



# Emerging Verbal Functions in Early Infancy: Lessons from Observational and Computational Approaches on Typical Development and Neurodevelopmental Disorders

Peter B. Marschik<sup>1,2</sup> · Claudius A. A. Widmann<sup>1</sup> · Sigrun Lang<sup>1</sup> · Tomas Kulvicius<sup>1</sup> · Sofie Boterberg<sup>3</sup> · Karin Nielsen-Saines<sup>4</sup> · Sven Bölte<sup>5,6</sup> · Gianluca Esposito<sup>7</sup> · Anders Nordahl-Hansen<sup>8</sup> · Herbert Roeyers<sup>3</sup> · Florentin Wörgötter<sup>9,10</sup> · Christa Einspieler<sup>2</sup> · Luise Poustka<sup>1</sup> · Dajie Zhang<sup>1,2</sup>

Accepted: 14 October 2022 / Published online: 25 October 2022  
© The Author(s) 2022

## Abstract

**Objectives** Research on typically developing (TD) children and those with neurodevelopmental disorders and genetic syndromes was targeted. Specifically, studies on autism spectrum disorder, Down syndrome, Rett syndrome, fragile X syndrome, cerebral palsy, Angelman syndrome, tuberous sclerosis complex, Williams-Beuren syndrome, Cri-du-chat syndrome, Prader-Willi syndrome, and West syndrome were searched. The objectives are to review observational and computational studies on the emergence of (pre-)babbling vocalisations and outline findings on acoustic characteristics of early verbal functions.

**Methods** A comprehensive review of the literature was performed including observational and computational studies focusing on spontaneous infant vocalisations at the pre-babbling age of TD children, individuals with genetic or neurodevelopmental disorders.

**Results** While there is substantial knowledge about early vocal development in TD infants, the pre-babbling phase in infants with neurodevelopmental and genetic syndromes is scarcely scrutinised. Related approaches, paradigms, and definitions vary substantially and insights into the onset and characteristics of early verbal functions in most above-mentioned disorders are missing. Most studies focused on acoustic low-level descriptors (e.g. fundamental frequency) which bore limited clinical relevance. This calls for computational approaches to analyse features of infant typical and atypical verbal development.

**Conclusions** Pre-babbling vocalisations as precursor for future speech-language functions may reveal valuable signs for identifying infants at risk for atypical development. Observational studies should be complemented by computational approaches to enable in-depth understanding of the developing speech-language functions. By disentangling features of typical and atypical early verbal development, computational approaches may support clinical screening and evaluation.

**Keywords** Autism · Developmental disorder · Infant · Speech-language · Vocalisation

Myriad studies have furthered our understanding of the ontogeny of human behaviour and early neurofunctions that underlie our later capacities and skills. Early human behaviours are complex, dynamic, and diverse. Given commonalities in emerging neurofunctions along development, there are undeniable individual distinctions. One of the most fascinating questions in development is if, which, and how individual oscillations lead to long-term favourable or adverse outcomes? Following a neurodevelopmentalist

perspective of development, acknowledging early functions as precursors and prerequisites for later ones, we presume that early deviations or impairments precede suboptimal traits or adverse outcomes, even if the core symptomatology of certain disorders may appear later in development (as for example in the case of autism spectrum disorder, ASD; e.g. Estes et al., 2019). This assumption, also known as deep constructivist notion (neuroconstructivism; e.g. Johnson, 2000; Johnson et al., 2021; Karmiloff-Smith, 1998; Mareschal, 2011; Westermann et al., 2011), is tightly linked to attempts at detecting and defining early functional markers of neurodiversity or atypicality, i.e. predictors of developmental trajectories (D'Souza & Karmiloff-Smith, 2017; Jones et al., 2014; Karmiloff-Smith, 1998, 2009; Marschik

✉ Peter B. Marschik  
peter.marschik@med.uni-goettingen.de

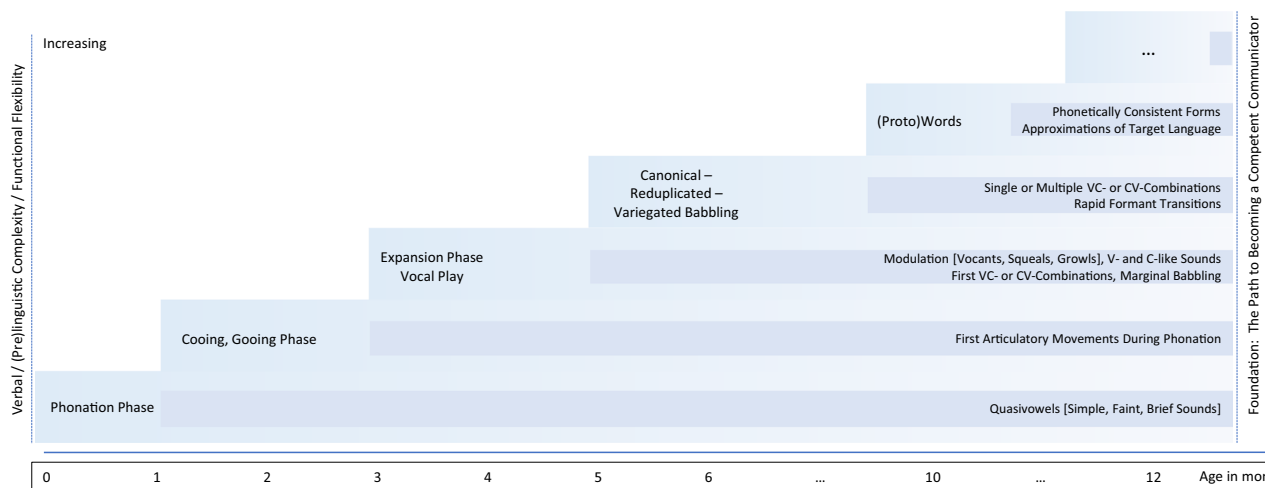
Extended author information available on the last page of the article

et al., 2014b, 2017; Micai et al., 2020). Concerns on whether behaviours reflect potential developmental atypicality or delay or mere diversity in typical development often result from recognising inter-individual discrepancies among peers caused by slowed or divergent functional acquisition within and across developmental domains, which might indicate stagnation or regression of intra-individual development. Notably, although the very early periods of speech-language development are not yet fully understood, atypicalities in the verbal domain are often one of the first perceived signs of neurodiversity during the first year of life.

Taking a closer look at the developing speech-language and communicative system, there is broad consensus regarding the essential role of prelinguistic vocalisations during early infancy for successful development of subsequent verbal functions (e.g. Karmiloff & Karmiloff-Smith, 2002; Locke, 1995; Oller, 2000; Vihman et al., 1985). Verbal development, meaning speech-language and communicative functions, follows a developmental trajectory of increasing complexity, accuracy and stability, thus building the complex human verbal capacity (e.g. Buder et al., 2013; Locke, 1995; Nathani et al., 2006; Oller, 1978, 2000; Papoušek, 1994; Stark, 1980). About four decades ago, stage-models were proposed describing developmental pathways from an infant's first cry to becoming a competent communicator (c.f. Karmiloff & Karmiloff-Smith, 2002; Koopmans-van Beinum & van der Stelt, 1986; Oller, 1978; Papoušek, 1994; Roug et al., 1989; Stark, 1980). While there are differences in exact definitions and labels for categorically distinct vocalisation types, reported age of onset and stages/phases, and the mastering of certain milestones, researchers have offered similar models which describe evolving verbal functions. In the initial developmental phase, most vocalisations are faint and brief quasi-vowels. This first phase is often referred to as phonation stage or uninterrupted phonation

stage (Fig. 1; Koopmans-van Beinum & van der Stelt, 1986; Oller, 2000). Thereafter, emerging at 1 to 2 months of age, vocalisations with articulatory movements of the tongue during phonation are uttered, a stage which was labelled “cooing” or “gooing” phase (Oller, 1978, 2000). Approximately 2 months later, an expansion of vocal and articulatory capacities can be observed. Vocalisation types at this expansion or vocal play stage, are vowel-/consonant-like sounds, squeals, and marginal syllables. These utterances are not yet produced with the articulatory accuracy and timing of adult-speech (Fig. 1; Nathani et al., 2006; Oller, 2000; Stark, 1980). The final stage of prelinguistic development, commonly referred to as canonical babbling stage, marks an infant's start to produce speech-like syllables, usually starting between 5 and 10 months of age (Oller, 2000). Vocalisations are single or multiple consonant–vowel-combinations with rapid formant transitions between the consonantal and vocalic part. In some stage models, reduplicated and variegated babbling have been proposed as subsequent stages (Oller, 1978; Roug et al., 1989; Stark, 1980). In summary, specific vocalisation types occur in a cascading fashion and become increasingly speech-like towards the end of the first year of life, when the first (proto-)words are uttered. Besides this shift to language-specific phonetic forms, vocalisation types and developmental stages during the first year of life have been considered as universal (cf. Buder et al., 2013 who provide an acoustic phonetic catalogue of pre-speech vocalisations).

The classical approach to assess whether the above-mentioned early speech-language milestones are met, follows a perceptual segmentation-annotation-classification procedure of infant utterances. In such studies (which are observational), vocalisation-entities are commonly defined through the breath group criterion (i.e. vocalisation(s) uttered in the exhalation/expiration phase of one breathing cycle;



**Fig. 1** The developing speech-language capacity

Lynch et al., 1995a; Nathani & Oller, 2001) and segmented accordingly. Other approaches segmenting infant speech have differentiated vocalisations through a pause criterion (e.g. pauses longer 300 ms subdivide vocalisation clusters; Oller et al., 2010). In either way, the segmentation step is usually followed by an annotation process, in which trained listeners assign vocalisations to the predefined vocalisation classes (e.g. Koopmans-van Beinum & van der Stelt, 1986; Lang et al., 2021; Lynch et al., 1995a; Nathani et al., 2006; Oller, 1978, 2000; Roug et al., 1989; Stark, 1980). Recently, a citizen science study externally validated the expert classification of babbling vocalisations and the onset of canonical babbling (Cychosz et al., 2021). Together with findings on auditory Gestalt perception of experts and naïve listeners differentiating early verbal functions of infants with neurodevelopmental disorders (NDDs), this points to the existence of an intrinsic human Gestalt of different vocal categories or typical vs. atypical pre-linguistic vocalisations (Marschik et al., 2012a). Human auditory Gestalt perception, or the adult capacity of intuitively recognising different vocal categories, becomes more robust when evaluating “higher order verbal functions” of infants. Explicitly, babbling vocalisations, being more salient in form, are easier to be categorised by listeners as compared to pre-babbling vocalisations uttered in the first 5 months of life (e.g. Marschik et al., 2012a; Pokorny et al., 2018).

