



Speech assessment tool methods for speech impaired children: a systematic literature review on the state-of-the-art in Speech impairment analysis

Gowri Prasood Usha¹ · John Sahaya Rani Alex¹ 

Received: 4 January 2022 / Revised: 23 December 2022 / Accepted: 12 February 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Speech is a powerful, natural mode of communication that facilitates effective interactions in human societies. However, when fluency or flow of speech is affected or interrupted, it leads to speech impairment. There are several types of speech impairment depending on the speech pattern and range from mild to severe. Childhood apraxia of speech (CAS) is the most common speech disorder in children, with 1 out of 12 children diagnosed globally. Significant advancements in speech assessment tools have been reported to assist speech-language pathologists diagnosis speech impairment. In recent years, speech assessment tools have also gained popularity among pediatricians and teachers who work with preschoolers. Automatic speech tools can be more accurate for detecting speech sound disorders (SSD) than human-based speech assessment methods. This systematic literature review covers 88 studies, including more than 500 children, infants, toddlers, and a few adolescents, (both male and female) (age = 0–17) representing speech impairment from more than 10 countries. It discusses the state-of-the-art speech assessment methods, including tools, techniques, and protocols for speech-impaired children. Additionally, this review summarizes notable outcomes in detecting speech impairments using said assessment methods and discusses various limitations such as universality, reliability, and validity. Finally, we consider the challenges and future directions for speech impairment assessment tool research.

Keywords Speech sound disorders · Speech assessment tools · Speech impairments · Speech assessment methods · Children

✉ John Sahaya Rani Alex
jsranialex@vit.ac.in

¹ School of Electronics Engineering, Vellore Institute of Technology, Chennai 600127, India

1 Introduction

Speaking difficulties, whether in producing sound or in other aspects of articulation, are collectively known as speech impairment. Speaking difficulties encompass several types of disorders and can range from mild to severe. Language Speech Impairment (LSI) [86] is a form of speech impairment that occurs without any evident underlying mental or physical disorder or direct neurological damage. More specifically, a language disorder describes an impairment in comprehension and spoken, written, and other symbol systems [26]. Speech sound disorders (SSD) are categorised into articulation, fluency, and voice disorders.

Previous studies have shown that childhood apraxia of speech is one of the most common disorders among children, with 1 out of 12 children globally affected by this condition [86]. Existing literature indicates that SSD prevalence in children is comparable in monolingual and multilingual communities [47]. Children with SSD present with low speech lucidity and have retarded speech sound acquisition [34]. Therefore, helpful speech evaluation tools must be developed to help speech-language pathologists (SLP) detect speech deficits in children as early possible to begin appropriate intervention. To better understand the current scenario and provide a foundation on which subsequent studies can be built on, this paper systematically reviews literature for voice assessment tool methods for children with speech impairments.

Earlier studies have dedicated little attention to understanding the morbid impacts of infectious diseases and epidemics in developing countries. There are several risks associated with these epidemics, such as the possibility of neurocognitive impairments in the children who survive the epidemic [17]. Therefore, SLPs must be culturally and linguistically competent to deliver effective patient service and not only cater to a specific demographic [28]. Traditional articulation treatment methods aim to rectify solitary speech sounds instead of phonological interventions that address speech sound systems [16]. Hence, the most desirable speech assessment tool methods are those that use the latter approach.

Adopting measures that reduce the need for further treatment will positively impact the children and their families, as well as the treatment systems itself [60]. During preschool, family members often misunderstand children with SSD since they are unintelligible [23]. The delay in their literacy competencies is often severe and present with concomitant language disorders [16]. Additionally, poor social relations among children with SSD might negatively impact their self-image [10, 12]. Despite such consequences, there is little evidence about the treatments SLPs employ when treating children with SSD [16]. Efficient and effective treatment methods must therefore be developed and promoted.

According to recent studies, the prevalence of speech and language impairments in children is rising [67, 68]. Speech therapy is the most common therapeutic intervention for SSD, but it is also one of the most expensive and challenging treatments available other than surgery. A speech-language evaluation normally costs between \$200 to \$300, and a half-hour therapy session may cost between \$50 to \$100 although the actual cost of speech therapy can vary depending on various factors [31]. In addition, it takes numerous sessions to observe a noticeable improvement in the children. According to research, intensive intervention is more successful and efficient for kids with SSD [43]. In other words, one might require multiple sessions each week. The research published in the literature thus indicates a likely increase in the number of children with SSD in the future, considering the costs of intervention and frequency, which are the limiting factors. A suitable, efficient, and cost-effective treatment should be available for these children to lower the rate of child SSD and help them navigate the condition.

