigm and to critically evaluate the gain of such a study, with respect to both methodological and theoretical impact. Using visual familiarization, we present signing children and age-matched non-signing peers with animated sign stimuli that consisted of a) KK target signs and b) modifications that are minimally different from the target, distinguished by a handshape manipulation. Our data did not yield any group-differences in looking time to minimally different stimuli,[8] but additional behavioral measures (attention, eye contact, hand activity) provided more fruitful insights into the children's communicative behavior responding to the stimuli. Nevertheless, the high variability certainly also has to do with different personalities, ages, possibly signing skills as well as the quantity and quality of signing input, which we do not measure here. In the following section, we (i) discuss various adaptations to design, procedure, and analysis, and how they might have affected the results, (ii) evaluate the suitability of the paradigm for testing signed stimuli with signing participants, due to the critical role of eye gaze both in the paradigm **and** in signed communication, (iii) propose an alternative design and measure, and (iv) critically reflect on the contribution of such small-scale experimental studies with rural population.

[8]We would like to acknowledge the possibility that this result may be influenced by the nature of the manipulations, primarily the modifications we used might have been obvious to all participants, regardless of their hearing or signing status. In order to rule out that the stimuli are the main contributor of this result, a separate test of all stimuli with adult and child non-signers would be necessary to evaluate whether manipulations can be detected visually without any signing experience.

Note that these points of discussion are all specifically linked to the phenomenon studied in this paper and the adaptations made within; it is worth noting that the conclusions drawn in this paper are unlikely to extend to the acquisition of sign phonology more generally, nor potentially to other experimental methods for measuring early phonological discrimination. Nevertheless, this case study reveals that gaze-based measures of implicit language knowledge have to be handled with care and that non-gaze-based alternatives may be more suitable when it comes to signing and deaf participant groups.

The dual adaptations challenge

In the following section, we address implications of the dual adaptation in these studies, (i) modality adaptations: adapting a technique from spoken language to signing participants, in particular concerning the stimuli and experimental setup, and (ii) field adaptations: adapting a technique commonly practiced with urban participants to participants from an understudied rural community, specifically concerning the experimental design, procedure, and analysis. Here, we reflect and discuss how adaptations to (i) the sample, (ii) the stimuli, and (iii) the analysis may have affected the results.

Compared to traditional studies of early speech perception, our sample is small and heterogeneous. As the implications of the small size are discussed elsewhere (Lutzenberger, 2023), we focus on the heterogeneity of the sample and how it might affect the results. Our single inclusion criterion for signing children was that children acquired KK as a first language. The results show that the age range is large (0;4–4;0) – for infant speech perception paradigms, the participants tested are typically between 0;0 and 1;0 (Fennell et al., 2017; Werker & Fennell, 2009). While traditionally, participants with the same hearing status would be selected to avoid confounds, e.g., with bimodal bilingualism, given the demographic reality, we included both bimodal bilinguals (hearing signing) and monolinguals (deaf signing), and their language input may vary alongside their family members' hearing status and signing fluency (Place & Hoff, 2016; Unsworth et al., 2019). Furthermore, unlike in larger populations where researchers can be reasonably sure that non-signing children have no exposure to a sign language, we cannot exclude the possibility that our non-signing children have no familiarity with signing at all as they are sampled from the same village community where KK is culturally entrenched. Each of these sampling conditions may affect the results of this study individually or in combination. For example, regarding age, within the signing group, the 4-month-old infant did not produce any signs or establish eye contact with an interlocutor, while the 4-year-old did both consistently. While the behavioral differences between the two participants certainly reflect the age difference and how developed their communicative behavior is, there is another added layer by the fact that the older child is deaf and the younger one hearing, which may also help account for the striking differences in their behavior. These sampling conditions reveal clearly that the key tenets of experimental approaches – creating control groups, isolating experimental factors, and so on - are bent to the point of breaking in such small communities. With small numbers, it is impossible to say with confidence that experimental results are due to specific participant factors, when participants with many confounding factors are included in the same group.

Signs (much like speech) unfold sequentially (Figure 2). As a result, the informativeness of a look varies alongside the timing rather than the duration of a look; looks during the preparatory phase are less informative than looks that coincide with the stroke. We presented children with looping videos of isolated signs where all three phases were equally long and excluded looks that covered less than half the stroke (21%; $N = 210/989$) to ensure that children had the chance to identify the sign based on the peak informativeness. Nevertheless, evidence from adult signers of German Sign Language suggests that signers make predictions about the sign already before seeing the stroke (Hosemann et al., 2013), and similarly, the word recognition of young speaking children occurs already 200-400 ms after word onset (Swingley et al., 1999). This would suggest that signing children might not need to see the stroke in order to identify the sign and that, therefore, our coding might have missed particularly short looks (of signing children). This presents a major challenge for looking-based paradigms, particularly one that hinges on determining the difference between interest and disinterest based on looking time. To fully understand any potential effect, more research is needed into the prediction, recognition, and processing of signs by young signing children. While one could consider measuring looking time more generously to retain more datapoints, it is unclear how to differentiate between short looks as being due to early recognition, disinterest, or communicative behavior.

