ASHA American Speech-Language-Hearing Association

AJSLP

Research Article

Speech Naturalness in the Assessment of Childhood Dysarthria

Theresa Schölderle,[a] Elisabet Haas,[a] and Wolfram Ziegler[a]

[a] Institute for Phonetics and Speech Processing, Ludwig-Maximilians-Universität München, München, Germany

ARTICLE INFO

Article History:
Received January 23, 2023
Revision received March 23, 2023
Accepted April 7, 2023

Editor-in-Chief: Katherine C. Hustad
Editor: Ignatius Nip

https://doi.org/10.1044/2023_AJSLP-23-00023

ABSTRACT

Purpose: This study investigated perceived speech naturalness estimated by adult listeners in typically developing children and children with dysarthria. We aimed to identify predictors of naturalness among auditory-perceptual parameters and to evaluate the concept of naturalness as a clinical marker of childhood dysarthria.

Method: In a listening experiment, naive adult listeners rated speech naturalness of 144 typically developing children (3–9 years old) and 28 children with neurological conditions (5–9 years old) on a visual analog scale. Speech samples were recorded using the materials of the Bogenhausen Dysarthria Scales–Childhood Dysarthria, which also provides for auditory-perceptual judgments covering all speech subsystems.

Results: Children with dysarthria obtained significantly lower naturalness ratings compared to typically developing children. However, there was a substantial age effect observable in the typically developing children; that is, younger typically developing children were also perceived as somewhat unnatural. The ratings of the typically developing children were influenced by the occurrence of developmental speech features; for the children with neurological conditions, specific symptoms of dysarthria had an additional effect. In both groups, the perception of naturalness was predominantly determined by the children's articulation and intelligibility.

Conclusions: Both symptoms of childhood dysarthria and developmental speech features (e.g., regarding articulation and intelligibility) were associated to some extent with unnatural speech by the listeners. Thus, perceived speech naturalness appears less suitable as a marker of dysarthria in children than in adults.

Dysarthria is a speech disorder that can be caused by various neurological conditions and may affect adults (e.g., with stroke or Parkinson's disease) as well as children (e.g., with cerebral palsy [CP] or genetic syndromes). By definition, dysarthria is characterized by impairments of the speech motor subsystems, that is, respiration, phonation, articulation, and prosody (Duffy, 2020). Clinical assessment is predominantly focused on the speech characteristics directly associated with these motor impairments (e.g., impaired voice quality, articulatory dysfunction), but for a comprehensive and ecologically valid description of the speech disorder, the manifold and highly individual consequences of the disorder for everyday communication must also be considered (International Classification of Functioning, Disability and Health; World Health Organization, 2001).

So far, research on communication-related parameters both in adults and children with dysarthria has predominantly focused on speech intelligibility. In adults, several studies also addressed perceived speech naturalness, showing that this concept may add relevant detail to the picture of communication abilities in dysarthria (Dagenais & Wilson, 2002; Dagenais et al., 2006; Klopfenstein et al., 2020; Schölderle et al., 2016; Southwood & Weismer, 1993; Yorkston et al., 1990). Speech naturalness refers to a rather broad perceptual impression representing the

Correspondence to Theresa Schölderle: theresa.schoelderle@ekn-muenchen.de. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

overall quality of a person's speech output in relation to what is conceptualized as normal or natural (Klopfenstein, 2015; Yorkston et al., 1990). Studies have shown that naturalness of speech is rated significantly lower in adults with dysarthria when compared to nondysarthric speakers (Dagenais et al., 2006; Whitehill et al., 2004). In a study by Lehner and Ziegler (2021), for instance, naturalness ratings separated adults with dysarthria from neurologically healthy speakers with an accuracy of 0.94. The deviation from the norm, which is equated with unnatural speech, can confuse interlocutors about a speaker's health condition or other personal issues and thus severely impair communication. In addition, unnatural speech may distract from what is being said and, in this way, contribute to an increased listener effort (Lehner et al., 2022). In this regard, naturalness represents a clinically important outcome measure for adult speakers with dysarthria.

Focusing on specific speech characteristics, perceived naturalness in adults with dysarthria has been shown to be influenced by the speakers' articulatory dysfunctions, but also by symptoms affecting respiration, phonation, and prosody (Dagenais et al., 2006; Klopfenstein, 2015; Klopfenstein et al., 2020; Lehner et al., 2022; Schölderle et al., 2016). In a study on individuals with Parkinson's disease, for instance, Klopfenstein (2015) identified various acoustic predictors of naturalness ratings, such as mean fundamental frequency, intensity range, articulation rate, and average syllable duration. Moreover, a previous auditory-perceptual study of our own revealed the relevance of imprecise articulation, inadequate inspirations, and monotonous speech for naturalness ratings of speech recorded from adults with CP (Schölderle et al., 2016). This study also emphasized the relevance of intelligibility for naturalness ratings.

Similar to adults, research on communication-related parameters in childhood dysarthria has primarily focused on intelligibility (Cahill et al., 2002; Haas et al., 2022; Hustad et al., 2012; Lee et al., 2014). However, the effects of speech dysfunctions on communication are not limited to intelligibility, and even children with preserved intelligibility can experience dysarthria-related communication difficulties (Pennington, 1999). Yet, speech naturalness in children with dysarthria has been addressed in only two studies so far (Jones et al., 2019; Moya-Galé et al., 2021), though it was not in the main focus of these investigations. That is, Moya-Galé et al. (2021) used speech naturalness as one of many outcome measures in a treatment design. In the work of Jones et al. (2019), on the other hand, speech-language therapists (SLTs) rated naturalness as one of 47 perceptual features in children with Down syndrome only. Given this sparse data basis, major questions remain open. For instance, little is known about speech naturalness in one of the most important pediatric groups with dysarthria, that is, children with CP (Mei et al., 2014). Moreover, there are several studies addressing the question of which dysarthric symptoms have the greatest impact on intelligibility (DuHadway & Hustad, 2012; Haas et al., 2022; Lee et al., 2014). With regard to naturalness, the same question has not been focused on yet. It remains unclear whether similar results can be expected in children as in adults.