Children presenting with cleft palate lip are likely to develop speech difficulties that will require speech and language therapy [13]. According to Cummins et al. (2015), speech is a sensitive output system due to the complexity of speech production; hence, slight physiological and cognitive changes potentially can produce noticeable acoustic changes [20]. Brookes and Bowley (2014) describe tongue-tie as a congenital state characterised by a short lingual frenulum that could restrict the tongue's movement and influence its function [14]. Studies have shown that tongue-tie is a common disorder with a documented 3–4% incidence among infants [9]. Therefore, a universal criterion for diagnosing children's language impairments is necessary to reduce present variations.

Fundamental elements of communicative competence encompass a framework that describes reasonably intelligible pronunciation [22]. Perceptual measures, which form a part of the comprehensive speech evaluation, are concerned with assessing the speaker's intelligibility [10], while a systematic speech pathology assessment tool uses articulation to predict the overall intelligibility score [12]. Intervention outcomes associated with speech impairments, such as increased sentence length, improved articulatory function, and use of grammatical markers, form the traditional focus of studies assessing speech-language therapies' effectiveness [21]. This paper aims to conduct a systematic literature review of the speech assessment tools for impaired speech children. Here, we review the speech impairment detection tools to establish current trends and findings in the educational relevant domain.

The current review presents the research and studies involved in speech assessment methods for children and adolescents with different speech impairments from 2010 to 2022. We present the methodology adopted in this study and literature review results in Sections 2 and 3, respectively. Section 4 presents the discussion, while Section 5 mentions future directions and challenges. Finally, we conclude the study in Section 6.

1.1 Purpose

In this review, we aim to address the following research questions:

1. Speech assessment methods and purpose:
 - a. What are the different types of assessment methods being used?
 - b. For what disordered language or disordered speech and the range of delay or disorder investigated?
2. Accuracy of analysis: How do these methods perform, and their efficiency/precision?
3. Is there room for improvement in these methods for the early detection of speech and language disorders?

Though these research questions are interrelated and discussed throughout the article, the speech assessment methods and purpose have been discussed mainly in Sections 3 and 4. Accuracy analysis of the methods is covered in Section 3 and especially in Table 3 on pages 9 and 10, but the efficiency of the method concerning the studies reported has been explained in Section 4. Finally, the challenges associated with existing methods and the ways to improve them are explored in Section 5.

2 Literature selection criteria

The authors searched for primary and secondary peer-reviewed articles that met the quality assessment criteria in this systematic literature review. Various digital databases were queried using keyword search to select the study's most appropriate and relevant papers. The criteria for exclusion and inclusion were met in the document studies that were analysed. Therefore, this paper's research design is a systematic approach that adheres to an outlined study protocol.

The research question was to establish whether methods can detect SSD using different techniques to develop practical speech assessment tools. The reliance on a well-defined methodology ensured that research bias is eliminated to result in fair and objective outcomes. The authors designed, reviewed, and revised the study protocol for the present review. Here, we analysed each peer-reviewed article twice to ascertain that the extracted data complied with the review protocol. The search strategy, criteria for integration and exclusion, and quality assessment process are described in detail in the following sections. We followed the PRISMA protocol to perform the systematic literature review to achieve higher transparency and reliability.

2.1 Search strategy

We established the existing studies in speech assessment tools for speech-impaired children by querying online databases such as Medline, ScienceDirect, CINAHL, EMBASE, IEEE Xplore, PsychInfo, Web of Search, SpringerLink, Scopus, First Search, ERIC, ACM Digital Library, Linguistics and Language Behaviour Abstracts, and DARE for articles that contained the keywords speech, speech impairments, speech assessment tools, speech impaired children, speech analysis and SSD in the title, abstract. Additionally, the authors queried Scopus and Web of Science to locate other published articles in little-known online libraries. The rationale behind the search strategy was to find significant peer-reviewed articles with full-text and conference proceedings related to the field of "Speech Impairment" and "Speech Assessment Tools". The keywords used during the search strategy were expected to yield most of the papers containing speech assessment tools. Google Scholar and Google search engines were also utilised to ensure no relevant article was omitted from the study. The author conducted the entire search process, and the process was finalised on 11th May 2022.

2.2 Inclusion and exclusion criteria

The researcher developed a pilot version of the selection criteria that targeted all relevant primary studies and finalised it after revising the review protocol. The authors' institutional affiliation and names were irrelevant when deciding on the inclusion and exclusion criteria. Exclusion criteria were as follows: Studies that did not include speech assessment tools and those that did not have robust speech assessment mechanisms for speech-impaired children; papers that failed incorporate the speech assessment tool's interrater reliability were not considered for the study; overlapping studies from various journals and online databases; and peer-reviewed studies published before 2010.