Another major adaptation of this study lies in the analysis. Traditionally, looking time is analyzed to test group differences. In this study, we provide a traditional analysis of looking times using a linear mixed-effects model. However, as stated in the results section, it seemed likely that our primary predictors would not reach statistical significance because of the characteristics of the sample explained above. Thus, the traditional way of analyzing data of this kind is not ideal for our data due to the sample size and its high variability. Instead of abandoning the analysis altogether, we extended the analysis with three behavioral measures that may reflect the quality of the data more appropriately. The current state of language acquisition research hinges heavily on inferential statistics and isolating specific response variables. Carving out space for supplementing such analyses with behavioral observations would be a challenge with larger sample sizes, however it would undoubtedly give readers a better understanding of data quality and advance the field.

Eye gaze – its role in visual familiarization and in signed communication

Eye gaze (or visual attention), the key operational measure in the familiarization paradigm, lies at the heart of signed communication. The familiarization paradigm builds on sustained eye gaze; we measure how long a child continues to look to a visual stimulus and interpret differences in looking behavior as evidence that they discriminate between stimuli. Eye gaze is essential for (non-tactile) signed communication, which builds on swift shifting of eye gaze and the coordination of eye contact between interlocutors to achieve joint attention (Lieberman et al., 2014). This creates a serious source of methodological conflict for gaze-based habituation measures of sign language perception: sustained eye gaze to the screen may be interrupted for the purpose of checking in with an interlocutor, for recruiting attention or for starting communication, all of which is qualitatively different from interrupting eye gaze to the screen because of stimulus habituation.

While this paradigm was developed for use with hearing speaking children who have extensive experience perceiving language without sustained eye gaze, it raises the question as to whether this paradigm is at all appropriate for signing children, who require eye contact to communicate. With signed stimuli, the measure and the mode of presentation are compressed into only the visual modality; signed stimuli require participants to look to the screen for sheer exposure, and, at the same time, looking time is measured, expecting longer looking times for both, higher interest, and discrimination. Indeed, this may suggest differences between signing children who are deaf or hearing: while their hearing peers have experience receiving language without eye gaze, their deaf peers do not. Thus, this raises the question as to whether their language experience drives children to the same behavior during the experiment, and in essence, whether short looks in (deaf) signing children actually reflect disinterest.

In this study, we have found establishing eye contact to be a particularly typical behavior for deaf signing children. Brooks et al. (2020, p. 5) report a similar result: the older participants in their study tend to shift visual attention from the stimulus to the experimenter, which they interpret as modality-typical checking-in behavior. Instead of marking decreasing interest, shorter looks among the signing participants may thus be caused by switching visual attention. This is further supported by findings from other research showing that deaf signing children achieve high eye gaze control early in life (Lieberman et al., 2014), exhibit adult-like gaze patterns already at age 5–14 months (Bosworth & Stone, 2021), and demonstrate enhanced gaze following as early as 7–20 months (Brooks et al., 2020). Clearly, they master the ecological niche of signed communication early in life. Nevertheless, research on hearing children suggests cultural differences in visual attention patterns (Kardan et al., 2017). Given that the aforementioned studies were conducted with deaf participants from the U.S., it remains unclear to what extent visual attention patterns in deaf infants are influenced by cultural differences.

Given the number of adaptations made in this study, it is unclear whether an adaptation affected the measure and whether there is a more general effect caused by the paradigm. Nevertheless, the role of eye gaze in signed communication may invalidate the measure and the interpretation of results; if we cannot rely on the fact that a short look is actually a short look because of boredom/familiarity with the stimulus but possibly instead because of switching visual attention to recruit the interlocutor's attention, the measure is not reliable. This suggests that paradigms that rely on sustained eye gaze may not be highly suitable for signed child participants. Instead, a preferential looking paradigm, where eye gaze is indicative of a choice, i.e., location of look rather than duration is measured, may be a more modality-neutral choice. Nevertheless, this conflict with eye gaze is not limited to signing participants, and it could be beneficial for researchers to use paradigms that rely on looking time as a measure, to closely examine the measure itself and to possibly add coding of communicative attempts during the experiment to gain deeper insights into how children respond to the task.

Alternative designs: Using touch as a more direct measure of interest

Another avenue that may be better suited to the ecological niche of signed communication might be using a more direct measure of interest, e.g., touch. Here, we report here an alternative design to collecting looking time equivalents with a tablet-based task that provides many attractive opportunities for the future. The reported design has only undergone initial piloting in the KK community and needs more extensive testing in the future for more reliable conclusions about its suitability and reliability.