In the first 5 months of life, before the canonical babbling stage, the various stage-models concordantly include descriptions of a developmental pathway from simple phonation to an expansion phase (Fig. 1, e.g. Kent, 2022; Nathani et al., 2006; Oller, 2000). Oller and colleagues introduced a classification scheme of three types of infant vocalisations: cry, laughter and protophones; the latter are defined as precursors to speech and subdivided into vocants, squeals and growls (Jhang & Oller, 2017; Oller et al., 2013). Interestingly, evidence showed spontaneously produced protophones to outnumber cries and laughter from early on (Jhang & Oller, 2017; Oller et al., 2019). The importance of protophones lies, in contrast to cry and laughter, in their functional flexibility. They can be used in variable contexts and may fulfil different communicative functions (Jhang & Oller, 2017; Oller et al., 2013). Besides flexibility in functioning, the ontogeny of vocalisations has been discussed in terms of physiological constraints. Physiological adaptation of peripheral anatomical structures, such as the larynx descent or vocal-tract shape (e.g. Fitch, 2010; Lieberman et al., 2001) as well as neurophysiological changes governing the functional output, shape the development and the increasing complexity of vocalisations (see Fig. 1; e.g. Kent, 2021, 2022; Oller, 2000; Zhang & Ghazanfar, 2020).

In infants with various developmental disorders (DDs), an increasing number of studies has investigated the prelinguistic development aiming to detect early atypical findings

and potential associations with later speech-language development (for reviews see for example Lang et al., 2019; Roche et al., 2018; Yankowitz et al., 2019). Canonical babbling, for example, was reported to be delayed or deviant in infants with hearing impairment (HI; Eilers & Oller, 1994; Koopmans-van Beinum et al., 2001; Moeller et al., 2007; Nathani Iyer & Oller, 2008; Shehata-Dieler et al., 2013; von Hapsburg & Davis, 2006), Down syndrome (DS; Lohmander et al., 2017; Lynch et al., 1995b), cerebral palsy (CP; Levin, 1999; Nyman & Lohmander, 2018), Williams-Beuren syndrome (WBS; Masataka, 2001), Cri-du-chat syndrome (CDS; Sohner & Mitchell, 1991), tuberous sclerosis complex (TSC; Gipson et al., 2021), autism spectrum disorder (ASD; Patten et al., 2014; Paul et al., 2011; Yankowitz et al., 2022), Rett syndrome (RTT; Einspieler et al., 2014; Marschik et al., 2012b, 2013), and fragile X syndrome (FXS; Belardi et al., 2017; Marschik et al., 2014a). Findings were however inconsistent and may depend on measures applied. For example, some infants with late detected developmental disorders (LDDs such as ASD, RTT, FXS) exhibited a delayed onset of canonical babbling whereas others have reached this milestone at an adequate age, i.e. between 5 and 10 months (Bartl-Pokorny et al., 2022; Lang et al., 2019; Marschik et al., 2013; Yankowitz et al., 2019, 2022).

As findings regarding achievement of developmental milestones in infants with DDs were inconclusive, recent research increasingly aimed at gaining in-depth knowledge about early vocal patterns through the extraction and characterisation of acoustic features of emerging verbal functions. For example, in cry but also in spontaneous infant vocal patterns acoustic features like fundamental frequency (lowest frequency of a periodic waveform, usually denoted as  $F_0$ ) or duration of vocalisations have been documented (Borysiak et al., 2017; Buder et al., 2013; Hamrick et al., 2019; Kent & Murray, 1982; Wermke & Robb, 2010). More complex models on analysing acoustic properties of infant vocalisations include machine learning approaches applied on a set of parameters or features on signal level (Pokorny et al., 2020; Schuller & Batliner, 2013). There are established parameter sets for analysing voice features such as the extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS; Eyben et al., 2015) and the Computational Paralinguistics Challenges parameter set (ComParE; Schuller et al., 2013).

The features that are included in such sets can be subdivided into three categories: parameters related to frequency aspects (e.g. pitch), parameters related to the energy or amplitude of the signal (e.g. harmonics-to-noise ratio; HNR) and spectral parameters (e.g. harmonic differences). Another common approach to produce a more specialised parameter set is the usage of the unsupervised Bag-of-Audio-Words (BoAW) approach to the best set of features according to a customised codebook quantisation of the low level descriptors (LLDs). In addition, machine learning models have

been applied to vocalisations including neural networks in different varieties, testing classification tasks (e.g. adult vs. infant speech, canonical vs. non-canonical utterances; Ebrahimpour et al., 2020; Warlaumont et al., 2010). In our group, we have utilised a machine learning approach (i.e. support vector machines), that focused on automatic pre-verbal vocalisation-based differentiation between typically developing infants and infants later diagnosed with RTT, FXS or ASD (Pokorny et al., 2016a, 2017, 2022). Studies evaluating acoustic features of early vocalisations or applying machine learning models or neural networks will be referred to as “computational studies” hereafter.

Given recent efforts to perceptually classify preverbal vocal patterns and characterise them acoustically, there is still a lack of synergised information in the field of prodromal or pre-diagnostic development in infants with neurodevelopmental or genetic disorders, especially concerning the pre-babbling phase. Therefore, the current article aimed to (i) outline characteristics of age-specific pre-linguistic vocalisations in the first 5 months of age (i.e. the pre-babbling phase), (ii) summarise computer-based approaches for the automated analysis of physiological and pathological pre-babbling vocalisations, and (iii) compare computer-based approaches on atypical early verbal functions and outline their potential to serve as neurofunctional marker of DDs.

## Methods

To address the above-mentioned issues, we systematically searched the existing literature for (a) characteristics of and (b) state-of-the-art computational and observational methods on prelinguistic vocalisations in infants with DDs. We conducted two rounds of paper extraction and selection, the first one in September 2021 and a second one in February and March 2022 in the following online electronic databases: PubMed, Web of Science, Science Direct, Scopus, and PsycINFO using the search strings “infan\* AND (prelinguistic OR preverbal OR cooing OR babbling OR vocal) AND (syndrome OR “genetic disorder” OR “developmental disorder”)” and “infan\* AND vocal\* AND (“computational analysis” OR “acoustic analysis” OR “audio analysis”)”.

Following this initial step, we performed an ancestral search for papers from the retrieved articles and searched Google Scholar for further publications. The retrieved articles were screened by two independent raters (CW and SL). Results were discussed with the co-authors, duplicates were removed, and articles were selected according to the following criteria: (1) peer-reviewed; (2) original studies or reviews and meta-analyses; (3) written in English; and (4) focusing on the pre-babbling age (0 to 5 months) in (4a) typically developing infants and (4b) infants at elevated likelihood for or diagnosed with neurodevelopmental disorders

(NDDs), late detected developmental disorders (LDDD), genetic syndromes, or developmental disorders (DDs). Articles of interest were those based on human coder-based assessments (observational studies) as well as articles on machine learning approaches (computational studies). We intended to focus on spontaneous infant vocalisations and excluded all studies analysing or reporting infant cry or distress vocalisations as well as vocalisations from parent–child interaction paradigms (PCI).

## Results

Our literature selection process led to a total of 27 papers, 17 of which are on pre-babbling in infants diagnosed with neurodevelopmental disorders or genetic syndromes applying observational methods (Table 1). Six articles focused on DS, seven on ASD (one of them also including infants with TSC), three on RTT or the preserved speech variant of RTT (PSV), and one on PWS. Two of the 17 articles reported acoustic features in addition to observational characteristics. The remaining ten articles focused on acoustic features/computational models, three studies applying computational methods on pre-babbling behaviour in TD infants (Table 2) and seven papers discussed the babbling stage in infants later diagnosed with a DD (i.e. ASD, CDS, PSV-RTT, RTT, WS and one study reporting on ASD, FXS, and RTT; Table 2). It is important to note that a differentiation between spontaneous vocalisations vs. vocalisations in interactive settings could not be reliably done for all articles. Thus, against the initially set exclusion criterion, we decided to report all observational studies of this age-range and outlined information on data sampling whenever possible (Tables 1 and 2).

Whilst there is a number of studies reporting early physiological development according to the established stage models (Fig. 1), reports of atypical development in infants with neurodevelopmental disorders or genetic syndromes in the younger ages are rare (Table 1). Most of the 17 included studies report on expanded age-bands up to 24 months; very few explicitly investigate the characteristics of early verbal functions emerging in the first 5 months of life (Brisson et al., 2014; Maestro et al., 2002; Pansy et al., 2019; Zappella et al., 2015). Most studies investigate developing verbal functions applying the classical approach of perceptual segmentation-annotation-classification. There is less effort present in delineating acoustic features (such as duration of vocalisations, syllables or phrases, pitch, fundamental frequency ( $F_0$ ) or intonation contours; Brisson et al., 2014; Lynch et al., 1995a). Observational studies reveal inconclusive results on behavioural differences in pre-babbling vocalisations in infants with DDs and typical development. Compared to TD infants several diverse behaviours have been reported for DD: e.g. longer duration of rhythmic units

**Table 1** Studies analysing pre-babbling behaviour in infants with neurodevelopmental disorders or genetic syndromes applying observational methods (ascending order)

Authors (year of publication)	Condition: sample size (M/F)	Age range (in months)	Data sampling	Behaviours analysed related to pre-babbling vocalisations	Key results
Smith and Oller (1981)	TD: <i>n</i> = 9 (7/2) DS: <i>n</i> = 10 (5/5)	T1: 0–3 T2: 3–6 T3: 6–9 T4: 9–12 T5: 12–15 T6 (DS): 15–18 T7 (DS): 18–21	<ul style="list-style-type: none"> <li>■ Longitudinal</li> <li>■ Prospective</li> <li>■ Laboratory audio recordings</li> <li>■ Every 3 months</li> <li>■ Interaction sequences</li> <li>■ 30 min duration</li> </ul>	<ul style="list-style-type: none"> <li>■ Vocalic productions</li> <li>■ Consonantal productions</li> <li>■ Transition from pre-babbling to babbling</li> </ul>	<ul style="list-style-type: none"> <li>■ Substantial similarities between TD and DS infants in terms of the place of articulation of vowels and consonants</li> <li>■ Transition to babbling at 6–10 (average: 7.9) months in TD and at 7–12 (average: 8.4) months in infants with DS</li> </ul>
Legerstee et al. (1992)	DS: <i>n</i> = 8 (4/4)	2–10	<ul style="list-style-type: none"> <li>■ Longitudinal</li> <li>■ Prospective</li> <li>■ Laboratory video recordings</li> <li>■ Biweekly (8–24 weeks)</li> <li>■ Monthly (25–40 weeks)</li> <li>■ Fixed experimental setup with mother, stranger, and object</li> </ul>	<ul style="list-style-type: none"> <li>■ Three vocal categories: vocalic (nonspeech-like), melodic (speech-like), emotional (laughing, crying, fussing)</li> </ul>	<ul style="list-style-type: none"> <li>■ From 17 weeks onwards, vocalisations could be categorised as vocalic, melodic, and emotional; more vocalisations were produced to people than to objects</li> <li>■ Only from 6 months onwards mother–stranger and active–passive differentiation</li> </ul>
Steffens et al. (1992)	TD: <i>n</i> = 27 (17/10) DS: <i>n</i> = 13 (4/9)	4–18	<ul style="list-style-type: none"> <li>■ Longitudinal</li> <li>■ Prospective</li> <li>■ Laboratory audio recordings</li> <li>■ Bimonthly</li> </ul>	<ul style="list-style-type: none"> <li>■ Four vocalisation types: quasi-resonant syllables, fully resonant syllables, marginal syllables, and canonical syllables</li> </ul>	<ul style="list-style-type: none"> <li>■ In both TD and DS infants increase of fully resonant and canonical syllables</li> <li>■ In both TD and DS infants decrease of quasi-resonant and marginal syllables</li> <li>■ No significant differences between DS and TD infants’ developmental patterns concerning pre-babbling</li> </ul>
Lynch et al., (1995b)	TD: <i>n</i> = 27 (17/10) DS: <i>n</i> = 13 (4/9)	4–18	<ul style="list-style-type: none"> <li>■ Longitudinal</li> <li>■ Prospective</li> <li>■ Laboratory audio recordings</li> <li>■ Monthly, merged to 2-month-sampling bins</li> <li>■ 20 min duration</li> </ul>	<ul style="list-style-type: none"> <li>■ Four vocalisation types: quasi-resonant syllables, fully resonant syllables, marginal syllables, and canonical syllables</li> <li>■ Focus on transition to canonical babbling</li> </ul>	<ul style="list-style-type: none"> <li>■ No details on pre-babbling</li> <li>■ As a group, infants with DS showed a delayed transition from pre-babbling to babbling and “less stable” canonical babbling</li> </ul>
Lynch et al., (1995a)	TD: <i>n</i> = 8 (–/–) DS: <i>n</i> = 8 (–/–)	T1: 2–4 T2: 6–8 T3: 10–12	<ul style="list-style-type: none"> <li>■ Longitudinal</li> <li>■ Prospective</li> <li>■ Laboratory audio recordings</li> <li>■ Monthly</li> <li>■ 20 min duration</li> </ul>	<ul style="list-style-type: none"> <li>■ Non-vegetative utterances identified as syllables, utterances, and prelinguistic phrases</li> <li>■ Perceptual characteristics of syllables: nucleus in isolation (vowel-like sound) or closant (consonant-like sound) and nucleus in sequence</li> </ul>	<ul style="list-style-type: none"> <li>■ Infants with DS showed similar amounts but longer duration of rhythmic units (syllables, utterances, phrases) when compared to TD infants</li> </ul>