In this context, however, it seems reasonable to take a step back and ask to what extent the assessment of perceived speech naturalness can and should be transferred from adults to children with dysarthria in the first place. In adults with dysarthria, speech symptoms leading to the impression of unnatural speech possess a high diagnostic specificity, as they can almost unequivocally be interpreted as (direct) results of the brain damage. In Lehner and Ziegler (2021), for example, only two out of 54 control speakers (< 5%) had naturalness ratings below the cutoff separating neurotypical speakers from adult patients with dysarthria; that is, the parameter naturalness had a specificity for dysarthria of 0.95.

Yet, this clear causal interpretation may not necessarily hold in children. Though the speech of children with dysarthria is characterized by specific markers that may make them sound unnatural, such as a strained–strangled voice or hypernasality (Schölderle et al., 2022), children's speech is also influenced by developmental processes, whose impact on naturalness ratings is still unknown. On a general account, listeners' internal models of undisturbed speech in children may be more variable as compared to adults, due to the large changes in speech taking place over the course of childhood. More specifically, however, developmental speech characteristics like a breathy voice, imprecise articulation, or limited intelligibility (Hustad et al., 2021; Schölderle et al., 2020, 2021) may overlap with specific markers of dysarthria in children and thereby influence listeners' assessment of how naturally a child speaks. Thus, one cannot exclude that listeners, especially when they are inexperienced with children's speech, interpret not only symptoms of dysarthria but also developmental characteristics as "unnatural," which may blur the distinction between typical and dysarthric speech in diagnostic applications of the naturalness criterion.

In order to evaluate the potential usefulness of naturalness ratings in assessing the communication-relevant consequences of childhood dysarthria, studies of typically developing children might provide valuable insights. The only study that has addressed naturalness of speech in typically developing children by Coughlin-Woods et al. (2005) demonstrated that rating scores differed between 8-year-olds, on the one hand, and 10- to 16-year-old children, on the other. For the group of 8-year-old children, the largest variability was found, and naturalness increased across the age range between 8 and 14 years. The finding of an age effect on naturalness, that is, lower naturalness scores in young children, may challenge the application of the concept of naturalness in children fundamentally. The interpretation that "natural" is a proxy for "normal," "undisturbed," or "unimpaired," whereas "unnatural" suggests the presence of some disorder, would no longer be valid in this case. Therefore, it is important to replicate this result and expand the analysis to children younger than 8 years, to see if their speech is scored even less natural. Moreover, if typically developing children are perceived as unnatural within a certain age range, it would be important to identify the features that render their speech unnatural, such as certain voice or prosodic characteristics, or reduced intelligibility (e.g., Schölderle et al., 2016). Ultimately, it needs to be investigated if naturalness ratings differ between typically developing children and children with dysarthria, that is, if the concept holds clinical value for the description of childhood dysarthria, and to determine the sensitivity and specificity of naturalness ratings in distinguishing between typically developing children and children with dysarthria. Also, from a clinical point of view, it will be important to know if naturalness can be predicted by similar speech parameters in children with dysarthria and typically developing children, or if specific markers of childhood dysarthria play a more important role. Previous research on naturalness in children with dysarthria neither addressed developmental impacts nor compared children with dysarthria to typically developing peers systematically (Jones et al., 2019; Moya-Galé et al., 2021).

The study presented here therefore aimed (a) to analyze naturalness ratings in typically developing children of a wide age range for a possible age effect, (b) to compare perceived naturalness between typically developing children and children with dysarthria in order to estimate the clinical value of the concept, and (c) to identify predictors of naturalness at the level of motor speech subsystems, including intelligibility, both for the typically developing children and the children with dysarthria.

Method

This study was approved by the Medical Faculty's Ethics Committee of the Ludwig-Maximilians-Universität München. As there is some overlap with previous publications, details on participants, assessment materials, and auditory-perceptual analyses beyond the information provided herein can be obtained from Schölderle et al. (2020, 2021).

Participants

One hundred forty-four children with typical development (in the following referred to as CTD) between 3;0 and 9;11 (years;months) participated in the study. The group was stratified over the age range and balanced for gender. All children were native speakers of German living in different areas in southern Germany. Exclusion criteria ensured that no child had a diagnosis of a neurological disease, structural abnormalities of the speech apparatus (e.g., cleft palates), speech-language delay or disorder, or other developmental disorders (e.g., generalized cognitive or motor disorders) according to a parent questionnaire. The presence of potential language impairment was also excluded through administration of the Test for the Reception of Grammar (German version), a tool for assessing the comprehension of grammatical structures (Fox-Boyer, 2020).

Moreover, 28 children with neurological conditions (CNC) participated in the study (nine girls, 19 boys; age median [range]: 7;7 [5;1–9;10]). They were recruited in kindergartens and schools for children with multiple disabilities in southern Germany. Inclusion required that children (a) have a medical diagnosis of a neurological condition with a nonprogressive course (e.g., CP, genetic syndromes), (b) have no structural abnormalities affecting the speech apparatus (e.g., bronchial asthma, cleft palates), (c) have hearing abilities within normal limits, (d) are native speakers of German, and (e) are able to communicate verbally and repeat short sentences in an elicitation task (the use of auxiliary augmentative and alternative communication systems did not lead to exclusion). Due to criterion (e), we set the lower age limit of children participating at age 5;0. Since we aimed to apply previously collected norm data from typically developing children between 3 and 9 years (see Schölderle et al., 2020), we only included children within this range. Accordingly, for this study, children had to be between ages 5;0 and 9;11 (see Table 1).

The sample included children of different etiologies, with CP being the most prevalent condition. The severity of the children's gross motor dysfunctions varied considerably (according to Gross Motor Function Classification System), as did their communication abilities (according to Communication Function Classification System, Bogenhausen Dysarthria Scales for Childhood Dysarthria [BoDyS-KiD] total score, and intelligibility; see Table 1). Notably, the BoDyS-KiD total scores and the intelligibility scores (see section on Listening Experiment for details) show that a substantial proportion of the CNC sample had severe speech deficits and intelligibility restrictions, respectively. On the other hand, the age-normalized BoDyS-KiD total score (see Haas et al., 2021)

Table 1. Demographic and clinical data of the children with neurological conditions (CNC group).