Aiming to circumvent some problems described above, we designed a tablet-based experiment that registers touch input to measure interest (i.e., replacing screen-look duration with screen-touch duration). The reason this attention measure is more direct is that children are required to actively touch the screen rather than just passively sit and watch. This experimental design is modeled on the core ideas of the habituation paradigm but using a tablet (in our case, a Samsung SM-T713). Touch input (and, if needed, touch area) initiates playback of a stimulus for the touch duration. That is, children regulate the screen of a tablet between the default screen (e.g., the picture of a curtain or an attention getter) and the stimulus through touch; as long as the screen is being touched, the stimulus video plays (Figure 11).[9] As a pilot, we tested the functionality of the application with 15 adults and 11 child signers of KK using six pairs of stimuli, each comprising a target sign and an off-target sign with manipulation. The application was written by Peter Withers (Max Planck Institute for Psycholinguistics, Nijmegen) as an Android application available through Google Play Store from where age-appropriate games can be used as filler activities. Data is saved automatically on the memory card and can be uploaded later to a server using internet connection. Unfortunately, the feature used to save and asynchronously upload touch data did not function reliably, which is why the collected data were not sufficient for any analysis. In addition, the sessions were videotaped using a Canon Legria HF G26 camera. While the pilot

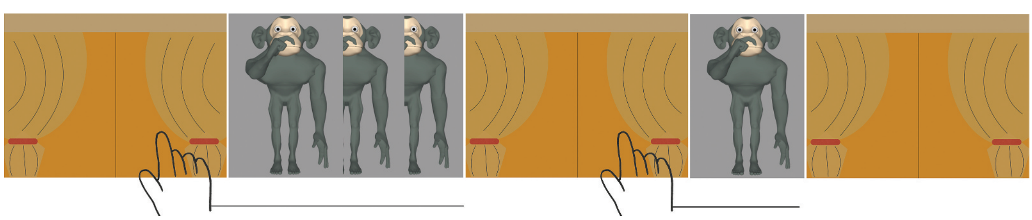

Figure 11. Example of tablet-based interaction.

[9]One possibility is that sustained touch might be counterintuitive as most tablet applications relay on short touches. We would like to stress that the children in our sample are not highly familiar with tablets as they are not commonly owned by people living in the village. In addition, the piloting suggests that, during the practice phase, children quickly attune to the fact that they see the stimulus only for as long as they keep on touching the screen.

version relied on controlling stimuli manually using a magicsee R1 Bluetooth remote controller, we planned to automatically calculate habituation based on a habituation criterion of 50% with a sliding window of three trials in the final dataset. Experimental sessions were held at the participants' homes.

The tablet-experiment holds several advantages to the method presented in this study: First, portable devices such as tablets expedite field-experiment through bringing them to the participant's homes (Frost & Casillas, 2021) and because even young children (Lytle et al., 2018) and participants with minimal prior experience with computing technologies (Frost & Casillas, 2021) use them intuitively. Second, hosting experimental infrastructure on Google Play Store encourages open science by making it available to other researchers, and facilitates the rapid development and spread of new versions through immediate online availability – this is helpful for field workers who need to adapt the app on the fly, piloting, and then testing in the same short field trip (e.g., Cristia et al., 2020; Frost & Casillas, 2021). A third advantage of touch input is that children generally use their dominant hand for navigating the tablet and for signing; thus, at least younger children, release their hand from the tablet when signing.

Nevertheless, continued touch might not always guarantee continuous exposure to the stimulus as children may cover the stimulus with their hand or look away while touching, e.g., to establish eye contact with an interlocutor. From the experiences of our pilot, we recommend that in order to optimize touch input data, tablets should have maximum screen size and be affixed to the floor in a 45-degree angle, preventing children from leaning on it or picking it up. Furthermore, the attention getter should not be placed in the center of the screen on a tablet since it encourages children to place their hand where the stimulus is about to appear. Despite its many advantages, it remains to be empirically tested whether touch indeed provides a reliable alternative measure to looking time or other, more traditional methods of testing discrimination. As explained above, touch-input does not circumvent all issues with sustained visual attention, one core principle of the habituation paradigm. Nevertheless, by providing a more direct measure, it opens up interesting avenues that warrant further exploration.

Where to go from here: Are experimental studies in rural communities a fruitful endeavor?

Sign language acquisition in rural settings is severely understudied (exceptions include de Vos, 2012a; Hou, 2016; Lutzenberger et al., 2023) and the use of experimental approaches in this context is unprecedented. In this paper, we have presented a study that implements a variety of methodological innovations, including changes to the experimental design, procedure, analysis, and sampling. In this final part of the discussion, we evaluate the question whether (or how) valuable these efforts are based on our study and the various points of discussion here above.

Adaptations or innovation. Experimental methods hinge on tight experimental control and subtle manipulations which may be compromised by the adaptations that are necessary when conducting them with (particularly small) signing and rural communities (summarized in Cristia et al., 2023 for rural, speaking communities). For example, using minimal pairs of full signs as signed stimuli rather than auditory stimuli of minimally different syllable pairs introduces the condition of a longer minimal time window in order to guarantee exposure to the stimulus. This raises the question whether continued adaptations of existing methods or rather innovations of new methods are more conducive to advancing and diversifying the field.

Cross-linguistic comparability. It is unclear how various adaptations and conducting experiments with different communities impact cross-linguistic comparability. Naturally, different adaptations are needed for different communities, including straightforward ones like using stimuli with phonological contrasts of the target language or more fundamental ones like changes to the experimental procedure or data analysis. Moreover, different communities might respond differently to certain tasks (e.g., Cristia et al., 2020; Kardan et al., 2017). On the one hand, community-appropriate methods make results more reliable; for example, Cristia's et al. (2020) decision of conducting a non-word repetition task in a group setting rather than with individual Tsimane' speakers. On the other hand, these differences may threaten study validity and limit how results can be compared across studies. Thus, results from different studies may have to be compared carefully especially given the wealth of different adaptations that traditional paradigms undergo when conducted with different rural communities.