In the end, only 92 items that were written in English from 2010 onwards about speech assessment tools, protocols and methods for speech-impaired children were selected for a systematic analysis. Only original articles were included in the review. Additionally, these studies include interrater reliability of the speech assessment tools between 2010 and 2022. A

significant proportion of the 92 articles selected for review had different authors, while a small number authoring more than one paper was found. Fig. 1 below shows the scientometric mapping of the type of research conducted by authors in the review articles.

Figure 2 shows the year-wise category of the papers selected from 2010 to 2022. The full-text paper's quality is assessed based on the sampling method, the study's sample size, and whether the survey is cohort or research-based.

3 Speech impairment analysis methods

3.1 Tools, techniques, and protocols

3.1.1 Tools

The planning and coordination of speech arises from complex neurological interactions occurring in certain brain regions while fold vibrations in the larynx generate signals that make speech audible [71]. According to Strand et al., Paediatric SSD results from various aetiologies and impairs speech production on several levels, including linguistic/phonological and motor speech [79]. Establishing the degree of contribution of motor speech impairment in the child's SSD is one of the principal difficulties during differential diagnosis [23]. Hence, it is necessary to develop a speech assessment tool that will eliminate these existing challenges. In the subsequent sections, we describe and compare the leading speech assessment methods currently employed by paediatricians, clinics, and therapists.

Dynamic Evaluation of Motor Speech Skill (DEMSS) DMESS tool is designed to help in differential detection of SSD among both young children and older children. It is challenging to isolate deficits in plan and program transitions between the volitional speech articulation positions of SSD children, partly because of the interactive speech and language processes. DMESS is a recent speech assessment tool designed to counter the abovementioned issue [79]. Strand et al., relied on expert opinions and current literature to conclude that there is consensus

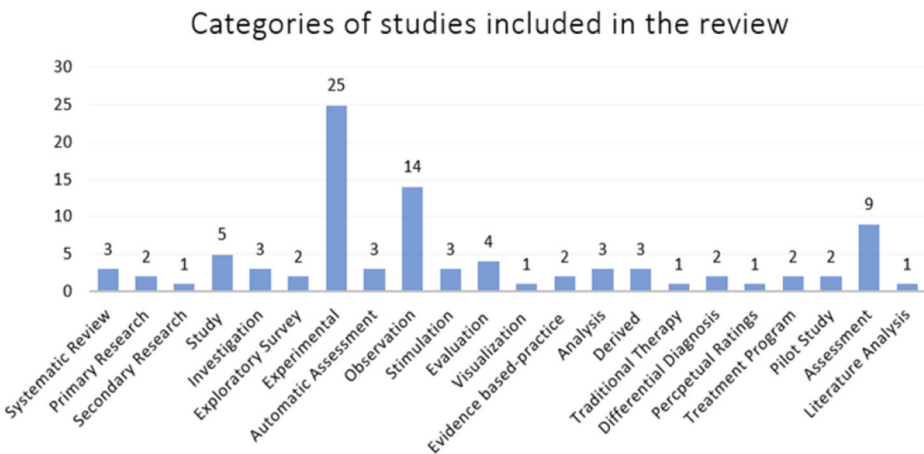


Fig. 1 Scientometric mapping of the categories of peer-reviewed papers included in the systematic review

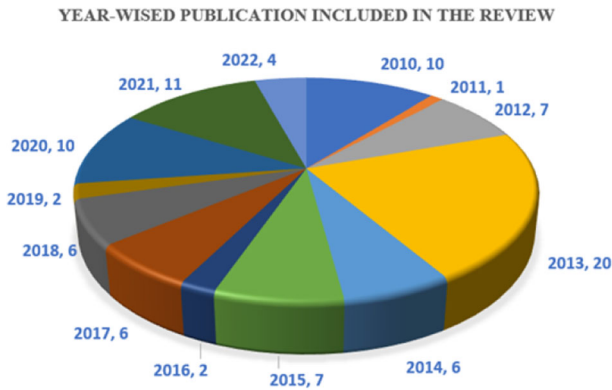


Fig. 2 Number of papers included in the review published year-wise from 2010 to 2022

among researchers about CAS [79]. CAS is in the recurrent construction of words or syllables through erroneous vowels and consonants; extended and disorderly co-articulatory shifts linked syllables and sounds; and ill-suited prosody achievement phrasal or lexical string [79]. Clinical assessment of children with SSD typically involves issuing oral structural-functional tests [66].

According to Strand et al., as a motor speech examination, the DEMSS systematically varies the length, vowel content, prosodic content, and phonetic complexity within sampled utterances [79]. DEMSS test is designed to test young children's speech movements and little ones with severe speech impairment. It does not act as a phonologic proficiency or articulation test which evaluates overall segments in a language. It is designed for children incapable of producing syllables, sounds, or words.