Case	Age (y;m)	Gender	Neurologic etiology	GMFCS[a] (I–V)	CFCS[a] (I–V)	Normalized BoDyS-KiD total score[a]	% intelligibility[b]
CNC01	5;1	m	CP	V	IV	3.22	77.4
CNC02	5;8	f	CP	III	IV	2.20	82.4
CNC03	6;0	m	CP, generalized epilepsy in West Syndrome	II	IV	1.45	72.1
CNC04	6;5	m	CP	III	IV	4.18	0.00
CNC05	6;7	m	CP, pachygyria	I	III	0.17*	95.3
CNC06	6;9	m	CP, microcephaly	II	III	1.14	77.3
CNC07	7;1	m	CP	III	V	5.29	2.8
CNC08	7;4	m	Early infantile autism, developmental motor delay of unknown neurological origin	I	III	0.33*	54.7
CNC09	7;7	f	CP	II	I	–0.01*	100.0
CNC10	7;7	m	CP	I	IV	1.12	92.3
CNC11	7;7	m	CP	II	V	4.38	25.0
CNC12	8;1	f	CP	IV	IV	3.24	17.9
CNC13	8;3	m	CP	II	IV	2.88	93.9
CNC14	8;4	f	CP	I	III	2.20	28.0
CNC15	7;2	m	CP	III	IV	4.14	59.1
CNC16	8;11	f	CP	II	IV	4.34	6.2
CNC17	9;5	m	CP	III	IV	1.14	91.9
CNC18	8;11	m	Generalized muscular hypotony	I	IV	5.23	26.7
CNC19	9;5	m	CP	II	III	4.82	8.2
CNC20	7;8	f	CP	III	III	1.85	67.3
CNC21	7;0	m	CP	II	IV	1.54	94.4
CNC22	7;9	f	CP	V	III	3.57	44.4
CNC23	8;8	m	CP	II	III	1.04	86.6
CNC24	7;7	m	Focal epilepsy	II	IV	3.12	20.0
CNC25	8;0	m	CP	IV	IV	2.99	68.7
CNC26	9;9	f	CP	II	V	3.24	76.1
CNC27	9;10	m	CP	V	V	2.54	93.7
CNC28	6;9	f	CP	II	IV	1.99	66.8

Note. y;m = years;months; GMFCS = Gross Motor Function Classification System (Palisano et al., 1997); CFCS = Communication Function Classification System (Hidecker et al., 2011); BoDyS-KiD = Bogenhausen Dysarthria Scales–Childhood Dysarthria (Haas et al., 2021); m = male; CP = cerebral palsy; f = female.

[a]Higher scores indicate more severe impairment. [b]See Listening experiment for details.

*BoDyS-KiD total score within the respective age norm (i.e., < 1).

demonstrated that three children from the CNC group ranged within their age norm and, hence, were not classified as dysarthric (i.e., CNC05, CNC08, and CNC09). They were included as part of the CNC sample in the following analyses. However, we have specifically addressed these three cases in the first part of our results, which was focused on a group comparison between CNC and CTD.

Speech Samples

Standard speech samples were obtained using the protocol of the BoDyS-KiD (see Haas et al., 2021; Schölderle et al., 2020). This tool builds upon the BoDyS, proposed by Ziegler et al. (2018) as an approach established for adults with dysarthria. Even though BoDyS-KiD was conceptualized along criteria applicable to dysarthric speech, its comprehensive and systematic auditory-perceptual descriptions of speech motor functions provide a reasonable template to characterize typical speech development as well (see section on Auditory-Perceptual Analysis).

The application of the BoDyS-KiD materials allows for a standardized recording of repetition sentences and spontaneous speech samples within the framework of a child-friendly and motivating computer game. The stimulus sentences presented in the course of the computer game were embedded in a story. Both sentence stimuli and the framing story were prerecorded by a trained male speaker instructed to speak clearly and understandably, yet in a natural way. He was not encouraged to speak in a child-directed manner. Children were instructed to repeat the sentences. Whenever the instructor had the impression that the child was trying to imitate the model speaker's manner of speaking rather than repeating the sentence in their own natural manner, the child was reinstructed. Two versions of the computer game were developed, one for children ≥ 5 years of age (Version A) and one for children < 5 years of age (Version B), to account for their different levels of language development and working memory. Version A contained 12 sentences of six to 12 syllables, whereas Version B comprised nine sentences of six to 10 syllables. Also, the sentences in Version B were syntactically less complex, did not include subordinate clauses, and contained fewer multisyllabic words with complex syllable structures. The sentences of both versions comprised all German phonemes and varied in intonation patterns. Four sentences were identical between Versions A and B. Note that the selection of either version according to age (≥ 5 years – Version A; < 5 years – Version B) is applicable for typically developing children, predominantly. In the CNC group, versions were selected individually according to a child's linguistic and cognitive abilities.

Both versions provide that at 3 points in the assessment, the examiner asks the child-specific questions in order to elicit spontaneous speech samples complementing the repetition sentences. For example, when the frame story of the computer game addresses a character's family, the child is encouraged to tell something about their own family at that point.

Auditory-Perceptual Analyses

As stipulated by the BoDyS-KiD protocol, the speech samples (repetition sentences and spontaneous speech samples) from both the CTD and CNC groups were subject to comprehensive auditory-perceptual analyses following the diagnostic criteria of the original BoDyS (see Ziegler et al., 2018). Results have been published previously (Haas et al., 2021; Schölderle et al., 2020, 2022).

The BoDyS protocol provides for analyses at two scoring levels—the BoDyS scales and the BoDyS features. At the BoDyS scale level, all relevant speech subsystems—that is, speech breathing, voice production, articulation/resonance, and prosody—are covered by nine scales (i.e., respiration, voice level, voice quality, voice stability, articulation, resonance, articulation rate, fluency, and prosodic modulation). Each of the nine BoDyS scales comprises a number of BoDyS features representing the clinically most relevant dysarthria symptoms. For instance, the voice quality scale comprises the features breathy voice, strained–strangled voice, rough/harsh voice, and fluctuating voice quality. BoDyS-KiD consists of a total of 29 features (Schölderle et al., 2020).