**Table 1** (continued)

Observational studies on pre-babbling in neurodevelopmental disorders or genetic syndromes	Condition: sample size (M/F)	Age range (in months)	Data sampling	Behaviours analysed related to pre-babbling vocalisations	Key results
Maestro et al. (2001)	TD: <i>n</i> = 15 (11/4) ASD: <i>n</i> = 15 (10/5)	T1: 0–6 T2: 6–12 T3: 12–18 T4: 18–24	■ RVA, home recordings ■ One video per child	■ Vocalisations in interactive settings ■ No details concerning vocalisation development during the pre-babbling phase	■ Infants with ASD showed deviations in socio-communicative development in their first month when compared to TD infants ■ No details concerning vocalisation development during the pre-babbling phase
Maestro et al. (2002)	TD: <i>n</i> = 15 (9/6) ASD or PDD-NOS: <i>n</i> = 15 (10/5)	0–6	■ RVA, home recordings ■ 10–62 min duration ■ One video per child	■ Vocalisations, sounds or babbling towards another person ■ Vocalisations, sounds or babbling towards objects	■ Infants with ASD showed significantly lower rates of vocalising towards another person or object as compared to TD children
Maestro et al. (2005)	TD: <i>n</i> = 13 (5/8) ASD: <i>n</i> = 15 (11/4)	T1: 0–6 T2: 6–12	■ RVA, home recordings ■ 10–82 min duration	■ Vocalisations towards another person (social) ■ Vocalisations towards objects (non-social)	■ Infants with ASD showed an atypical trajectory of social and non-social activities
Marschik et al. (2009)	PSV-RTT: <i>n</i> = 1 (0/1)	0–24	■ Longitudinal ■ RVA, home recordings ■ Additional: parental diaries, medical history	■ Vocalisations	■ Typical vocalisations interspersed with atypical episodes of forceful and/or inspiratory vocalisations
Apicella et al. (2013)	TD: <i>n</i> = 9 (–/–) ASD: <i>n</i> = 10 (9/1)	T1: 0–6 T2: 6–12	■ RVA, home recordings ■ Dyadic interaction sequences ■ T1 = 0–6 months: 46 min ■ T2 = 6–12 months: 51 min	■ Caregiver–infant reciprocity scale (CIRS) <sup>a</sup> : ■ Vocalisations to elicit caregivers' response ■ Vocalisations to respond to caregiver	■ Infants with ASD showed similar amounts of vocalisations in the first half year of life compared to TD infants but a significant decrease of total vocalisations from T1 (0–6 months) to T2 (6–12 months)
Marschik et al. (2013)	RTT: <i>n</i> = 10 (0/10) PSV-RTT: <i>n</i> = 5 (0/5)	0–24	■ RVA, home recordings ■ Total: 2431 min, typical RTT: 2041 min, PSV-RTT: 390 min	■ Realisations of age-specific speech-language milestones (cooing, babbling, (proto-) words, word combinations) ■ Exclusion of fixed vocal signals (e.g. crying, laughing), vegetative sounds (e.g. sneezing, coughing)	■ Infants with RTT or PSV-RTT showed diverging behaviours when compared to TD infants (delayed or non-achievement of milestones)

**Table 1** (continued)

Authors (year of publication)	Condition: sample size (M/F)	Age range (in months)	Data sampling	Behaviours analysed related to pre-babbling vocalisations	Key results
Brissou et al. (2014)	TD: $n = 13$ (9/4) ASD: $n = 13$ (11/2)	0–6	<ul style="list-style-type: none"> <li>■ RVA, home recordings</li> <li>■ Dyadic interaction sequences</li> </ul>	<ul style="list-style-type: none"> <li>■ Prosodic analysis: duration, mean pitch, fundamental frequency</li> <li>■ Analysis of intonation contours</li> <li>■ Exclusion of non-speech productions (e.g. cries, fusses, laughter)</li> <li>■ Vocalisations</li> </ul>	<ul style="list-style-type: none"> <li>■ Infants with ASD produced vocalisations with similar duration and pitch</li> <li>■ Infants with ASD used simple intonation contours more often; vocalisations contained less complex modulations</li> <li>■ Infants with RTT showed indications of speech-language atypicalities in the first year of life before showing clear signs of regression</li> </ul>
Einspieler et al. (2014)	RTT: $n = 2$ (0/2) monozygotic twins	0–22	<ul style="list-style-type: none"> <li>■ Longitudinal</li> <li>■ RVA, home recordings</li> <li>■ Additionally: parental diaries, medical history</li> <li>■ Total twin 1: 55 min/38 clips, total twin 2: 54 min/36 clips</li> </ul>		
Zappella et al. (2015)	Suspected ASD: $n = 18$ (18/0) ASD + : $n = 10$ ASD - : $n = 1$ TSC: $n = 7$	T1: 1–2 T2: 3–4 T3: 5–6	<ul style="list-style-type: none"> <li>■ RVA, home recordings</li> <li>■ Total: 314 min, 3–11 min per infant per 2-month-range</li> </ul>	<ul style="list-style-type: none"> <li>■ Unspecific vocalisations (basic sounds or vowel-like sounds)</li> <li>■ Pleasure vocalisations/laughter</li> <li>■ Cooing (melodic modulated vocalisations)</li> </ul>	<ul style="list-style-type: none"> <li>■ The majority of infants later diagnosed with ASD or TSC showed age-adequate vocalisations (i.e. cooing and pleasure vocalisations) during the first 6 months</li> <li>■ Six infants out of the whole sample did not develop cooing during the first 6 months</li> </ul>
Chericoni et al. (2016)	TD: $n = 10$ (8/2) ASD: $n = 10$ (9/1)	T1: 0–6 T2: 6–12 T3: 12–18	<ul style="list-style-type: none"> <li>■ Longitudinal</li> <li>■ RVA, home recordings</li> <li>■ Dyadic interaction sequences</li> <li>■ Each participant 3 min per time period T1, T2, T3</li> <li>■ Total: 142 sequences (T1: 47, T2: 47, T3: 48 sequences)</li> </ul>	<ul style="list-style-type: none"> <li>■ Vocalisations consisting of vowels/non-reduplicated consonants and vowels</li> <li>■ Transition to babbling and words</li> <li>■ Exclusion of non-speech sounds (such as squeals, yells, growls, or grunts)</li> </ul>	<ul style="list-style-type: none"> <li>■ At T1 (0–6 months) no differences in the rate of vocalisations per minute were observed between infants with ASD and TD infants</li> </ul>
Pansy et al. (2019)	PWS: $n = 1$ (1/0)	0–6	<ul style="list-style-type: none"> <li>■ Longitudinal</li> <li>■ Prospective</li> <li>■ Audio–video recording at 27 weeks</li> <li>■ 10 min duration</li> </ul>	<ul style="list-style-type: none"> <li>■ Classification of vocalisations according to the Stark Assessment of Early Vocal Development-Revised (SAEVD-R)<sup>6</sup></li> </ul>	<ul style="list-style-type: none"> <li>■ Age-adequate distribution of SAEVD-R-levels with high proportion of level-3 vocalisations</li> <li>■ Vocalisation quality was observed to be monotonous and with peculiar harmonic structure</li> <li>■ Transition to level-4 vocalisations (canonical babbling) was not reached during the first 6 months</li> </ul>

**Table 1** (continued)

Authors (year of publication)	Condition: sample size (M/F)	Age range (in months)	Data sampling	Behaviours analysed related to pre-babbling vocalisations	Key results
Omnivello et al. (2021)	DS: $n = 74$ (38/36)	4–18	<ul style="list-style-type: none"> <li>■ Cross-sectional</li> <li>■ Prospective</li> <li>■ Laboratory test situation (approximately 40 min)</li> </ul>	<ul style="list-style-type: none"> <li>■ BSID-III<sup>c</sup> Expressive communication scale:               <ul style="list-style-type: none"> <li>■ Vocalisations of mood</li> <li>■ Two vowel sounds</li> <li>■ Two consonant sounds</li> <li>■ Further milestones (babbling and word use)</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>■ Most prominent delays in development of infants with DS were observed in the communicative capabilities               <ul style="list-style-type: none"> <li>■ The first 6 months of vocal development were similar in infants with DS and TD infants concerning vocalisation of mood and social vocalising</li> <li>■ The onset of more speech-like communicative functions (i.e. babbling and first words) was delayed in infants with DS</li> </ul> </li> </ul>

ASD autism spectrum disorder, DS Down syndrome, PDD-NOS Pervasive Developmental Disorder-Not Otherwise Specified, PSV-RTT preserved speech variant-Rett syndrome, PWS Prader-Willi syndrome, RTT Rett syndrome, RVA retrospective video analysis, TD typically developing, TSC tuberous sclerosis complex

<sup>a</sup>Only verbal scale attributes of CIRS (for details see Apicella et al., 2013)

<sup>b</sup>SAEVD-R (for details see Nathani et al., 2006)

<sup>c</sup>Bayley Scales of Infant and Toddler Development Third Edition (Bayley, 2006)



**Table 2** Studies applying computational methods on pre-babbling behaviour in typically developing infants and on babbling behaviour in infants with neurodevelopmental disorders or genetic syndromes (ascending order)