Fundamental biases. Efforts to adapt paradigms across cultures expose fundamental biases about how children are socialized that are embedded in our experimental design and analyses. Working with child signers from a rural community taps into multiple of these biases. For example, most studies are designed for participants who are hearing and acquiring spoken languages, and thus use auditory stimuli. As there is a large number of hearing children available to most spoken language researchers, it is relatively uncomplicated to reach a homogenous sample of hearing children with similar characteristics. In research with deaf children, sampling typically becomes more heterogeneous simply because the population is smaller, for example with wide variation in signing input. Requiring the same experimental conditions may favor urban communities and may push communities with more heterogeneous characteristics to the outskirts of experimental inquiry (Cristia et al., 2023). Not accepting methods that are well-suited to specific samples, e.g. rural communities or deaf, signing populations, or orienting to them as alternatives or nonstandard ways to study linguistic phenomena create exclusionary research. Holding research up to high ethical standards, these biases need to be addressed.

The question remains: Where to go from here? Are experimental studies in rural communities a fruitful endeavor? Our answer is yes, as long as the design of such studies is well adapted to the participant population being studied. We cannot expect to use a narrow range of design types (originating from hearing and urban populations) to study deaf children in rural communities. More broadly, experimental and analytical designs must be flexible enough to incorporate a variety of language communities, or else we will find ourselves limited to learning and drawing conclusions from a small, homogenous group of languages and cultures (Jaeger & Norcliffe, 2009; Kidd & Garcia, 2022; Singh, 2022; Singh et al., 2022). This allows for fundamental biases to manifest in psycholinguistic theories. Broadening the typological landscape of languages that are studied experimentally is the only way to test and corroborate our theories about how children acquire language and to better represent the world's linguistic diversity. This effort cannot be carried out by researchers working with small and rural communities alone; what is needed, in our opinion, is a joint effort in which researchers working with over-represented populations and using high-tech experimental set-ups start to include alternative measures that can parallel rural adaptations (such as the behavioral measures in this study) or even seek collaborations to innovate methods that are optimally suited to specific rural communities and then adapt them to urban communities.

Acknowledgments

We would like to express our gratitude to all participants and their parents for their joyful participation in this research. Further, we thank our research assistants Ni Made Sumarni and Ketut Kanta for their support and collaboration. We also thank Susanne Brouwer for a great discussion on the considerations about analysis options in this paper and Rehana Omardeen for helpful comments on previous drafts. Lastly, we are very grateful to the anonymous reviewers and the editors for their helpful suggestions and feedback that greatly improved the paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was part of the doctoral project of Hannah Lutzenberger at Radboud University Nijmegen, supported by FWO-NWO under Grant [NWO 326-70-002] "The emergence of phonology within six generations" awarded to Bart de Boer, Paula Fikkert, and Connie de Vos. The research was further supported by NWO under Grant [VICI 277-70-014] "Deaf communication without a shared language" awarded to Onno Crasborn and under Grant [VENI 275-89-028] "The face in sign language interaction" awarded to Connie de Vos, as well as by the European Research Council under

ERC Starting Grant Emergence of Language in Social Interaction (ELISA - 852352) awarded to Connie de Vos and ERC Advanced Grant (SignMorph - 885220) awarded to Adam Schembri.

ORCID

Hannah Lutzenberger http://orcid.org/0000-0003-0574-4060
Marisa Casillas http://orcid.org/0000-0001-5417-0505
Paula Fikkert http://orcid.org/0000-0001-7317-2100
Onno Crasborn http://orcid.org/0000-0003-4985-9097
Connie de Vos http://orcid.org/0000-0002-4800-4313

Data availability statement

All materials, source code, export queries, and scripts may be found in the Open Science Framework (https://osf.io/w62tm/).

References

Alday, P. M. (n.d.). Coding schemes for categorical variables. *Contrast Coding*. https://phillipalday.com/stats/coding.html

Baker, S., Golinkoff, R. M., & Petitto, L.-A. (2006). New insights into old puzzles from infants' categorical discrimination of soundless phonetic units. *Language Learning and Development*, *2*(3), 147–162. https://doi.org/10.1207/s15473341lld0203_1

Baker, S., Idsardi, W. J., Golinkoff, R. M., & Petitto, L.-A. (2005). The perception of handshapes in American sign language. *Memory & Cognition*, *33*(5), 887–904. https://doi.org/10.3758/BF03193083

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

Best, C. T., Mathur, G., Miranda, K. A., & Lillo-Martin, D. (2010). Effects of sign language experience on categorical perception of dynamic ASL pseudosigns. *Attention, Perception, & Psychophysics*, *72*(3), 747–762. https://doi.org/10.3758/APP.72.3.747

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

Blau, S. (2019, September 26). *Perceptual narrowing in deaf infants* [Poster]. Theoretical Issues in Sign Language Research (TISLR13), Hamburg, Germany.