All BoDyS scales and features were rated uniformly on a 5-point equal-appearing interval scale representing the salience of auditory-perceptual characteristics. For this purpose, an intuitive concept of an adult norm served as a reference (i.e., 0 = highly salient, i.e., very different from how a healthy adult person would speak; 4 = adultlike). Since it seemed unreasonable to relate a child's pitch to an adult norm (i.e., due to anatomic–physiologic conditions, all children have much higher voice levels than adults), the features "high pitch" and "low pitch," originally part of the BoDyS protocol, were removed from the voice-level scale and were not rated as part of the analyses. The steps described so far and the results of these analyses have all been elaborated on in the works of Schölderle et al. (2020, 2022), together with the corresponding information on listener agreement.

Listening Experiment

A comprehensive listening experiment involving naive adult listeners was conducted to assess several communication parameters (e.g., intelligibility, naturalness, as well as estimations of age and gender not reported here) in the CTD and the CNC groups. For this purpose, 240 adult listeners (120 female [f], 120 male [m]; median age = 26 years, range: 18–59) were recruited. All were native speakers of German, had self-reported normal-hearing abilities, and were recruited via postings on notice boards at LMU Munich, mailing lists for students of all departments at LMU Munich, and personal contacts. We included only adults who had no experience with the BoDyS-KiD materials and had no regular professional contact with children between the ages of 3 and 10 years (such as kindergarten or primary school teachers, pediatricians, etc.) or CNC.

For the assessment of intelligibility, listeners performed a sentence transcription task. The naturalness ratings were provided in the second part of the experiment. Since the BoDyS-KiD repetition sentences were used for both tasks, it was essential to adhere to this order to avoid familiarity effects on intelligibility.

Intelligibility. To further avoid familiarity with the speakers or materials during the course of the intelligibility experiment, each listener was presented with a maximum of 12 different sentences spoken by 12 different children including children from both the CTD and CNC groups. The sentences were presented via headphones at a comfortable loudness level. After each stimulus, listeners were instructed to transcribe the sentence orthographically and mark unintelligible segments by a dash. The transcripts were compared to master transcripts of the individual realizations of the target sentences. The proportion of correctly transcribed syllables across all sentences spoken by each child was then calculated as a measure of intelligibility. Results on intelligibility have already been published (see Schölderle et al. [2021] for the CTD and Haas et al. [2022] for the CNC). For more detailed descriptions of the applied procedures, we refer to these previous articles.

Naturalness. Listeners were presented the four sentences that were identical in Versions A and B. The sentences varied regarding intonation pattern (declarative vs. interrogative sentence) and length (six to 10 syllables). To keep an influence of intelligibility on the rating of naturalness as little as possible, all sentences were also provided in printed form. The sentences were embedded in listener-specific PowerPoint presentations. For each listener, the sentences were played in the same order to avoid order effects.

Each child was judged by a total of 20 listeners, and it was ensured that throughout the entire listening experiment, each listener heard each child only once across both the intelligibility and the naturalness experiment. To conduct this experimental setup, the 240 naive listeners were assigned to 12 groups of 20 listeners each. Listener groups were balanced for age and gender. Consequently, speaker sets of 20 children each were compiled that were balanced with regard to the age of the children and the ratio of CNC versus CTD. Within each speaker set, pseudorandomization of speaker order was performed for each listener.

Since the naturalness ratings always followed the intelligibility experiment, it was ensured that all listeners had some amount of exposure to different varieties of child speech (i.e., younger and older children, CNC of different severities, and CTD) prior to making naturalness ratings. To further minimize familiarization effects, the speech samples of the first five children were repeated at the end and the first ratings of these five children were discarded subsequently. Listeners were allowed to listen to the four sentences a maximum of 2 times per child.

Listeners rated naturalness on a visual analog scale provided with the two anchors “completely unnatural” and “natural.” They were instructed to imagine themselves sitting in a subway or walking past a playground, casually listening to a child speaking. They were not asked to focus on what was being said, but rather to pay attention to whether the child’s way of speaking seemed unnatural (striking, peculiar) in any way, or whether the way of speaking seemed completely natural or inconspicuous to them (the way they would usually imagine children speaking). They were encouraged to make intuitive and spontaneous judgments and not to hesitate to exploit the entire scale, including the two extremes. Re-instructions were given during the rating of the first five children, if necessary. Note that the listeners were not informed of the children’s age, regarding neither the age of each individual case nor the age range of the presented group. The ratings on the visual analog scales were completed on paper, hand-measured (with a millimeter resolution on scales with a length of 5 cm), and transformed into scores ranging from 0 (“completely unnatural”) to 5 (“natural”).

Consistency of naturalness ratings among listeners was determined for the 12 groups of 20 listeners separately. Cronbach’s alpha was used as a measure of between-listeners agreement across samples of 20 speakers in each cohort of speakers assessed by the same listener group (function *alpha* of the R-package *psych*; Revelle, 2022). To reduce the risk of overestimation of Cronbach’s alpha, consistency was analyzed for randomly sampled subgroups of six out of 20 listeners per group, with 1,000 bootstrapping samples. The resulting estimates of alpha ranged between .86 and .96 (.90 ± .03) across the 12 cohorts, indicating overall good internal consistency of naturalness ratings.

Results

Age Effect and Group Comparison

As for the CTD group, naturalness showed a high variability with average scores (across 20 listeners per child) ranging from 1.75 to 4.83 (M = 3.83, SD = 0.69). Figure 1 illustrates that this variability was partly ascribable to an age effect; that is, younger children tended to have lower naturalness scores.