Computational studies on pre-babbling in TD					
Authors (year of publication)	Condition: sample size (M/F)	Age range (in months)	Data sampling	Behaviours analysed related to pre-babbling vocalisations	Methods/results
Warlaumont et al. (2010)	TD: $n=6$ (2/4)	T1: 3–6 T2: 6–9 T3: 9–12	<ul style="list-style-type: none"> <li>■ Longitudinal</li> <li>■ Prospective</li> <li>■ Laboratory audio–video recordings</li> <li>■ Interaction and non-interaction sequences</li> <li>■ At least two 20-min sessions per age interval, first 49 utterances coded</li> </ul>	<ul style="list-style-type: none"> <li>■ Interactive and spontaneously produced vocalisations</li> <li>■ Vocalisation types: vocants, squeals, and growls</li> <li>■ Exclusion of crying, distress, vegetative sounds</li> </ul>	<ul style="list-style-type: none"> <li>■ Segmented infant non-cry vocalisations with a duration of 1 s were transformed into a spectrogram using the fast Fourier transform and passed into the self-organising map with a 4 × 4 grid which was then used as the input for a single-layer perceptron</li> <li>■ Classification of vocalisation types: vocants, squeals, and growls with an accuracy of 0.55</li> </ul>
Ebrahimpour et al. (2020)	TD: $n=15$ (–/–)	3–18	<ul style="list-style-type: none"> <li>■ Longitudinal</li> <li>■ Prospective</li> <li>■ LENA<sup>a</sup>: all day recordings</li> <li>■ Three 5-min periods of active vocalisations</li> </ul>	<ul style="list-style-type: none"> <li>■ Vocalisation categories: laugh/cry, non-canonical babbling (vowels or consonants), canonical babbling (well-structured syllables with consonants and vowels)</li> </ul>	<ul style="list-style-type: none"> <li>■ The dataset presented by Warlaumont et al. (2010) was used</li> <li>■ Two end-to-end convolutional neural networks using time-domain waveform data as input were used. Different conditions were tested: infant vs. adult, vocalisation vs. non-vocalisation, infant-directed speech (IDS) vs. adult-directed speech (ADS), laugh/cry vs. canonical/non-canonical vs. IDS vs. ADS, laugh/cry vs. canonical vs. non-canonical vs. IDS/ADS, laugh/cry vs. canonical vs. non-canonical vs. IDS vs. ADS</li> <li>■ Classification accuracy of 0.41–0.94 in the above-mentioned cases was obtained (above chance)</li> <li>■ Non-canonical babbling vs. canonical babbling and IDS vs. ADS changed as a function of age and became more accurate with increasing age</li> </ul>

Table 2 (continued)

Authors (year of publication)	Condition: sample size (M/F)	Age range (in months)	Data sampling	Behaviours analysed related to pre-babbling vocalisations	Methods/results
Li et al. (2021)	TD: $n = 119$ (-/-)	3–12	<ul style="list-style-type: none"> <li>■ Longitudinal</li> <li>■ Prospective</li> <li>■ Home and laboratory audio–video recordings</li> <li>■ Home recordings: LENA<sup>a</sup></li> <li>■ Lab recordings: semi-structured interaction sequences and still face procedure</li> </ul>	<ul style="list-style-type: none"> <li>■ Vocalisation types: cry, fuss, laughter, babbling, screech</li> </ul>	<ul style="list-style-type: none"> <li>■ 1582 spectral and prosodic features were used</li> <li>■ Multiple different machine learning methods (i.e. linear discriminant analysis, convolutional neural network with self-attention, feed-forward fully-connected neural network) were applied to differentiate infant and adult vocalisation types</li> <li>■ From three types of classifiers evaluated, convolutional neural network with self-attention achieved the highest classification accuracy of 87%</li> </ul>
Computational studies on babbling in neurodevelopmental disorders or genetic syndromes					
Sohner and Mitchell (1991)	CDS: $n = 1$ (-/-)	8–26	<ul style="list-style-type: none"> <li>■ Longitudinal</li> <li>■ Prospective</li> <li>■ Audio recordings</li> </ul>	<ul style="list-style-type: none"> <li>■ Comfort state vocalisations (fundamental frequency, intonation contours)</li> </ul>	<ul style="list-style-type: none"> <li>■ Analysis of the fundamental frequency of non-cry vocalisation across the period of 8–26 months were performed</li> <li>■ High fundamental frequency in comfort state vocalisations (521 – 622 Hz) with no downward trend as can be observed in TD infants with a fundamental frequency of 410 Hz at 8 months decreasing to 310 Hz at 26 months</li> <li>■ Limited inter-utterance variation of fundamental frequency with a predominance of falling intonation contours</li> </ul>

**Table 2** (continued)

Authors (year of publication)	Condition: sample size (M/F)	Age range (in months)	Data sampling	Behaviours analysed related to pre-babbling vocalisations	Methods/results
Pokorny et al. (2016a)	TD: n=4 (0/4)	6–12	<ul style="list-style-type: none"> <li>■ RVA, home video recordings</li> <li>■ Total: &gt; 1000 min</li> <li>■ RTT: 2199 vocalisations</li> <li>■ TD: 2479 vocalisations</li> </ul>	<ul style="list-style-type: none"> <li>■ Vocalisations (vocal breathing groups)</li> <li>■ Exclusion of vegetative sounds (e.g. hiccups, sneezes)</li> </ul>	<ul style="list-style-type: none"> <li>■ Feature representation of the signal using the ComParE<sup>c</sup></li> <li>■ Classification of TD vs. RTT using a support vector machine</li> <li>■ Comparison of two approaches: audio signal with speaker normalisation and without speaker normalisation (speaker normalisation being infant dependent normalisation to the interval [0,1])</li> <li>■ Classification accuracy was 76.5% for classification of TD infants vs. infants with RTT vocalisations using speaker normalisation</li> </ul>
	RTT: n=4 (0/4)				
Pokorny et al. (2016b)	ASD: n=14 (13/-) IT	7–12	<ul style="list-style-type: none"> <li>■ RVA, home video recordings</li> <li>■ Total: &gt; 1000 min</li> <li>■ ASD: 696 vocalisations</li> <li>■ FXS: 87 vocalisations</li> <li>■ RTT: 2585 vocalisations</li> <li>■ TD: 1535 vocalisations</li> </ul>	<ul style="list-style-type: none"> <li>■ Voice-active segments</li> <li>■ Vocalisations/utterances (vocal breathing groups)</li> <li>■ Exclusion of vegetative sounds (e.g. hiccups, sneezes)</li> </ul>	<ul style="list-style-type: none"> <li>■ Multiple different automatic voice detection algorithms (i.e. cVaDV1, hidden Markov models, support vector machines, and random forests) were compared against manual annotation</li> <li>■ The best performance was obtained using the hidden Markov model with the Gaussian mixture model achieving acoustic event detection accuracy of 0.29</li> </ul>
	FXS: n=1 (1/-) AT RTT: n=2 (-/2) AT RTT: n=2 (-/2) DE RTT: n=5 (-/5) UK TD: n=9 (5/4) AT				

Table 2 (continued)

Authors (year of publication)	Condition: sample size (M/F)	Age range (in months)	Data sampling	Behaviours analysed related to pre-babbling vocalisations	Methods/results
Pokorny et al. (2017)	TD: <i>n</i> = 10 (5/5) ASD: <i>n</i> = 10 (5/5)	10	<ul style="list-style-type: none"> <li>■ Longitudinal</li> <li>■ Prospective</li> <li>■ Laboratory audio–video recordings</li> <li>■ Parent–child interaction sequences</li> <li>■ 12 min duration</li> </ul>	<ul style="list-style-type: none"> <li>■ Vocalisations</li> <li>■ Types of vocalisations more complex than single canonical syllables, based on Stark Assessment of Early vocal development Revised (SAEVD-R)<sup>b</sup></li> <li>■ Exclusion of vegetative sounds</li> </ul>	<ul style="list-style-type: none"> <li>■ eGeMAPS<sup>d</sup> features were compared between infants with ASD and TD infants applying Mann–Whitney <i>U</i> test</li> <li>■ A linear kernel support vector machine and a single-layer bidirectional long-short term memory neural network were tested</li> <li>■ For 54 out of 88 extracted features a significant difference between vocalisations of TD infants vs. infants with ASD was found</li> <li>■ Both approaches (i.e. linear kernel support vector machine and single-layer bidirectional long short-term memory neural network) achieved an accuracy of 75% in a subject-wise identification of ASD and TD infants</li> </ul>
Pokorny et al. (2018)	PSV-RTT: <i>n</i> = 1 (–/1)	7–12	<ul style="list-style-type: none"> <li>■ RVA, home video recordings</li> <li>■ Total: 363 vocalisations (61 min)</li> </ul>	<ul style="list-style-type: none"> <li>■ Typical vs. atypical vocalisations</li> <li>■ Exclusion of vegetative sounds</li> </ul>	<ul style="list-style-type: none"> <li>■ ComParE<sup>e</sup> features were compared with the atypicality factor calculated in the listener experiment (proportion of listeners that have rated the vocalisation as atypical)</li> <li>■ Atypicality was mainly related to the characteristic “timbre” and prosodic, spectral, and voice quality features in the acoustic domain</li> <li>■ More than 50% of the vocalisations of the infant with PSV-RTT were rated as atypical by at least one listener</li> </ul>

**Table 2** (continued)

Authors (year of publication)	Condition: sample size (M/F)	Age range (in months)	Data sampling	Behaviours analysed related to pre-babbling vocalisations	Methods/results
Pokorny et al. (2020)	TD: $n=3$ (-/-) RIT: $n=3$ (-/-)	6–12	<ul style="list-style-type: none"> <li>■ RVA, home video recordings</li> <li>■ Total: 3502 vocalisations</li> </ul>	<ul style="list-style-type: none"> <li>■ Typical vs. atypical vocalisations</li> </ul>	<ul style="list-style-type: none"> <li>■ Two parameter sets, eGeMAPS<sup>d</sup> and ComParE<sup>c</sup>, and a linear kernel support vector machine were used to classify typical vs. atypical vocalisations</li> <li>■ The best approach reached a performance of 0.879 unweighted average recall using the eGeMAPS<sup>d</sup> feature set</li> </ul>
Ouss et al. (2020)	TD: $n=19$ (-/-) WS: $n=32$ (-/-) Outcome 48 M: WS with ASD/ID: $n=10$ WS without ASD/ID: $n=22$	9–12	<ul style="list-style-type: none"> <li>■ Prospective</li> <li>■ Laboratory audio–video recordings</li> <li>■ 3 sequences of 3-min interaction sequences (free play without toy, free play with toy, singing to baby)</li> </ul>	<ul style="list-style-type: none"> <li>■ Infant vocalisations (babbling vocalisations, laugh, and cry)</li> <li>■ Atypical vocalisations (other noise)</li> </ul>	<ul style="list-style-type: none"> <li>■ Annotation of mother–infant speech turn taking</li> <li>■ Reduction of dataset dimensionality using principal component analysis</li> <li>■ Classification of TD vs. WS and WS with ASD/ID vs. WS without ASD/ID using a decision tree with a depth of one</li> <li>■ The approach achieved an accuracy of 76% when classifying between infants with WS and TD infants and an accuracy of 81% when distinguishing between WS with ASD and WS without ASD based on assessed vocalisation types</li> <li>■ It was found that especially in the classification of WS with ASD and WS without ASD, synchronicity and reciprocity were important factors in the interactive setting</li> </ul>

ASD autism spectrum disorder, *AT* Austria, *CDS* Cri-du-Chat syndrome, *DE* Germany, *FXS* fragile X syndrome, *ID* intellectual disorder/intellectual disability, *IT* Italy, *PSY-RIT* preserved speech variant–Rett syndrome, *RIT* Rett syndrome, *RVA* retrospective video analysis, *TD* typical development/typically developing, *UAR* unweighted average recall, *UK* United Kingdom, *WS* West syndrome

<sup>a</sup>LENA (for details see Oller et al., 2010)

<sup>b</sup>SAEVD-R (for details see Nathani et al., 2006)

<sup>c</sup>ComParE (for details see Schuller et al., 2013)

<sup>d</sup>eGeMAPS (for details see Eyben et al., 2015)

in infants with DS (Lynch et al., 1995a); divergent intonation contours and less vocal response in interactive settings (Brisson et al., 2014); some participants with ASD failed to achieve the developmental milestone “cooing” (Maestro et al., 2002; Zappella et al., 2015); typical vocalisations interspersed with atypical forceful and/or inspiratory vocalisations in infants with RTT (Marschik et al., 2009); more details on age-specific vocalisations and characteristics of this period are outlined in Table 1.