Bonvillian, J., & Siedlecki, T. (1996). Young children's acquisition of the location aspect of American sign language signs: Parental report findings. *Journal of Communication Disorders*, *29*(1), 13–35. https://doi.org/10.1016/0021-9924(94)00015-8

Bosworth, R. G., & Stone, A. (2021). Rapid development of perceptual gaze control in hearing native signing infants and children. *Developmental Science*, *24*(4), e13086. https://doi.org/10.1111/desc.13086

Brentari, D. (1998). *A prosodic model of sign language phonology*. MIT Press. https://doi.org/10.7551/mitpress/5644.001.0001

Brentari, D. (2019). *Sign language phonology* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781316286401

Brentari, D., Coppola, M., Cho, P. W., & Senghas, A. (2017). Handshape complexity as a precursor to phonology: Variation, emergence, and acquisition. *Language Acquisition*, *24*(4), 283–306. https://doi.org/10.1080/10489223.2016.1187614

Brooks, R., Singleton, J. L., & Meltzoff, A. N. (2020). Enhanced gaze-following behavior in deaf infants of deaf parents. *Developmental Science*, *23*(2), e12900. https://doi.org/10.1111/desc.12900

Carroll, J. J., & Gibson, E. J. (1986). Infant perception of gestural contrasts: Prerequisites for the acquisition of a visually specified language. *Journal of Child Language*, *13*(1), 31–49. https://doi.org/10.1017/S0305000900000283

Casillas, M., Brown, P., & Levinson, S. C. (2020a). Early language experience in a Papuan community. *Journal of Child Language*, 1–23. https://doi.org/10.1017/S0305000920000069

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

Cheek, A., Cormier, K., Repp, A., & Meier, R. P. (2001). Prelinguistic gesture predicts mastery and error in the production of early signs. *Language*, *77*(2), 292–323. https://doi.org/10.1353/lan.2001.0072

Conlin, K. E., Mirus, G. R., Mauk, C., & Meier, R. P. (2000). The acquisition of first signs: Place, handshape, and movement. In C. Chamberlain, J. P. Morford, & R. Mayberry (Eds.), *Language acquisition by eye* (pp. 51–69). Lawrence Erlbaum Associates.

Coppola, M. (2002). *The emergence of grammatical categories in home sign: Evidence from family-based gesture systems in Nicaragua* [PhD Thesis, University of Rochester]. https://production.wordpress.uconn.edu/languagecreationlab/wp-content/uploads/sites/644/2014/03/coppola_diss_2002.pdf

Covarrubias, M. (1937). *Island of Bali*. Cassell and Company Limited.

Cristia, A., & Casillas, M. (2022). Non-word repetition in children learning Yélî Dnye. *Language Development Research*, *2*(1), Article 1. https://doi.org/10.34842/zr2q-1x28

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., Farabolini, G., Scaff, C., Havron, N., Stieglitz, J., & Mulak, K. E. (2020). Infant-directed input and literacy effects on phonological processing: Non-word repetition scores among the tsimane'. *PLOS ONE*, *15*(9), e0237702. https://doi.org/10.1371/journal.pone.0237702

Cristia, A., Foushee, R., Aravena-Bravo, P., Cychosz, M., Scaff, C., & Casillas, M. (2023). Combining observational and experimental approaches to the development of language and communication in rural samples: Opportunities and challenges. *Journal of Child Language*, *50*(3), 495–517. https://doi.org/10.1017/S0305000922000617

de Vos, C. (2012a). The kata kolok perfective in child signing: Coordination of manual and non-manual components. In U. Zeshan & C. de Vos (Eds.), *Sign languages in village communities*. De Gruyter. https://doi.org/10.1515/9781614511496.127

de Vos, C. (2012b). *Sign-spatiality in Kata Kolok: How a village sign language of Bali inscribes its signing space* [PhD Thesis, Radboud University]. https://doi.org/10.1075/sll.16.2.08vos

de Vos, C., & Pfau, R. (2015). Sign language typology: The contribution of rural sign languages. *Annual Review of Linguistics*, *1*(1), 265–288. https://doi.org/10.1146/annurev-linguist-030514-124958

Eccarius, P., & Brentari, D. (2010). A formal analysis of phonological contrast and iconicity in sign language handshapes. *Sign Language & Linguistics*, *13*(2), 156–181. https://doi.org/10.1075/sll.13.2.02ecc

Edward, M. (2022). Child language acquisition research on indigenous African sign languages: A commentary on Kidd and Garcia (2022). *First Language*, *42*(6), 756–759. https://doi.org/10.1177/01427237221104049

ELAN [Computer software] (5.9). (2020). [Computer Software]. Max Planck Institute for Psycholinguistics, The Language Archive. https://archive.mpi.nl/tla/elan

Emmorey, K., McCullough, S., & Brentari, D. (2003). Categorical perception in American sign language. *Language and Cognitive Processes*, *18*(1), 21–45. https://doi.org/10.1080/01690960143000416

Fennell, C. T. (2017). Habituation techniques. In A. M. B., Groot & P.Hagoort (Eds.), *Research methods in psycholinguistics and the Neurobiology of Language: A practical guide* (p. 1–18). Wiley-Blackwell.