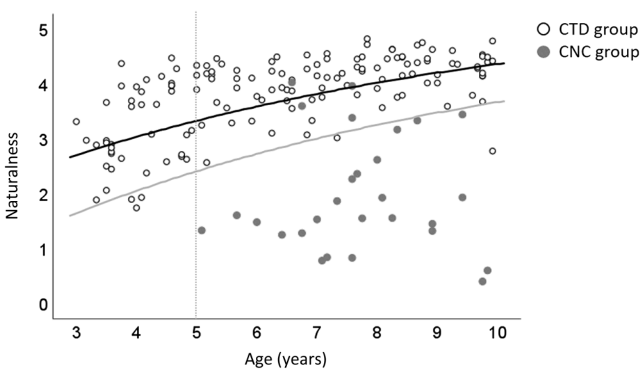
The development of naturalness scores over age was analyzed by a nonlinear exponential regression model, as established in two earlier studies (Schölderle et al., 2020, 2021).[1] In this model, age was entered as an independent variable in a negative exponential function, that is,

$$\text{nat} \sim 5 - \exp(a - b \times (\text{age} - 3.0)). \tag{1}$$

This model expresses the decay rate of the score’s distance from the maximally achievable score of 5 by a

[1]Note that a t test conducted with the CTD group’s raw data indicated that there was no significant difference between boys and girls (T142 = 1.29; p > .05). As a consequence, all data were collapsed; gender is not marked in Figure 1.

Figure 1. Naturalness ratings plotted against chronological age of the 144 CTD and 28 CNC speakers. Black solid line: median of the CTD group; gray solid line: 5th percentile of the CTD group; dashed vertical line: age boundary for CTD and CNC groups included in *t* test (see text). CTD = children with typical development; CNC = children with neurological conditions.



negative exponential law. Age was included in years, with a day-wise resolution (e.g., 5.027 years for a 5-year to 10-day-old child). As starting values in the regression analyses, the model coefficients a and b were set to 1.0 and 0.5, respectively. The intercept of the model was shifted to 3.0 years, since data were available from 3 years onward, and the approximation value was set to 5.0, that is, the maximum value of the naturalness scale.

In order to obtain age-dependent estimates of the central tendency and the lower boundary of the developmental trajectory of a parameter, exponential regression curves were computed for the median values and the 5th percentile of the naturalness scores. Gliding windows with a window size of 21 points, shifted by 1 point, were applied using the "runquantile" function (5th percentile) from the "caTools" package in R (Tuszynski, 2018; for further explanations, see Schölderle et al., 2020, 2021). Statistical computations were performed using the "nls" function from the R package "nlstools" (Baty et al., 2015) for least square nonlinear estimation.

The scores of the CNC group also covered a wide range (0.41–4.03) but were overall lower than the CTD group's naturalness scores (M = 1.99, SD = 1.06; see Figure 1). To normalize naturalness scores for age, each participant's raw score was subtracted from the median model estimate (black solid line in Figure 1) and divided by the distance between the median and the 5% exponential curves at the participant's age. For a direct comparison of the age-normalized naturalness scores of the two groups, we selected all CTD of the CNC's age range (i.e., all children from the CTD group between ages 5;1 and 9;10; see dashed vertical line in Figure 1). A t test comparing this CTD subsample (n = 102) with the CNC group (n = 28) showed a significant difference, $t(162.7) = 33.27$, $p < .001$.

In order to further elucidate how accurately the naturalness parameter separates between CTD and CNC, we calculated specificity and sensitivity coefficients based on the models plotted in Figure 1. Note that the 5% cut-score used in the normalization implies a model-based specificity of 0.95. As only 5/144 CTD fell below the 5% cutoff, the empirical specificity was 0.97. Four out of 28 children from the CNC group ranged above the CTD group's 5% threshold, corresponding to a sensitivity of 0.86. Among those four children were the three cases with a BoDyS-KiD total score < 1, that is, who did not have dysarthria according to their age norm.

Predictions of Naturalness Ratings by Speech Characteristics

In a further step, we aimed to identify the BoDyS scales that influenced the speech naturalness ratings provided by the listener group. To avoid the problem of overfitting in the CNC group (n = 28), we used the variable selection procedures implemented in the R package "olsrr" (Hebbali, 2020). In particular, stepwise forward regression based on the Akaike information criterion was applied in combination with the linear modeling function "lm" in R.

First, we determined the BoDyS variables that predicted the developmentally driven variation of naturalness scores in the CTD group (n = 144) by entering the nine BoDyS scales in a linear regression model of the CTD group's naturalness scores. The upper part of Table 2 reports the four variables that were included in the model in serial order, together with their corresponding model characteristics and the value of the Akaike information criterion.

The model explained ca. 57% of the variance in the naturalness scores of the CTD group (R^2_{adj} = .57). Inclusion of only the articulation and voice level scales yielded an R^2_{adj} value of .55, whereas the prosodic modulation and respiration scales added only a small margin of ca. 2%.

To determine the BoDyS predictors of naturalness scores in the CNC group, the selection of variables included in the regression model was limited to three to avoid overfitting. The lower part of Table 2 reports the best among all possible three-variable regression models, which yielded an overall fit of R^2_{adj} = .71. The largest contribution was made by the articulation scale (R^2_{adj} = .57), followed by articulation rate, and, with a small margin, resonance.

To finally analyze the relevance of intelligibility for the perception of naturalness, we ran two additional linear regressions for the CTD and the CNC groups, respectively, with intelligibility serving as the only predictor variable. In the CTD group, intelligibility explained 57% of

Table 2. Linear regression models of the response variable *naturalness* with the BoDyS scales as predictors.

Group	Variable	Akaike information criterion (AIC)	Sum of squares	Residual sum of squares (RSS)	R^2	Adjusted R^2
CTD (*n* = 144)	Articulation	201.82	36.05	32.85	.52	.52
	Voice level	191.73	38.56	30.23	.56	.55
	Prosodic modulation	189.61	39.42	29.40	.57	.56
	Respiration	188.44	40.06	28.76	.58	.57
CNC (*n* = 28)	Articulation	62.80	17.79	12.46	.58	.57
	Articulation rate	55.43	21.34	8.92	.71	.68
	Resonance	54.02	22.36	7.90	.74	.71

Note. Top: CTD group (*n* = 144). Bottom: CNC group (*n* = 28). Variable selection was performed using forward regression modeling based on Akaike information criterion (AIC). BoDyS = Bogenhausen Dysarthria Scales; CTD = children with typical development; CNC = children with neurological conditions.

the observed variance of naturalness (R^2 = .57, p < .001). In the CNC group, R^2 reached .66 (p < .001).