More advanced methods such as digital measurement instruments and computational analyses open new possibilities for earlier identification of atypical development, as they surpass human capabilities of perception. Most approaches identified aim to describe and investigate trends in the typical development of vocalisations throughout the first 5 months of life. Very early studies focus on a categorical analysis of vocalisations, applying spectral analysis to gain additional insights in addition to the verbal Gestalt-perception (Buder et al., 2008; Lynch et al., 1995a; Oller et al., 2019; Warlaumont et al., 2010). The spectra analysed were acquired through the application of a window function. Most commonly, a fast-Fourier transformation is used to present results as a graphical visualisation, showing the intensity of frequencies at a point in time (Heideman et al., 1985). With the resulting graphical representation, one can visually determine fundamental and formant frequencies ( $F_0$  and  $F_n$ , respectively) and the general “shape” of a vocalisation (Bauer & Kent, 1987; Kent & Murray, 1982; Oller et al., 2019). The method of spectrography has been applied in studies over the last 3 decades, finding specific intonation patterns in pre-babbling vocalisations and a developmental trajectory of the  $F_0$  and  $F_n$  (Kent & Murray, 1982). Oller and colleagues used spectrograms to visualise examples of vocants, squeals, growls, and cries at specific ages, providing a visual description of the noise found in the signal as well as other unique features (e.g.  $F_0$  contour) of the analysed classes of utterances (Oller et al., 2019). Another feature, which can be identified through inspection of the spectrogram or the waveform of a vocalisation, is the duration of a single utterance. The duration is used in several studies to gain an understanding of how utterance durations change with age (Apicella et al., 2013; Brisson et al., 2014; Lynch et al., 1995a; Smith & Oller, 1981).

More in depth analyses of audio signals require multi-dimensional parameter sets to provide feature-based representations of the underlying audio segment to a classifier, which can then build an optimal predictor for the classification scheme provided. There are pre-defined parameter sets that are commonly utilised in linguistic and acoustic analyses. Such parameter sets are for example the Computational Pralinguistics ChallengeEs parameter set (ComParE; Schuller et al., 2013) or the eGeMAPS (Eyben et al., 2015). These parameter sets consist of low-level descriptors (LLDs). LLDs

are parameters that are very closely related to the signal itself (e.g. fundamental frequency  $F_0$ , loudness). To gain further insights about the general occurrence and statistical behaviour of those LLDs, functionals (e.g. mean, kurtosis, variance) are used on top of these (Schuller & Batliner, 2013).

Yet, in the field of pre-babbling vocalisations, most studies rely on basic features such as duration or fundamental frequency to gain a more in depth understanding of infant vocalisations (Apicella et al., 2013; Brisson et al., 2014; Lynch et al., 1995a; Smith & Oller, 1981). Visual spectrogram analysis has been used to evaluate different vocalisation shapes and help estimate signal to noise ratios in certain vocalisation types (Oller et al., 2019). These approaches, whilst not utilizing advanced computational methods, highlight the importance of particular features for identification of certain vocalisation types and analysis of developmental trajectories. Lynch and colleagues, who focused on a comparison between TD children and children with DS, present the only study that employs a feature-based approach in the analysis of pre-babbling vocalisations in infants with DDs (Lynch et al., 1995a). In this study, the duration of utterances was compared between DS and TD children across respective timelines. For the first 5 months of life, no significant difference was found between TD infants and infants with DS. Nevertheless, the duration of utterances increases until 8 months of age and then decreases until 12 months of age, continuously diverging between TD and DS groups (Lynch et al., 1995a). Although the methodology is not sensitive enough for an accurate differentiation between the two studied groups, it provides a starting point in the identification of possible features that can be used for future analysis of pre-babbling vocalisations (Lynch et al., 1995a). This early phase of verbal development is not yet very well researched in terms of the effectiveness of the aforementioned parameter sets (i.e. ComParE & eGeMAPS). So far, there is a lack of studies applying advanced computational approaches as well as comparative studies that enable rendering a verdict on their applicability (see Table 2). Deep learning approaches have been applied to different settings (e.g. interactive settings, home recordings; Pokorný et al., 2020) of pre-segmented infant audio signals to solve superficial classification tasks (e.g. infant vs. adult, canonical vs. non-canonical). However, none of these studies focused on infants with DDs in the first few months of life (Ebrahimpour et al., 2020; Warlaumont et al., 2010).

Several studies on machine learning approaches applied to vocalisations in the first year of life (pre-babbling and babbling) were identified. In the pre-babbling phase, only three studies utilised approaches beyond the manual analysis of LLDs in the assessment of vocalisations in TD infants (Table 2). To the best of our knowledge, there are no studies available in infants at risk or with a later diagnosis of DDs. These approaches investigate the effectiveness of

different neural network architectures (i.e. convolutional neural network, self-organising map and perceptron hybrid network), input features (i.e. spectrograms, waveform, parametric representation), and classification schemes (i.e. infant-directed speech vs. adult-directed speech, infant vs. adult, vocalisation vs. non-vocalisation, canonical vs. non-canonical; vocant vs. squeal vs. growl; Ebrahimpour et al., 2020; Li et al., 2021; Warlaumont et al., 2010). Opposed to that, in the babbling phase, a number of studies analyse verbal capacities utilizing computational approaches (e.g. Pokorny et al., 2018, 2020, 2022). In general, manual analysis of LLDs such as fundamental frequency ( $F_0$ ) is not very common for babbling vocalisations. Spectrographic analysis is very often used only for representational purposes, e.g. to represent different syllable types (e.g. Poeppel & Assaneo, 2020). For analysis and detection of atypical development by utilising computational methods, the number of approaches described is limited (Table 2).

## Discussion

Some 40 years ago, the field of early infant vocalisation study was revolutionised with new ways to assess, measure and interpret early development (Koopmans-van Beinum & van der Stelt, 1986; Oller, 1978; Papoušek, 1994; Roug et al., 1989; Stark, 1980). Since then, we have learned a lot about infant prelinguistic development and vocalisation categories. Most studies, however, focused on babbling and the emergence of first words (second half of the first year of life) whilst the pre-babbling phase (first months of life), especially in infants at elevated likelihood for or diagnosed with neurodevelopmental disorders and genetic syndromes, was less researched.

The very early phase of verbal development is mostly described through the achievement of certain milestones (e.g. phonation, cooing, expansion) or via perceptual assignment of infant vocalisations to certain types (e.g. vocant, canonical syllable). Another, albeit still rarely used approach is the description of infant vocalisations through acoustic features (e.g. duration, mean pitch,  $F_0$ ). Studies have only recently focused on the investigation of quantitative changes of different vocalisation types in the first 5 months of life (Jhang & Oller, 2017; Oller et al., 2013, 2019). However, these studies have not assessed infants with developmental disorders or genetic syndromes so far. Threshold definitions, such as the canonical babbling ratio (CBR) applied in the second half of the first year of life, have to the best of our knowledge, not yet been developed or used for types of pre-babbling vocalisations. For the later stages of development, a number of different approaches to define the onset of certain functions (e.g. canonical babbling) providing similar critical time periods in which milestones are achieved (Lang et al., 2021; Molemans et al., 2012; Oller, 2000), have been

proposed. Oller and colleagues (Oller et al., 1998, 1999) reported that delayed onset of canonical babbling is a precursor to later adverse linguistic functioning. Whether precursors of atypical development may already be detected in earlier vocalisations has not yet been investigated. Further research observing typical verbal development is still needed for a basis to understand deviant patterns and trajectories.

Besides pioneering the field of perceptively evaluating infant vocalisations, Oller and colleagues were also at the forefront to propose semi-automated recording and analytical tools for the assessment of infant vocalisations (e.g. LENA system; Oller et al., 2010). Challenges of recording preverbal data as well as advantages of automated tools for the acquisition and analyses of acoustic features have been increasingly discussed (Pokorny et al., 2020). The aim of this article is not to discuss pros and cons of automated data acquisition approaches but to focus on whether such undertakings have been utilised in the study of infant vocalisations in the first half year of life, in typical cohorts, in individuals at elevated likelihood for DDs, or groups with DDs or pre-/perinatally diagnosed disorders.

When looking beyond behavioural observations and general perceptual evaluations of early infant vocalisations, there is a lack of computational methods that study, substantiate, and support the findings of observational studies. We found that despite the existence of thoroughly tested computational approaches for babbling-vocalisations, there are no attempts to use these methods in the evaluation of pre-babbling vocalisations. These perceptually less salient vocalisations, as compared to canonical babbling, have preferably been studied through simple LLDs such as  $F_0$  and duration. Only a few studies have used more advanced computational approaches to prove the applicability and value of such approaches in the field of pre-babbling vocalisations (Ebrahimpour et al., 2020; Li et al., 2021; Warlaumont et al., 2010). Besides missing analytical approaches, there is also a lack of standardisation of coding-schemes and datasets, which impedes the comparability of performance between applied computational models in the field of speech-language analysis in the first 5 months of life. Additionally, the sample sizes investigated in observational and computational studies are usually small (i.e. 1–119; see Table 2). Generalisation capabilities of machine learning approaches applied on small dataset sets are questionable. Computational or feature-based approaches are underrepresented in studying pre-babbling vocalisations, especially in infants with NDDs (Brisson et al., 2014; Lynch et al., 1995a). To fingerprint early neurofunctional development and its deviations (Marschik et al., 2017), we need in-depth understanding of physiological functioning as well as disorder specific characteristics. Early verbal development is one domain of interest cluing in the integrity of the developing nervous system. Recent development of analytical tools appear well suited for analysing pre-linguistic vocalisations

at pre-babbling age to enhance our insights into emerging early verbal functions. Pioneer work is required to verify computational tools in identifying disorder-specific features in early vocalisations, which may inform future clinical diagnoses and be used for monitoring therapeutic success.

**Acknowledgements** We would like to thank the members of the *Systemic Ethology and Developmental Science Team (SEE)*, especially Christiane Theodossiou-Wegner for critically discussing the manuscript; PBM and FW are supported by the Volkswagen Foundation – IDENTIFIED and a Leibniz Science Campus Audacity Award; CW was funded through the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), SFB1528 – Project C03; SL by Rett Elternhilfe e.V.; DZ and LP were supported by DFG 456967546. We would like to extend our sincere gratitude to all colleagues and experts of early vocal development who have discussed the idea of this manuscript with us and helped to further develop our approach.