Friedman, T. B., Liang, Y., Weber, J. L., Hinnant, J. T., Barber, T. D., Winata, S., Arhya, I. N., & Asher, J. H. (1995). A gene for congenital, recessive deafness DFNB3 maps to the pericentromeric region of chromosome 17. *Nature Genetics*, *9*(1), 86–91. https://doi.org/10.1038/ng0195-86

Frost, R., & Casillas, M. (2021). Investigating statistical learning of nonadjacent dependencies: Running statistical learning tasks in non-WEIRD populations. In *SAGE methods case studies*. https://doi.org/10.4135/9781529759181

Gagne, D. L., & Coppola, M. (2017). Visible social interactions do not support the development of false belief understanding in the absence of linguistic input: Evidence from deaf adult homesigners. *Frontiers in Psychology*, *8*, 8. https://www.frontiersin.org/articles/10.3389/fpsyg.2017.00837

Gleitman, L., Senghas, A., Flaherty, M., Coppola, M., & Goldin-Meadow, S. (2019). The emergence of the formal category "symmetry" in a new sign language. *Proceedings of the National Academy of Sciences*, *116*(24), 11705–11711. https://doi.org/10.1073/pnas.1819872116

Golinkoff, R. M., Soderstrom, M., Can, D. D., & Hirsh, K. (2017). Visual preference techniques. In A. M. B. Groot & P. Hagoort (Eds.), *Research methods in psycholinguistics and the Neurobiology of Language: A practical guide* (pp. 18–39). Wiley Blackwell.

Horton, L. (2018). *Conventionalization of shared homesign systems in Guatemala: Social, lexical, and morphophonological dimensions* [PhD Thesis, University of Chicago]. https://doi.org/10.6082/uchicago.1321

Hosemann, J., Herrmann, A., Steinbach, M., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013). Lexical prediction via forward models: N400 evidence from German sign language. *Neuropsychologia*, *51*(11), 2224–2237. https://doi.org/10.1016/j.neuropsychologia.2013.07.013

Hou, L. (2016). *"Making hands": Family Sign Languages in the San Juan Quiahije Community* [UT Austin]. http://hdl.handle.net/2152/46879

Hunter, M. A. & Ames, E. W. (1988). *A multifactor model of infant preferences for novel and familiar stimuli in advances in infancy research* (Vol. 5) (pp. 69–95). Ablex Publishing.

Jaeger, T. F., & Norcliffe, E. J. (2009). The cross-linguistic study of sentence production. *Language and Linguistics Compass*, *3*(4), 866–887. https://doi.org/10.1111/j.1749-818X.2009.00147.x

Johnston, T., & Schembri, A. (1999). On defining lexeme in a signed language. *Sign Language & Linguistics*, *2*(2), 115–185. https://doi.org/10.1075/sll.2.2.03joh

Kardan, O., Shneidman, L., Krogh-Jespersen, S., Gaskins, S., Berman, M. G., & Woodward, A. (2017). Cultural and developmental influences on overt visual attention to videos. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/s41598-017-11570-w

Karnopp, L. B. (2008). Sign phonology acquisition in Brazilian sign Language. Sign languages: Spinning and unraveling the past, present and future. TISLR9, forty five papers and three posters from the 9th. Theoretical Issues in Sign Language Research Conference, Florianopolis, Brazil, December. Sign phonology acquisition in Brazilian sign Language. Sign languages: Spinning and unraveling the past, present and future. TISLR9, forty five papers and three posters from the 9th. M. de Quadros (Ed.). Theoretical Issues in Sign Language Research Conference. Brazil. http://www.editora-arara-azul.com.br/EstudosSurdos.php

Kidd, E., & Garcia, R. (2022). How diverse is child language acquisition research? *First Language*, 014272372110664. https://doi.org/10.1177/01427237211066405

Kita, S., van Gijn, I., & van der Hulst, H. (1998). Movement phases in signs and co-speech gestures, and their transcription by human coders. In I. Wachsmuth & M. Fröhlich (Eds.), *Gesture and sign language in human-computer interaction* (Vol. 1371, pp. 23–35). Springer. https://doi.org/10.1007/BFb0052986

Kooij, E. V. D. (2002). *Phonological categories in Sign Language of the Netherlands. The role of phonetic implementation and iconicity* [PhD Thesis].

Krentz, U. C., & Corina, D. P. (2008). Preference for language in early infancy: The human language bias is not speech specific. *Developmental Science*, *11*(1), 1–9. https://doi.org/10.1111/j.1467-7687.2007.00652.x

Lieberman, A. M., Hatrak, M., & Mayberry, R. I. (2014). Learning to look for language: Development of joint attention in young deaf children. *Language Learning and Development: The Official Journal of the Society for Language Development*, *10*(1), 19–35. https://doi.org/10.1080/15475441.2012.760381

Lieven, E., & Stoll, S. (2009). Language. In M. H. Borinstein (Ed.), *The handbook of cultural developmental science* (pp. 143–160). Psychology Press.