Discussion

This study aimed to investigate perceived speech naturalness in children with dysarthria in comparison to typically developing children and to identify predictors of naturalness among auditory-perceptual parameters. One of the key findings of this study was that naturalness was subject to an age effect in the typically developing children between 3 and 9 years old. Especially the younger children received lower naturalness ratings, but there was still considerable variability even in the 9-year-olds. In this context, it is important to recall that in our listening experiment, participants were always presented with both typically developing children and children with dysarthria. Hence, if listeners perceived typically developing children as unnatural, it cannot be due to a set-effect, that is, that they were unaware of the entire range of speech qualities across the CTD and CNC groups.

Our results complement the study conducted by Coughlin-Woods et al. (2005), who analyzed speech naturalness in typically developing children between 8 and 16 years old and also reported that naturalness increased with age until the age of 14 years. This age effect, which has now been shown to be even stronger in younger children, fundamentally calls into question the concept of speech naturalness as a clinical marker of dysarthria or of any speech disorder whatsoever in children. In adult populations with dysarthria, who were in the center of naturalness studies so far, we can assume that anything that contributes to a listener's impression of unnatural speech is predominantly related to their dysarthria. This has been confirmed through the finding that adults with dysarthria are reliably perceived as unnatural, whereas neurologically healthy speakers are not (Dagenais et al., 2006; Lehner & Ziegler, 2021; Whitehill et al., 2004). As a consequence, "natural" can be used as a proxy for "normal," "undisturbed," or "unimpaired" in adults, whereas "unnatural" has been understood as an indicator of a speech disorder (e.g., Schölderle et al., 2016). Our result that typically developing children were not always perceived as natural reveals that this implication is not valid in children.

The speech dimension that seemed to influence the listeners' assessment of the typically developing children's speech naturalness in the first place was the quality of articulation as measured by the articulation scale of BoDyS-KiD, followed by the voice level scale, which mainly represents the feature *low voice volume*. Both variables have shown major developmental dynamics (Schölderle et al., 2020). Moreover, developmental features of speech breathing, as captured by the respiration scale, had some influence on the listeners' perception of naturalness, which is also explainable by the large change of respiration in children between 3 and 9 years old (Schölderle et al., 2020), whereas the influence of the prosodic modulation scale must be ascribed to the contributions of only few children with a mildly monotonous speech who were perceived as unnatural. It is worth mentioning that some of the speech dimensions that are known to demonstrate a high variability in typical development, such as voice quality or fluency, seemed to be largely neglected by the listeners in their assessment of speech naturalness. A finding that is compatible with the strong influence of the articulation scores is the strong predictive value of intelligibility. A high intercorrelation between intelligibility and naturalness has already been shown in several studies of pathologic speech in adults (Dagenais et al., 2002; Lehner et al., 2022; Schölderle et al., 2016).

Thus, it was not unexpected that accuracy of articulation and its close covariate, intelligibility, were also strong predictors of unnatural speech in the CNC group. In contrast, the remaining variables in the developmental regression model of speech naturalness, that is, voice level, prosodic modulation, and respiration, were not particularly influential in the CNC model. Instead, listeners' naturalness ratings were more dependent on slow articulation rate and hypernasal resonance, a finding that is partly in line with previous research on adults with dysarthria. For instance, articulation rate has been identified as a key factor determining naturalness in individuals with Parkinson's disease (Klopfenstein, 2015). The articulation and articulation rate scores observed in the children with neurological disorders were markedly lower than those seen in typically developed children (Schölderle et al., 2020), which may explain their great importance for the impression of unnatural speech. As regards the resonance scores, hypernasal resonance was among the specific markers that distinguished dysarthric from developmental features in childhood dysarthria (Schölderle et al., 2022). The fact that it emerged here among the predictors of unnatural speech, though with only a small contribution, suggests that the naive listeners were sensitive to this feature.

Perceived naturalness was thus influenced by developmental speech characteristics in the typically developing children, whereas for the CNC, specific symptoms of dysarthria played a role as well. In both groups, however, the perception of naturalness was determined most strongly by the children's articulation and intelligibility. This commonality points at the overlap in the perceptual impression of speech between typically developing children and children with dysarthria. That is, restrictions in articulatory accuracy and intelligibility are not only indicators of speech motor development but also markers of neurogenic speech impairment (Schölderle et al., 2020), which made a differentiation between age-typical and impaired speech very hard for the listeners. This explains the low naturalness ratings in the typically developing children, particularly the youngest ones, who still showed prominent articulatory deficits and intelligibility limitations. It should be noted that, even when assessed by professional SLTs, a differentiation between developmental speech features and symptoms of dysarthria can only be accomplished in consideration of the salience of the respective feature and by taking into account the child's age. For example, slow articulation rate of a certain degree can be age appropriate in a 3-year-old child but indicate a disorder in a 7-year-old. In our study, listeners were not informed of the children's age, which may have had an influence here (see Limitations section).

The sensitivity and specificity of the laypersons' naturalness ratings were high in our study. With a score of 0.97, specificity was even slightly higher than reported for adult speakers by Lehner and Ziegler (2021). Only four out of 28 CNC ranged within the norm, and three of them were classified as nondysarthric. At first sight, this finding contradicts what has been pointed out before, as it shows that naturalness did in fact separate children with dysarthria from children without a speech disorder. We assume, however, that this finding is very specifically connected to the sample of children with dysarthria participating in this investigation. As has been pointed out before, a considerable proportion of the children enrolled in this study had severe dysarthria. The fact that their naturalness scores ranged quite far from typically developing children does not necessarily permit conclusions as to how well perceived naturalness can distinguish dysarthric from nondysarthric children in milder cases. Thus, the high sensitivity and specificity reported here must be put into perspective. The finding that typically developing children are perceived as unnatural still represents a major challenge for the concept as a clinically relevant descriptor of dysarthria in children.