**Author Contribution** Peter B Marschik, Claudius AA Widmann, and Sigrun Lang share first authorship. PBM provided the idea for study, designed and conceptualised it, and wrote the paper. CW, SL, and SB conducted the literature search and wrote the paper with DZ. DZ and SB co-supervised the literature search, conceptualisation, and writing. TK, KNS, SB, GE, ANH, HR, FW, CE, and LP collaborated with critically discussing the idea, refining the design, adding relevant literature, double checking search strategies, and writing of the study. All authors collaborated in the writing and editing of the final manuscript.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

## Declarations

**Conflict of Interest** All authors declare no direct conflict of interest related to this article. Bölte discloses that he has in the last 3 years acted as an author, consultant, or lecturer for Medice and Roche. He receives royalties for textbooks and diagnostic tools from Hogrefe and Liber. Marschik and Lang receive royalties from Elsevier, Springer, and Urban & Fischer. Bölte is shareholder in SB Education/Psychological Consulting AB and NeuroSupportSolutions International AB.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

All references marked with an \* are included in the review

- \*Apicella, F., Chericoni, N., Costanzo, V., Baldini, S., Billeci, L., Cohen, D., & Muratori, F. (2013). Reciprocity in interaction: A window on the first year of life in autism. *Autism Research and Treatment*, 2013, 705895. <https://doi.org/10.1155/2013/705895>.
- Bartl-Pokorny, K. D., Pokorny, F. B., Garrido, D., Schuller, B. W., Zhang, D., & Marschik, P. B. (2022). Vocalisation repertoire at the end of the first year of life: An exploratory comparison of Rett syndrome and typical development. *Journal of Developmental and Physical Disabilities*. <https://doi.org/10.1007/s10882-022-09837-w>
- Bauer, H. R., & Kent, R. D. (1987). Acoustic analyses of infant fricative and trill vocalizations. *The Journal of the Acoustical Society of America*, 81(2), 505–511. <https://doi.org/10.1121/1.394916>
- Bayley, N. (2006). *Bayley scales of infant and toddler development* (3rd ed.). Psychological Corporation.
- Belardi, K., Watson, L. R., Faldowski, R. A., Hazlett, H., Crais, E., Baranek, G. T., McComish, C., Patten, E., & Oller, D. K. (2017). A retrospective video analysis of canonical babbling and volubility in infants with fragile X syndrome at 9–12 months of age. *Journal of Autism and Developmental Disorders*, 47(4), 1193–1206. <https://doi.org/10.1007/s10803-017-3033-4>
- Borysiak, A., Hesse, V., Wermke, P., Hain, J., Robb, M., & Wermke, K. (2017). Fundamental frequency of crying in two-month-old boys and girls: Do sex hormones during mini-puberty mediate differences? *Journal of Voice*, 31(1), 128.e21–128.e28. <https://doi.org/10.1016/j.jvoice.2015.12.006>
- \*Brisson, J., Martel, K., Serres, J., Sirois, S., & Adrien, J. L. (2014). Acoustic analysis of oral productions of infants later diagnosed with autism and their mother. *Infant Mental Health Journal*, 35(3), 285–295. <https://doi.org/10.1002/imhj.21442>
- Buder, E. H., Chorna, L. B., Oller, D. K., & Robinson, R. B. (2008). Vibratory regime classification of infant phonation. *Journal of Voice: Official Journal of the Voice Foundation*, 22(5), 553–564. <https://doi.org/10.1016/j.jvoice.2006.12.009>
- Buder, E. H., Warlaumont, A. S., & Oller, D. K. (2013). An acoustic phonetic catalog of prespeech vocalizations from a developmental perspective. In B. Peter & A. A. N. MacLeod (Eds.), *Comprehensive perspectives on speech sound development and disorders: Pathways from linguistic theory to clinical practice* (pp. 103–134). Nova Publishers.
- \*Chericoni, N., de Brito Wanderley, D., Costanzo, V., Diniz-Goncalves, A., Leitgel Gille, M., Parlato, E., Cohen, D., Apicella, F., Calderoni, S., & Muratori, F. (2016). Pre-linguistic vocal trajectories at 6–18 months of age as early markers of autism. *Frontiers in Psychology*, 7, 1595. <https://doi.org/10.3389/fpsyg.2016.01595>.
- Cychosz, M., Cristia, A., Bergelson, E., Casillas, M., Baudet, G., Warlaumont, A. S., Scaff, C., Yankowitz, L., & Seidl, A. (2021). Vocal development in a large-scale crosslinguistic corpus. *Developmental Science*, 24(5), e13090. <https://doi.org/10.1111/desc.13090>
- D'Souza, H., & Karmiloff-Smith, A. (2017). Neurodevelopmental disorders. *WIREs. Cognitive Science*, 8(1–2), e1398. <https://doi.org/10.1002/wcs.1398>
- \*Ebrahimpour, M. K., Schneider, S., Noelle, D. C., & Kello, C. T. (2020). *InfantNet: A deep neural network for analyzing infant vocalizations*, arXiv preprint arXiv:2005.12412. <https://doi.org/10.48550/arXiv.2005.12412>
- Eilers, R. E., & Oller, D. K. (1994). Infant vocalizations and the early diagnosis of severe hearing impairment. *Journal of Pediatrics*, 124(2), 199–203. [https://doi.org/10.1016/s0022-3476\(94\)70303-5](https://doi.org/10.1016/s0022-3476(94)70303-5)
- \*Einspieler, C., Marschik, P. B., Domingues, W., Talisa, V. B., Bartl-Pokorny, K. D., Wolin, T., & Sigafos, J. (2014). Monozygotic twins with Rett syndrome: Phenotyping the first two years of life. *Journal of Developmental and Physical Disabilities*, 26(2), 171–182. <https://doi.org/10.1007/s10882-013-9351-3>
- Estes, A., St. John, T., & Dager, S. R. (2019). What to tell a parent who worries a young child has autism. *JAMA Psychiatry*, 76(10), 1092–1093. <https://doi.org/10.1001/jamapsychiatry.2019.1234>
- Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., André, E., Busso, C., Devillers, L. Y., Epps, J., Laukka, P., & Narayanan,



- S. S. (2015). The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. *IEEE Transactions on Affective Computing*, 7(2), 190–202. <https://doi.org/10.1109/TAFFC.2015.2457417>
- Fitch, W. T. (2010). The evolution of language. *Cambridge University Press*. <https://doi.org/10.1017/CBO9780511817779>
- Gipson, T. T., Ramsay, G., Ellison, E. E., Bene, E. R., Long, H. L., & Oller, D. K. (2021). Early vocal development in tuberous sclerosis complex. *Pediatric Neurology*, 125, 48–52. <https://doi.org/10.1101/2021.01.06.21249364>
- Hamrick, L. R., Seidl, A., & Tonnsen, B. L. (2019). Acoustic properties of early vocalizations in infants with fragile X syndrome. *Autism Research*, 12(11), 1663–1679. <https://doi.org/10.1002/aur.2176>
- Heideman, M. T., Johnson, D. H., & Burrus, C. S. (1985). Gauss and the history of the fast Fourier transform. *Archive for History of Exact Sciences*, 34(3), 265–277.
- Jhang, Y., & Oller, D. K. (2017). Emergence of functional flexibility in infant vocalizations of the first 3 months. *Frontiers in Psychology*, 8, 300. <https://www.frontiersin.org/article/https://doi.org/10.3389/fpsyg.2017.00300>
- Johnson, M. H. (2000). Functional brain development in infants: Elements of an interactive specialization framework. *Child Development*, 71(1), 75–81. <https://doi.org/10.1111/1467-8624.00120>
- Johnson, M. H., Charman, T., Pickles, A., & Jones, E. J. H. (2021). Annual research review: Anterior modifiers in the emergence of neurodevelopmental disorders (AMEND)—A systems neuroscience approach to common developmental disorders. *Journal of Child Psychology and Psychiatry*, 62(5), 610–630. <https://doi.org/10.1111/jcpp.13372>
- Jones, E. J. H., Gliga, T., Bedford, R., Charman, T., & Johnson, M. H. (2014). Developmental pathways to autism: A review of prospective studies of infants at risk. *Neuroscience & Biobehavioral Reviews*, 39, 1–33. <https://doi.org/10.1016/j.neubiorev.2013.12.001>
- Karmiloff-Smith, A. (2009). Nativism versus neuroconstructivism: Rethinking the study of developmental disorders. *Developmental Psychology*, 45(1), 56–63. <https://doi.org/10.1037/a0014506>
- Karmiloff, K., & Karmiloff-Smith, A. (2002). *Pathways to language: From fetus to adolescent*. Harvard University Press.
- Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends in Cognitive Sciences*, 2(10), 389–398.
- Kent, R. D. (2021). Developmental functional modules in infant vocalizations. *Journal of Speech, Language, and Hearing Research*, 64(5), 1581–1604. [https://doi.org/10.1044/2021\\_JSLHR-20-00703](https://doi.org/10.1044/2021_JSLHR-20-00703)
- Kent, R. D. (2022). The maturational gradient of infant vocalizations: Developmental stages and functional modules. *Infant Behavior and Development*, 66, 101682. <https://doi.org/10.1016/j.infbeh.2021.101682>
- Kent, R. D., & Murray, A. D. (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *The Journal of the Acoustical Society of America*, 72(2), 353–365. <https://doi.org/10.1121/1.388089>
- Koopmans-van Beinum, F. J., Clement, C. J., & van den Dikkenberg-Pot, I. (2001). Babbling and the lack of auditory speech perception: A matter of coordination? *Developmental Science*, 4(1), 61–70. <https://doi.org/10.1111/1467-7687.00149>
- Koopmans-van Beinum, F. J., & van der Stelt, J. M. (1986). Early stages in the development of speech movements. In B. Lindblom & R. Zetterström (Eds.), *Precursors of early speech* (pp. 37–50). Stockton.
- Lang, S., Bartl-Pokorny, K. D., Pokorny, F. B., Garrido, D., Mani, N., Fox-Boyer, A. V., Zhang, D., & Marschik, P. B. (2019). Canonical babbling: A marker for earlier identification of late detected developmental disorders? *Current Developmental Disorders Reports*, 6(3), 111–118. <https://doi.org/10.1007/s40474-019-00166-w>
- Lang, S., Willmes, K., Marschik, P. B., Zhang, D., & Fox-Boyer, A. (2021). Prelexical phonetic and early lexical development in German-acquiring infants: Canonical babbling and first spoken words. *Clinical Linguistics & Phonetics*, 35(2), 185–200. <https://doi.org/10.1080/02699206.2020.1731606>
- \*Legerstee, M., Bowman, T. G., & Fels, S. (1992). People and objects affect the quality of vocalizations in infants with Down syndrome. *Early Development and Parenting*, 1(3), 149–156. <https://doi.org/10.1002/edp.2430010304>
- Levin, K. (1999). Babbling in infants with cerebral palsy. *Clinical Linguistics & Phonetics*, 13(4), 249–267. <https://doi.org/10.1080/026992099299077>
- \*Li, J., Hasegawa-Johnson, M., & McElwain, N. L. (2021). Analysis of acoustic and voice quality features for the classification of infant and mother vocalizations. *Speech Communication*, 133, 41–61. <https://doi.org/10.1016/j.specom.2021.07.010>
- Lieberman, D. E., McCarthy, R. C., Hiemea, K. M., & Palmer, J. B. (2001). Ontogeny of postnatal hyoid and larynx descent in humans. *Archives of Oral Biology*, 46(2), 117–128. [https://doi.org/10.1016/S0003-9969\(00\)00108-4](https://doi.org/10.1016/S0003-9969(00)00108-4)
- Locke, J. L. (1995). *The child's path to spoken language*. Harvard University Press.
- Lohmander, A., Holm, K., Eriksson, S., & Lieberman, M. (2017). Observation method identifies that a lack of canonical babbling can indicate future speech and language problems. *Acta Paediatrica*, 106(6), 935–943. <https://doi.org/10.1111/apa.13816>
- \*Lynch, M. P., Oller, D. K., Steffens, M. L., & Buder, E. H. (1995a). Phrasing in prelinguistic vocalizations. *Developmental Psychology*, 28(1), 3–25.
- \*Lynch, M. P., Oller, D. K., Steffens, M. L., Levine, S. L., Basinger, D. L., & Umbel, V. (1995b). Onset of speech-like vocalizations in infants with Down syndrome. *American Journal of Mental Retardation*, 100(1), 68–86.
- \*Maestro, S., Muratori, F., Barbieri, F., Casella, C., Cattaneo, V., Cavallaro, M. C., Cesari, A., Milone, A., Rizzo, L., Viglione, V., Stern, D. D., & Palacio-Espasa, F. (2001). Early behavioral development in autistic children: The first 2 years of life through home movies. *Psychopathology*, 34(3), 147–152. <https://doi.org/10.1159/000049298>
- \*Maestro, S., Muratori, F., Cavallaro, M. C., Pecini, C., Cesari, A., Paziente, A., Stern, D., Golse, B., & Palacio-Espasa, F. (2005). How young children treat objects and people: An empirical study of the first year of life in autism. *Child Psychiatry and Human Development*, 35(4), 383–396. <https://doi.org/10.1007/s10578-005-2695-x>
- \*Maestro, S., Muratori, F., Cavallaro, M. C., Pei, F., Stern, D., Golse, B., & Palacio-Espasa, F. (2002). Attentional skills during the first 6 months of age in autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, 41(10), 1239–1245. <https://doi.org/10.1097/00004583-200210000-00014>
- Mareschal, D. (2011). From NEOconstructivism to NEUROconstructivism. *Child Development Perspectives*, 5(3), 169–170. <https://doi.org/10.1111/j.1750-8606.2011.00185.x>
- \*Marschik, P. B., Einspieler, C., Oberle, A., Laccone, F., & Prechtel, H. F. (2009). Case report: Retracing atypical development: A preserved speech variant of Rett syndrome. *Journal of Autism and Developmental Disorders*, 39(6), 958–961. <https://doi.org/10.1007/s10803-009-0703-x>
- Marschik, P. B., Einspieler, C., & Sigafos, J. (2012a). Contributing to the early detection of Rett syndrome: The potential role of auditory Gestalt perception. *Research in Developmental Disabilities*, 33(2), 461–466. <https://doi.org/10.1016/j.ridd.2011.10.007>
- Marschik, P. B., Pini, G., Bartl-Pokorny, K. D., Duckworth, M., Gugatschka, M., Vollmann, R., Zappella, M., & Einspieler, C. (2012b). Early speech-language development in females with Rett syndrome: Focusing on the preserved speech variant. *Developmental Medicine and Child Neurology*, 54(5), 451–456. <https://doi.org/10.1111/j.1469-8749.2012.04123.x>