Lillo-Martin, D., & Henner, J. (2021). Acquisition of sign languages. *Annual Review of Linguistics*, *7*(1), 395–419. https://doi.org/10.1146/annurev-linguistics-043020-092357

Lutzenberger, H. (2018). Manual and nonmanual features of name signs in Kata Kolok and sign language of the Netherlands. *Sign Language Studies*, *18*(4), 546–569. https://doi.org/10.1353/sls.2018.0016

Lutzenberger, H. (2019, September): B1AS5 - a preliminary investigation of the phonological feature inventory in Kata Kolok, a rural sign language of Bali. Poster at 13th Theoretical Issues in Sign Language Research Conference (TISLR13), Hamburg University. Retrieved from: https://www.idgs.uni-hamburg.de/forschung/tagungen/tislr2019/programm/poster-lutzenberger.pdf

Lutzenberger, H. (2019). Threat or natural fluctuation? Revisiting language vitality of Kata Kolok, the sign language of a village in Bali. In *State of the art of indigenous languages in research: A collection of selected research papers* (p. 12). UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381556

Lutzenberger, H. (Director). (2020). *Kata Kolok Dataset*. https://signbank.science.ru.nl/datasets/Kata%20Kolok

Lutzenberger, H. (2022). *Kata Kolok phonology—Variation & acquisition* [PhD Thesis, Radboud University]. https://pure.mpg.de/rest/items/item_3370313/component/file_3373920/content

Lutzenberger, H. (2023). The acquisition of sign languages in rural contexts – what can we do when samples will always be 'too small'? *Journal of Child Language*, *50*(3), 527–531. https://doi.org/10.1017/S0305000922000757

Lutzenberger, H., de Vos, C., Kanta, K., Primantara, G. I., Astini, N. D. & Sumarni, N. M. (2007-2023). Item "Kata Kolok Child Signing Corpus (KKCSC)". Item "Kata Kolok Child Signing Corpus (KKCSC)". https://hdl.handle.net/1839/f47a1a94-2d09-4e19-b721-7e9547cc796c.

Lutzenberger, H., Fikkert, P., de Vos, C., & Crasborn, O. (2023). Development of sign phonology in Kata Kolok. *Journal of Child Language*, 1–34. https://doi.org/10.1017/S0305000922000745

Lutzenberger, H., Vos, C. D., Crasborn, O., & Fikkert, P. (2021). Formal variation in the kata kolok lexicon. *Glossa: A Journal of General Linguistics*, *6*(1), Article 1. https://doi.org/10.16995/glossa.5880

Lytle, S. R., Garcia-Sierra, A., & Kuhl, P. K. (2018). Two are better than one: Infant language learning from video improves in the presence of peers. *Proceedings of the National Academy of Sciences*, *115*(40), 9859–9866. https://doi.org/10.1073/pnas.1611621115

Mann, W., Marshall, C. R., Mason, K., & Morgan, G. (2010). The acquisition of sign language: The impact of phonetic complexity on phonology. *Language Learning and Development*, *6*(1), 60–86. https://doi.org/10.1080/15475440903245951

Marsaja, I. G. (2008). *Desa kolok: A deaf village and its sign language in Bali, Indonesia*. Ishara Press.

Meier, R. P. (2006). The form of early signs: Explaining signing children's articulatory development. In B. Schick, M. Marschark, & P. E. Spencer (Eds.), *Advances in the sign language development of deaf children* (pp. 202–230). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195180947.003.0009

Morgan, G. (2006). 'Children are just lingual': The development of phonology in British sign language (BSL). *Lingua*, *116*(10), 1507–1523. https://doi.org/10.1016/j.lingua.2005.07.010

Morgan, H. (2017). *The phonology of Kenyan Sign Language (southwestern dialect)* [PhD Thesis, UC San Diego]. https://escholarship.org/uc/item/9bp3h8t4

Mudd, K., Lutzenberger, H., de Vos, C., Fikkert, P., Crasborn, O., & de Boer, B. (2020). The effect of sociolinguistic factors on variation in the kata kolok lexicon. *Asia-Pacific Language Variation*, *6*(1), 53–88. https://doi.org/10.1075/aplv.19009.mud

Mulak, K. E., Sarvasy, H. S., Tuninetti, A., Escudero, P., & Schiller, N. O. (2021). Word learning in the field: Adapting a laboratory-based task for testing in remote Papua New Guinea. *PLOS ONE*, *16*(9), e0257393. https://doi.org/10.1371/journal.pone.0257393

Oakes, L. M. (2010). Using habituation of looking time to assess mental processes in infancy. *Journal of Cognition and Development: Official Journal of the Cognitive Development Society*, *11*(3), 255–268. https://doi.org/10.1080/15248371003699977

Palmer Baker, S., Fais, L., Golinkoff, R. M., & Werker, J. F. (2012). Perceptual narrowing of linguistic sign occurs in the 1st year of life. *Child Development*, *83*(2), 543–553. https://doi.org/10.1111/j.1467-8624.2011.01715.x