Limitations

The sample of children with neurologic conditions was relatively small and not by any means representative of the population. As already mentioned, the fact that a large subgroup of children in this sample was severely affected from dysarthria reduces the informative value of our specificity and sensitivity coefficients. Furthermore, our sample contained only children of 5 years and older, whereas we had assessed typically developing children from 3 years on. In the typically developing children, especially the youngest were not perceived as completely natural. It remains unclear how well naturalness would have separated dysarthric from not dysarthric children below 5 years old. Due to the fact that many developmental speech features occur particularly strongly in the first years of life, the confusion between symptoms of childhood dysarthria and age-appropriate features of typical speech could have been even more critical. This issue requires further empirical investigation. On the other hand, standard speech assessment can be difficult in CNC younger than 5 years old, which led us to include older children in the first place.

In our study, listeners were not informed of the children's age, regarding neither the age range of the group nor the age of each individual child. If they had known the age of each individual child, they might have been able to apply an intuitive age normalization, at least to some extent, which may have led to higher naturalness ratings also in the younger typically developed children.

Clinical Implications

For a comprehensive and ecologically valid assessment of dysarthria, parameters beyond the auditory-perceptual description of the speech subsystems are needed. Particularly, clinical decision making should also take into account the consequences of the speech disorder for a person's everyday communication. In adults, speech naturalness as perceived by naive listeners has been shown to provide valuable information complementing the more common assessment of speech intelligibility. Our results have shown, however, that the application of the concept of perceived naturalness to children with dysarthria involves issues merely associated with developmental age. The assumption that naive listeners are able to intuitively distinguish the range of normal, undisturbed speech from the auditory characteristic of dysarthria, which has been confirmed repeatedly for adult speakers, might not hold true for children. Future research is needed to find and evaluate categories representing laypersons' perceptions of children's speech that are robust for typical developmental changes and thereby explore new ways of assessing communicative consequences of childhood dysarthria.

Data Availability Statement

Data can be requested from the authors. The possibility of sharing data will be scrutinized in each individual case in line with recommendations made by the Medical Faculty's Ethics Committee of the Ludwig-Maximilians-Universität München.

Acknowledgments

This research was supported by a grant from the German Research Foundation awarded to the first and the last author (ZI 469/17-1, SCHO 17421-1), as well as a PhD fellowship from the German National Academic Foundation awarded to the second author. The authors thank the children and their families as well as all listeners who participated in the study. Further thanks are due to Reha Hilfe e.V. for their support.

References

Baty, F., Ritz, C., Charles, S., Brutsche, M., Flandrois, J.-P., & Delignette-Muller, M.-L. (2015). A toolbox for nonlinear regression in R: The package nlstools. *Journal of Statistical Software, 66*(5), 1–21. https://doi.org/10.18637/jss.v066.i05

Cahill, L. M., Murdoch, B. E., & Theodoros, D. G. (2002). Perceptual analysis of speech following traumatic brain injury in childhood. *Brain Injury, 16*(5), 415–446. https://doi.org/10.1080/02699050110119871

Coughlin-Woods, S., Lehman, M. E., & Cooke, P. A. (2005). Ratings of speech naturalness of children ages 8–16 years. *Perceptual and Motor Skills, 100*(2), 295–304. https://doi.org/10.2466/pms.100.2.295-304

Dagenais, P. A., Brown, G. R., & Moore, R. E. (2006). Speech rate effects upon intelligibility and acceptability of dysarthric speech. *Clinical Linguistics & Phonetics, 20*(2–3), 141–148. https://doi.org/10.1080/02699200400026843

Dagenais, P. A., & Wilson, A. F. (2002). Acceptability and intelligibility of moderately dysarthric speech by four types of listeners. *Investigations in Clinical Phonetics and Linguistics, 8,* 363–372.

Duffy, J. R. (2020). *Motor speech disorders: Substrates, differential diagnosis, and management* (4th ed.). Mosby Incorporated.

DuHadway, C. M., & Hustad, K. C. (2012). Contributors to intelligibility in preschool-aged children with cerebral palsy. *Journal of Medical Speech-Language Pathology, 20*(4), 11.

Fox-Boyer, A. (2020). *TROG-D: Test zur Überprüfung des Grammatikverständnisses* [Test for Reception of Grammar–German Version]. Edition Steiner im Schulz-Kirchner Verlag.

Haas, E., Ziegler, W., & Schölderle, T. (2021). Developmental courses in childhood dysarthria: Longitudinal analyses of auditory-perceptual parameters. *Journal of Speech, Language, and Hearing Research, 64*(5), 1421–1435. https://doi.org/10.1044/2020_JSLHR-20-00492

Haas, E., Ziegler, W., & Schölderle, T. (2022). Intelligibility, speech rate, and communication efficiency in children with neurological conditions: A longitudinal study of childhood dysarthria. *American Journal of Speech-Language Pathology, 31*(4), 1817–1835. https://doi.org/10.1044/2022_AJSLP-21-00354

Hebbali, A. (2020). *Olsrr: Tools for building OLS regression models* (R package Version 0.5) [Computer software].