- \*Marschik, P. B., Kaufmann, W. E., Sigafos, J., Wolin, T., Zhang, D., Bartl-Pokorny, K. D., Pini, G., Zappella, M., Tager-Flusberg, H., & Einspieler, C. (2013). Changing the perspective on early development of Rett syndrome. *Research in Developmental Disabilities*, 34(4), 1236–1239. <https://doi.org/10.1016/j.ridd.2013.01.014>.
- Marschik, P. B., Bartl-Pokorny, K. D., Sigafos, J., Urlesberger, L., Pokorny, F., Didden, R., Einspieler, C., & Kaufmann, W. E. (2014a). Development of socio-communicative skills in 9- to 12-month-old individuals with fragile X syndrome. *Research in Developmental Disabilities*, 35(3), 597–602. <https://doi.org/10.1016/j.ridd.2014.01.004>
- Marschik, P. B., Bartl-Pokorny, K. D., Tager-Flusberg, H., Kaufmann, W. E., Pokorny, F., Grossmann, T., Windpassinger, C., Petek, E., & Einspieler, C. (2014b). Three different profiles: Early socio-communicative capacities in typical Rett syndrome, the preserved speech variant and normal development. *Developmental Neurorehabilitation*, 17(1), 34–38. <https://doi.org/10.3109/17518423.2013.837537>
- Marschik, P. B., Pokorny, F. B., Peharz, R., Zhang, D., O'Muircheartaigh, J., Roeyers, H., Bolte, S., Spittle, A. J., Urlesberger, B., Schuller, B., Poustka, L., Ozonoff, S., Pernkopf, F., Pock, T., Tammimies, K., Enzinger, C., Kriebler, M., Tomantschger, I., Bartl-Pokorny, K. D., Bee-Pri Study Group. (2017). A novel way to measure and predict development: A heuristic approach to facilitate the early detection of neurodevelopmental disorders. *Current Neurology and Neuroscience Reports*, 17(5), 43. <https://doi.org/10.1007/s11910-017-0748-8>.
- Masataka, N. (2001). Why early linguistic milestones are delayed in children with Williams syndrome: Late onset of hand banging as a possible rate-limiting constraint on the emergence of canonical babbling. *Developmental Science*, 4, 158–164. <https://doi.org/10.1111/1467-7687.00161>
- Micai, M., Fulceri, F., Caruso, A., Guzzetta, A., Gila, L., & Scattoni, M. L. (2020). Early behavioral markers for neurodevelopmental disorders in the first 3 years of life: An overview of systematic reviews. *Neuroscience & Biobehavioral Reviews*, 116, 183–201. <https://doi.org/10.1016/j.neubiorev.2020.06.027>
- Moeller, M. P., Hoover, B., Putman, C., Arbataitis, K., Bohnenkamp, G., Peterson, B., Lewis, D., Estee, S., Pittman, A., & Stelmachowicz, P. (2007). Vocalizations of infants with hearing loss compared with infants with normal hearing: Part II—transition to words. *Ear and Hearing*, 28(5), 628–642. <https://doi.org/10.1097/AUD.0b013e31812564c9>
- Molemans, I., van den Berg, R., van Severen, L., & Gillis, S. (2012). How to measure the onset of babbling reliably? *Journal of Child Language*, 39(3), 523–552. <https://doi.org/10.1017/S0305000911000171>
- Morgan, L., & Wren, Y. E. (2018). A systematic review of the literature on early vocalizations and babbling patterns in young children. *Communication Disorders Quarterly*, 40(1), 3–14. <https://doi.org/10.1177/1525740118760215>
- Nathani Iyer, S., & Oller, D. K. (2008). Prelinguistic vocal development in infants with typical hearing and infants with severe-to-profound hearing loss. *Volta Review*, 108(2), 115–138.
- Nathani, S., Ertmer, D. J., & Stark, R. E. (2006). Assessing vocal development in infants and toddlers. *Clinical Linguistics & Phonetics*, 20(5), 351–369. <https://doi.org/10.1080/02699200500211451>
- Nathani, S., & Oller, D. K. (2001). Beyond ba-ba and gu-gu: Challenges and strategies in coding infant vocalizations. *Behavior Research Methods, Instruments, & Computers*, 33(3), 321–330. <https://doi.org/10.3758/bf03195385>
- Nyman, A., & Lohmander, A. (2018). Babbling in children with neurodevelopmental disability and validity of a simplified way of measuring canonical babbling ratio. *Clinical Linguistics & Phonetics*, 32(2), 114–127. <https://doi.org/10.1080/02699206.2017.1320588>
- Oller, D. K. (1978). Infant vocalization and the development of speech. *Allied Health and Behavioral Sciences*, 1(4), 523–549.
- Oller, D. K. (2000). *The emergence of the speech capacity*. Lawrence Erlbaum Associates.
- Oller, D. K., Buder, E. H., Ramsdell, H. L., Warlaumont, A. S., Chorna, L., & Bakeman, R. (2013). Functional flexibility of infant vocalization and the emergence of language. *Proceedings of the National Academy of Sciences*, 110(16), 6318–6323. <https://doi.org/10.1073/pnas.1300337110>
- Oller, D. K., Caskey, M., Yoo, H., Bene, E. R., Jhang, Y., Lee, C.-C., Bowman, D. D., Long, H. L., Buder, E. H., & Vohr, B. (2019). Preterm and full term infant vocalization and the origin of language. *Scientific Reports*, 9(1), 1–10. <https://doi.org/10.1038/s41598-019-51352-0>
- Oller, D. K., Eilers, R. E., Neal, A. R., & Cobo-Lewis, A. B. (1998). Late onset canonical babbling: A possible early marker of abnormal development. *American Journal of Mental Retardation*, 103(3), 249–263. [https://doi.org/10.1352/0895-8017\(1998\)103%3c0249:LOCBAP%3e2.0.CO;2](https://doi.org/10.1352/0895-8017(1998)103%3c0249:LOCBAP%3e2.0.CO;2)
- Oller, D. K., Eilers, R. E., Neal, A. R., & Schwartz, H. K. (1999). Precursors to speech in infancy: The prediction of speech and language disorders. *Journal of Communication Disorders*, 32(4), 223–245. [https://doi.org/10.1016/s0021-9924\(99\)00013-1](https://doi.org/10.1016/s0021-9924(99)00013-1)
- Oller, D. K., Niyogi, P., Gray, S., Richards, J. A., Gilkerson, J., Xu, D., Yapanel, U., & Warren, S. (2010). Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Biological Sciences*, 107, 13354–13359. <https://doi.org/10.1073/pnas.1003882107>
- \*Onnivello, S., Schworer, E. K., Daunhauer, L. A., & Fidler, D. J. (2021). Acquisition of cognitive and communication milestones in infants with Down syndrome. *Journal of Intellectual Disability Research*, jir.12893. <https://doi.org/10.1111/jir.12893>.
- \*Ouss, L., Palestra, G., Saint-Georges, C., Leitgel Gille, M., Afshar, M., Pellerin, H., Bailly, K., Chetouani, M., Robel, L., Golse, B., Nabbout, R., Desguerre, I., Guergova-Kuras, M., & Cohen, D. (2020). Behavior and interaction imaging at 9 months of age predict autism/intellectual disability in high-risk infants with West syndrome. *Translational Psychiatry*, 10(1), 1–7. <https://doi.org/10.1038/s41398-020-0743-8>.
- \*Pansy, J., Barones, C., Urlesberger, B., Pokorny, F. B., Bartl-Pokorny, K. D., Verheyen, S., Marschik, P. B., & Einspieler, C. (2019). Early motor and pre-linguistic verbal development in Prader-Willi syndrome—A case report. *Research in Developmental Disabilities*, 88, 16–21. <https://doi.org/10.1016/j.ridd.2019.01.012>.
- Papoušek, M. (1994). *Vom ersten Schrei zum ersten Wort. Anfänge der Sprachentwicklung in der vorsprachlichen Kommunikation*. Hans Huber.
- Patten, E., Belardi, K., Baranek, G. T., Watson, L. R., Labban, J. D., & Oller, D. K. (2014). Vocal patterns in infants with autism spectrum disorder: Canonical babbling status and vocalization frequency. *Journal of Autism and Developmental Disorders*, 44(10), 2413–2428. <https://doi.org/10.1007/s10803-014-2047-4>
- Paul, R., Fuerst, Y., Ramsay, G., Chawarska, K., & Klin, A. (2011). Out of the mouths of babes: Vocal production in infant siblings of children with ASD. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 52(5), 588–598. <https://doi.org/10.1111/j.1469-7610.2010.02332.x>
- Poeppl, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience*, 21(6), 322–334. <https://doi.org/10.1038/s41583-020-0304-4>
- \*Pokorny, F. B., Bartl-Pokorny, K. D., Einspieler, C., Zhang, D., Vollmann, R., Bölte, S., Gugatschka, M., Schuller, B. W., & Marschik, P. B. (2018). Typical vs. atypical: Combining auditory Gestalt perception and acoustic analysis of early vocalisations in Rett syndrome. *Research in Developmental Disabilities*, 82, 109–119. <https://doi.org/10.1016/j.ridd.2018.02.019>.
- \*Pokorny, F. B., Bartl-Pokorny, K. D., Zhang, D., Marschik, P. B., Schuller, D., & Schuller, B. W. (2020). Efficient collection and