Place, S., & Hoff, E. (2016). Effects and noneffects of input in bilingual environments on dual language skills in 2 ½-year-olds*. *Bilingualism: Language and Cognition*, *19*(5), 1023–1041. https://doi.org/10.1017/S1366728915000322

Pye, C. (2021). Documenting the acquisition of indigenous languages. *Journal of Child Language*, *48*(3), 454–479. https://doi.org/10.1017/S0305000920000318

Pyers, J. E., & Senghas, A. (2009). Language promotes false-belief understanding: Evidence from learners of a new sign language. *Psychological Science*, *20*(7), 805–812. https://doi.org/10.1111/j.1467-9280.2009.02377.x

R Core Team. (2019). *R: A language and environment for statistical computing*. (1.0.143) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Sandler, W. (1989). *Phonological representation of the sign: Linearity and nonlinearity in American sign Language*. De Gruyter Mouton. https://doi.org/10.1515/9783110250473

Sandler, W. (2012). The phonological organization of sign languages. *Language and Linguistics Compass*, *6*(3), 162–182. https://doi.org/10.1002/lnc3.326

Sarvasy, H. S., Morgan, A. M., Yu, J., Ferreira, V. S., & Momma, S. (2023). Cross-clause planning in Nungon (Papua New Guinea): Eye-tracking evidence. *Memory & Cognition*, *51*(3), 666–680. https://doi.org/10.3758/s13421-021-01253-3

Sarvasy, H. S., Tuninetti, A., Kung, C., & Escudero, P. (2019). *EEG in PNG: A report on the neurolinguistics of clause chain processing*. International Cognitive Linguistics Conference, Nishinomiya, Japan.

Schley, S. (1991). Infant discrimination of gestural classes: Perceptual precursors of ASL acquisition. *Sign Language Studies*, *1072*(1), 277–296. https://doi.org/10.1353/sls.1991.0015

Senghas, A., & Coppola, M. (2001). Children creating language: How Nicaraguan sign language acquired a spatial grammar. *Psychological Science*, *12*(4), 323–328. https://doi.org/10.1111/1467-9280.00359

Siedlecki, T., & Bonvillian, J. D. (1993). Location, handshape & movement: Young children's acquisition of the formational aspects of American sign language. *Sign Language Studies*, *1078*(1), 31–52. https://doi.org/10.1353/sls.1993.0016

Singh, L. (2022). From information to action: A commentary on Kidd and Garcia (2022). *First Language*, *42*(6), 814–817. https://doi.org/10.1177/01427237221090024

Singh, L., Rajendra, S. J., & Mazuka, R. (2022). Diversity and representation in studies of infant perceptual narrowing. *Child Development Perspectives*, *16*(4), 191–199. https://doi.org/10.1111/cdep.12468

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*(6640), 381–382. Article 6640 https://doi.org/10.1038/41102

Stokoe, W. C. (1960). *Sign language structure: An outline of the visual communication systems of the American Deaf* (Vol. 8). Linstok Press.

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, *71*(2), 73–108. https://doi.org/10.1016/S0010-0277(99)00021-9

Unsworth, S., Brouwer, S., de Bree, E., & Verhagen, J. (2019). Predicting bilingual preschoolers' patterns of language development: Degree of non-native input matters. *Applied Psycholinguistics*, *40*(5), 1189–1219. https://doi.org/10.1017/S0142716419000225

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology*, *34*(6), 1289–1309. https://doi.org/10.1037/0012-1649.34.6.1289

Werker, J. F., & Fennell, C. T. (2009). Infant speech perception and later language acquisition: Methodological underpinnings. In J. Colombo, P. McCardle, & L. Freund (Eds.), *Infant pathways to language: Methods, models, and research directions* (pp. 85–98). Psychology Press.

Wilbourn, M. P., & Casasola, M. (2007). Discriminating signs: Perceptual precursors to acquiring a visual-gestural language. *Infant Behavior and Development*, *30*(1), 153–160. https://doi.org/10.1016/j.infbeh.2006.08.006

Winata, S., Arhya, I. N., Moeljopawiro, S., Hinnant, J. T., Liang, Y., Friedman, T. B., & Asher, J. H. (1995). Congenital non-syndromal autosomal recessive deafness in Bengkala, an isolated Balinese village. *Journal of Medical Genetics*, *32*(5), 336–343. https://doi.org/10.1136/jmg.32.5.336

Winter, B. (2019). *Statistics for linguists: An introduction using R*. Routledge.

Zeshan, U., & de Vos, C. (Eds.). (2012). *Sign languages in village communities: Anthropological and linguistic insights*. De Gruyter Mouton; Ishara Press.

Appendix A:

Detailed overview of stimuli

		stimulus			
block	sign gloss	target		mispronunciation	
practice	COW	L		5	
Type 1 (fingers extended or bent at the finger base joint)	FATHER	1_Curved		C_Spread	
	GRANDPARENT	5		1	
	MOTHER	5		1	
	PRAY	5		A	
Type 2 (fingers extended vs. curved)	BATHE	5		C_Spread	
	CRY	B		C_Spread	
	DOG	C_Spread		5	
	SHY	5		C_Spread