Hidecker, M. J. C., Paneth, N., Rosenbaum, P. L., Kent, R. D., Lillie, J., Eulenberg, J. B., Chester, K., Jr., Johnson, B., Michalsen, L., Evatt, M., & Taylor, K. (2011). Developing and validating the Communication Function Classification System for individuals with cerebral palsy. *Developmental Medicine & Child Neurology, 53*(8), 704–710. https://doi.org/10.1111/j.1469-8749.2011.03996.x

Hustad, K. C., Mahr, T. J., Natzke, P., & Rathouz, P. J. (2021). Speech development between 30 and 119 months in typical children I: Intelligibility growth curves for single-word and multiword productions. *Journal of Speech, Language, and Hearing Research, 64*(10), 3707–3719. https://doi.org/10.1044/2021_JSLHR-21-00142

Hustad, K. C., Schueler, B., Schultz, L., & DuHadway, C. (2012). Intelligibility of 4-year-old children with and without cerebral palsy. *Journal of Speech, Language, and Hearing Research, 55*(4), 1177–1189. https://doi.org/10.1044/1092-4388(2011/11-0083)

Jones, H. N., Crisp, K. D., Kuchibhatla, M., Mahler, L., Risoli, T., Jr., Jones, C. W., & Kishnani, P. (2019). Auditory-perceptual speech features in children with down syndrome. *American Journal on Intellectual and Developmental Disabilities, 124*(4), 324–338. https://doi.org/10.1352/1944-7558-124.4.324

Klopfenstein, M. (2015). Relationship between acoustic measures and speech naturalness ratings in Parkinson's disease: A within-speaker approach. *Clinical Linguistics & Phonetics, 29*(12), 938–954. https://doi.org/10.3109/02699206.2015.1081293

Klopfenstein, M., Bernard, K., & Heyman, C. (2020). The study of speech naturalness in communication disorders: A systematic review of the literature. *Clinical Linguistics & Phonetics, 34*(4), 327–338. https://doi.org/10.1080/02699206.2019.1652692

Lee, J., Hustad, K. C., & Weismer, G. (2014). Predicting speech intelligibility with a multiple speech subsystems approach in children with cerebral palsy. *Journal of Speech, Language, and Hearing Research, 57*(5), 1666–1678. https://doi.org/10.1044/2014_JSLHR-S-13-0292

Lehner, K., & Ziegler, W. (2021). Clinical measures of communication limitations in dysarthria assessed through crowdsourcing:

Specificity, sensitivity, and retest-reliability. *Clinical Linguistics & Phonetics, 36*(11), 988–1009. https://doi.org/10.1080/02699206.2021.1979658

Lehner, K., Ziegler, W., & KommPaS Study Group. (2022). Indicators of communication limitation in dysarthria and their relation to auditory-perceptual speech symptoms: Construct validity of the KommPaS web app. *Journal of Speech, Language, and Hearing Research, 65*(1), 22–42. https://doi.org/10.1044/2021_JSLHR-21-00215

Mei, C., Reilly, S., Reddihough, D., Mensah, F., & Morgan, A. (2014). Motor speech impairment, activity, and participation in children with cerebral palsy. *International Journal of Speech-Language Pathology, 16*(4), 427–435. https://doi.org/10.3109/17549507.2014.917439

Moya-Galé, G., Keller, B., Escorial, S., & Levy, E. S. (2021). Speech treatment effects on narrative intelligibility in French-speaking children with dysarthria. *Journal of Speech, Language, and Hearing Research, 64*(6S), 2154–2168. https://doi.org/10.1044/2020_JSLHR-20-00258

Palisano, R., Rosenbaum, P., Walter, S., Russell, D., Wood, E., & Galuppi, B. (1997). Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Developmental Medicine & Child Neurology, 39*(4), 214–223. https://doi.org/10.1111/j.1469-8749.1997.tb07414.x

Pennington, L. (1999). Assessing the communication skills of children with cerebral palsy: Does speech intelligibility make a difference? *Child Language Teaching and Therapy, 15*(2), 159–169. https://doi.org/10.1177/026565909901500204

Revelle, W. (2022). *Psych: Procedures for personality and psychological research* (Version 2.2.5). Northwestern University. https://CRAN.R-project.org/package=psych

Schölderle, T., Haas, E., Baumeister, S., & Ziegler, W. (2021). Intelligibility, articulation rate, fluency, and communicative efficiency in typically developing children. *Journal of Speech, Language, and Hearing Research, 64*(7), 2575–2585. https://doi.org/10.1044/2021_JSLHR-20-00640

Schölderle, T., Haas, E., & Ziegler, W. (2020). Age norms for auditory-perceptual neurophonetic parameters: A prerequisite for the assessment of childhood dysarthria. *Journal of Speech, Language, and Hearing Research, 63*(4), 1071–1082. https://doi.org/10.1044/2020_JSLHR-19-00114

Schölderle, T., Haas, E., & Ziegler, W. (2022). Childhood dysarthria: Auditory-perceptual profiles against the background of typical speech motor development. *Journal of Speech, Language, and Hearing Research, 65*(6), 2114–2127. https://doi.org/10.1044/2022_JSLHR-21-00608

Schölderle, T., Staiger, A., Strecker, K., Lampe, R., & Ziegler, W. (2016). Dysarthria in adults with cerebral palsy: Clinical presentation and impacts on communication. *Journal of Speech, Language, and Hearing Research, 59*(2), 216–229. https://doi.org/10.1044/2015_JSLHR-S-15-0086

Southwood, M. H., & Weismer, G. (1993). Listener judgments of the bizarreness, acceptability, naturalness, and normalcy of the dysarthria associated with amyotrophic lateral sclerosis. *Journal of Medical Speech-Language Pathology, 1*(3), 151–161.

Tuszynski, J. (2018). *caTools: Tools: moving window statistics, GIF, Base64, ROC AUC, etc.*. (Version R package Version 1.17.1.1.). https://CRAN.R-project.org/package=caTools

Whitehill, T. L., Ciocca, V., & Yiu, E. M.-L. (2004). Perceptual and acoustic predictors of intelligibility and acceptability in Cantonese speakers with dysarthria. *Journal of Medical Speech-Language Pathology, 12*(4), 229–234.

World Health Organization. (2001). *International classification of functioning, disability and health: ICF.*

Yorkston, K. M., Hammen, V. L., Beukelman, D. R., & Traynor, C. D. (1990). The effect of rate control on the intelligibility and naturalness of dysarthric speech. *Journal of Speech and Hearing Disorders, 55*(3), 550–560. https://doi.org/10.1044/jshd.5503.550

Ziegler, W., Schölderle, T., Staiger, A., & Vogel, M. (2018). *BoDyS—Bogenhausener Dysarthrieskalen* [Bogenhausen Dysarthria Scales]. Hogrefe Verlag.