- representation of preverbal data in typical and atypical development. *Journal of Nonverbal Behavior*, 44(4), 419–436. <https://doi.org/10.1007/s10919-020-00332-4>.
- \*Pokorny, F. B., Marschik, P. B., Einspieler, C., & Schuller, B. W. (2016a). Does she speak RTT? Towards an earlier identification of Rett Syndrome through intelligent pre-linguistic vocalisation analysis. *Proceedings of Interspeech 2016*. 1953–1957. <https://doi.org/10.21437/Interspeech.2016-520>.
- \*Pokorny, F. B., Peharz, R., Roth, W., Zöhrer, M., Pernkopf, F., Marschik, P. B., & Schuller, B. (2016b). Manual versus automated: The challenging routine of infant vocalisation segmentation in home videos to study neuro(mal)development. *Proceedings of Interspeech 2016*, 2997–3001. <https://doi.org/10.21437/Interspeech.2016-1341>.
- \*Pokorny, F. B., Schuller, B. W., Marschik, P. B., Brueckner, R., Nyström, P., Cummins, N., Bölte, S., Einspieler, C., & Falck-Ytter, T. (2017). Earlier identification of children with autism spectrum disorder: An automatic vocalisation-based approach. *Proceedings of Interspeech 2017*, 309–313. <https://doi.org/10.21437/Interspeech.2017-1007>.
- Pokorny, F. B., Schmitt, M., Egger, M., Bartl-Pokorny, K. D., Zhang, D., Schuller, B. W., & Marschik, P. B. (2022). Automatic vocalisation-based detection of fragile X syndrome and Rett syndrome. *Scientific Reports*, 12(1), 1–13. <https://doi.org/10.1038/s41598-022-17203-1>
- Roche, L., Zhang, D., Bartl-Pokorny, K. D., Pokorny, F. B., Schuller, B. W., Esposito, G., Bolte, S., Roeyers, H., Poustka, L., Gugatschka, M., Waddington, H., Vollmann, R., Einspieler, C., & Marschik, P. B. (2018). Early vocal development in autism spectrum disorder, Rett syndrome, and fragile X syndrome: Insights from studies using retrospective video analysis. *Advances in Neurodevelopmental Disorders*, 2(1), 49–61. <https://doi.org/10.1007/s41252-017-0051-3>
- Roug, L., Landberg, I., & Lundberg, L. J. (1989). Phonetic development in early infancy: A study of four Swedish children during the first eighteen months of life. *Journal of Child Language*, 16(1), 19–40. <https://doi.org/10.1017/s0305000900013416>
- Schuller, B. W., & Batliner, A. M. (2013). Computational paralinguistics: Emotion, affect and personality in speech and language processing. *John Wiley & Sons Ltd*. <https://doi.org/10.1002/9781118706664>
- Schuller, B. W., Steidl, S., Batliner, A., Vinciarelli, A., Scherer, K., Ringeval, F., Chetouani, M., Weninger, F., Eyben, F., Marchi, E., Mortillaro, M., Salamin, H., Polychroniou, A., Valente, F., & Kim, S. (2013). Computational paralinguistics challenge: Social signals, conflict, emotion, autism. *Proceedings of Interspeech, 2013*, 148–152.
- Shehata-Dieler, W., Ehrmann-Mueller, D., Wermke, P., Voit, V., Cebulla, M., & Wermke, K. (2013). Pre-speech diagnosis in hearing-impaired infants: How auditory experience affects early vocal development. *Speech, Language and Hearing*, 16(2), 99–106. <https://doi.org/10.1179/2050571x13z.0000000011>
- \*Smith, B. L., & Oller, D. K. (1981). A comparative study of pre-meaningful vocalizations produced by normally developing and Down's syndrome infants. *Journal of Speech and Hearing Disorders*, 46(1), 46–51. <https://doi.org/10.1044/jshd.4601.46>.
- \*Sohner, L., & Mitchell, P. (1991). Phonatory and phonetic characteristics of prelinguistic vocal development in cri du chat syndrome. *Journal of Communication Disorders*, 24(1), 13–20. [https://doi.org/10.1016/0021-9924\(91\)90030-M](https://doi.org/10.1016/0021-9924(91)90030-M).
- Stark, R. E. (1980). Stages of speech development in the first year of life. In G. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child Phonology* (Vol. 1, pp. 73–90). Academic Press.
- \*Steffens, M. L., Oller, D. K., Lynch, M., & Urbano, R. C. (1992). Vocal development in infants with Down syndrome and infants who are developing normally. *American Journal of Mental Retardation*, 97(2), 235–246.
- Vihman, M. M., Macken, M. A., Miller, R., Simmons, H., & Miller, J. (1985). From babbling to speech: A re-assessment of the continuity issue. *Language*, 61(2), 397–445.
- von Hapsburg, D., & Davis, B. L. (2006). Auditory sensitivity and the prelinguistic vocalizations of early-amplified infants. *Journal of Speech, Language, and Hearing Research*, 49(4), 809–822. [https://doi.org/10.1044/1092-4388\(2006\)057](https://doi.org/10.1044/1092-4388(2006)057)
- \*Warlaumont, A. S., Oller, D. K., Buder, E. H., Dale, R., & Kozma, R. (2010). Data-driven automated acoustic analysis of human infant vocalizations using neural network tools. *The Journal of the Acoustical Society of America*, 127(4), 2563–2577. <https://doi.org/10.1121/1.3327460>.
- Wermke, K., & Robb, M. P. (2010). Fundamental frequency of neonatal crying: Does body size matter? *Journal of Voice*, 24(4), 388–394. <https://doi.org/10.1016/j.jvoice.2008.11.002>
- Westermann, G., Thomas, M. S. C., & Karmiloff-Smith, A. (2011). Neuroconstructivism. In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed., pp. 723–747). Wiley-Blackwell.
- Yankowitz, L. D., Schultz, R. T., & Parish-Morris, J. (2019). Pre- and paralinguistic vocal production in ASD: Birth through school age. *Current Psychiatry Reports*, 21(12), 126. <https://doi.org/10.1007/s11920-019-1113-1>
- Yankowitz, L. D., Petrulla, V., Plate, S., Tunc, B., Guthrie, W., Meera, S. S., Tena, K., Pandey, J., Swanson, M. R., Pruett, J. R., Cola, M., Russel, A., Marrus, N., Hazlett, H. C., Botteron, K., Constantino, J. N., Dager, S. R., Estes, A., Zwaigenbaum, L., ... Network, T. I. B. I. S. (2022). Infants later diagnosed with autism have lower canonical babbling ratios in the first year of life. *Molecular Autism*, 13, 28. <https://doi.org/10.1186/s13229-022-00503-8>
- \*Zappella, M., Einspieler, C., Bartl-Pokorny, K. D., Kriebler, M., Coleman, M., Bölte, S., & Marschik, P. B. (2015). What do home videos tell us about early motor and socio-communicative behaviours in children with autistic features during the second year of life—An exploratory study. *Early Human Development*, 91(10), 569–575. <https://doi.org/10.1016/j.earlhumdev.2015.07.006>.
- Zhang, Y. S., & Ghazanfar, A. A. (2020). A hierarchy of autonomous systems for vocal production. *Trends in Neurosciences*, 43(2), 115–126. <https://doi.org/10.1016/j.tins.2019.12.006>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

**Peter B. Marschik<sup>1,2</sup> · Claudius A. A. Widmann<sup>1</sup> · Sigrun Lang<sup>1</sup> · Tomas Kulvicius<sup>1</sup> · Sofie Boterberg<sup>3</sup> · Karin Nielsen-Saines<sup>4</sup> · Sven Bölte<sup>5,6</sup> · Gianluca Esposito<sup>7</sup> · Anders Nordahl-Hansen<sup>8</sup> · Herbert Roeyers<sup>3</sup> · Florentin Wörgötter<sup>9,10</sup> · Christa Einspieler<sup>2</sup> · Luise Poustka<sup>1</sup> · Dajie Zhang<sup>1,2</sup>**

<sup>1</sup> Child and Adolescent Psychiatry and Psychotherapy, Göttingen, Germany and Leibniz ScienceCampus Primate Cognition, University Medical Center Göttingen, Göttingen, Germany

<sup>2</sup> iDN - Interdisciplinary Developmental Neuroscience, Division of Phoniatics, Medical University of Graz, Graz, Austria

<sup>3</sup> Research in Developmental Disorders Lab, Department of Experimental Clinical and Health Psychology, Faculty of Psychology and Educational Sciences, Ghent University, Ghent, Belgium

<sup>4</sup> Department of Pediatrics, David Geffen UCLA School of Medicine, Los Angeles, CA, USA

<sup>5</sup> Center of Neurodevelopmental Disorders (KIND), Centre for Psychiatry Research, Department of Women's and Children's Health, Child and Adolescent Psychiatry,

Region Stockholm, Karolinska Institutet & Stockholm Health Care Services, Stockholm, Sweden

<sup>6</sup> Curtin Autism Research Group, Curtin School of Allied Health, Curtin University, Perth, WA, Australia

<sup>7</sup> Affiliative Behavior and Physiology Lab, Department of Psychology and Cognitive Science, University of Trento, Trento, Italy

<sup>8</sup> Department of Education, ICT and Learning, Østfold University College, Halden, Norway

<sup>9</sup> Third Institute of Physics-Biophysics, Georg-August University Göttingen, Göttingen, Germany

<sup>10</sup> Bernstein Center for Computational Neuroscience Göttingen, Göttingen, Germany