The DEMSS is concerned with earlier developing consonant sounds matched with an array of vowels in numerous evolving syllable shapes [59]. The DEMSS comprises nine subtests consisting of 66 utterances, as shown in this Table 1. The 66 pronunciations contain 171 judgmental items that make four sets of sub-scores [79]. The severity of the childhood apraxia of speech is determined based on the child's overall score after taking the test.

Table 1 Dynamic Evaluation of Motor Speech Skills (DEMSS) Content Coverage

| Utterance type (examples) | No. of utterances | Overall articulatory accuracy | Vowel accuracy | Prosodic accuracy | Consistency |
|---|-------------------|-------------------------------|----------------|-------------------|-------------|
| CV (me, hi) | 8 | 8 (0–32) | 8 (0–16) | - | 4 (0–4) |
| VC (up, eat) | 8 | 8 (0–32) | 8 (0–16) | - | 4 (0–4) |
| Reduplicated Syllables (mama, booboo) | 4 | 4 (0–16) | - | 4 (0–4) | - |
| CVC1 (mom, peep, pop) | 6 | 6 (0–24) | 6 (0–12) | - | 6 (0–6) |
| CVC2 (mad, bed, hop) | 8 | 8 (0–32) | 8 (0–16) | - | 8 (0–8) |
| Bisyllabic 1 (baby, puppy) | 5 | 5 (0–20) | 5 (0–10) | 5 (0–5) | - |
| Bisyllabic 2 (bunny, happy) | 6 | 6 (0–24) | - | 6 (0–6) | - |
| Multisyllabic (banana, kangaroo) | 6 | 6 (0–24) | 6 (0–12) | 6 (0–6) | 6 (0–6) |
| Utterances of increasing length (dad, hi dad, hi daddy) | 15 | 15 (0–60) | 15 (0–30) | - | - |
| Total utterances | 66 | 66 | 56 | 21 | 28 |

The DEMSS is the most influential speech assessment tool among children with impaired speech. It incorporates the dynamic assessment for judgments about severity and prognosis. The medical practitioner administering the DEMSS test instructs the child to fixate their eyes on the instructor's face as much as possible while uttering a series of words. Considering the child's first imitation, the pediatrician might use various levels of cuing to elicit more imitative attempts before compiling the final score. Evidence shows that the DEMSS tool is one of the most suitable speech assessment tools since it indicates the SSD severity. Since the tool utilises a dynamic assessment, the pediatrician incorporates cues and other techniques, such as simultaneous production or slowed rate, to elicit several scoring attempts. The prosody and vowel accuracy scoring are done when the child first attempts an utterance. Overall articulatory accuracy is not scored based on the initial effort but on subsequent trials [79]. Table 2 illustrates the basic rules clinicians follow when scoring the child within the four sub-scores: vowel accuracy, consistency, overall articulatory accuracy, and prosodic accuracy (lexical stress accuracy), with poor performance symbolised by higher scores [79].

Motor Speech Examination MSE, often used to establish the presence or absence of speech motor programming and planning in adults, can also be adapted to diagnose SSD in young children [79]. MSE enables a pediatrician to detect speech construction across utterances that differ in phonetic complexity and length using organised stimuli systematically to vary programming demands. Previous studies have shown that only the Verbal Motor Production for children, among the six documented assessment tools for diagnosing SSD, passed the validity test, although none of the tests recorded reliability [79]. Therefore, there is a need to develop an MSE tool that provides proof of validity and reliability.

According to Strand et al., providing evidence of reliability is critical to developing speech assessment examinations. Validity in MSE tools can be described as the extent to which the study measures the elements it seeks to evaluate [79]. Several approaches can document the validity of a given test used in SSD diagnosis. Therefore, the validity and reliability measures of a particular speech assessment tool are critical in determining the overall acceptance of its outcomes.

The most frequently used validity measures methods are the gold standard (acknowledged valid measure) and contrasting correlations and groups between the examinations under investigation [79]. Another technique used to measure the validity evidence of an MSE test

Table 2 DMESS Scoring

| Scores | 5-point multidimensional scoring | 3-point multidimensional scoring | Binary scoring | |
|--------|--|----------------------------------|-------------------|--|
| | | | Prosodic accuracy | Consistency |
| 0 | Correct on first attempt | Correct | Correct | Consistent across all trials |
| 1 | Consistent developmental substitution error (e.g., /t/ for /k/, /w/ for /r/) without slowness or distortion of movement gestures | Mild distortion | Incorrect | Inconsistent across any 2 or more trials |
| 2 | Correct after first cued attempt | Frank distortion | - | - |
| 3 | Correct after 2 or 3 additional cued attempts | - | - | - |
| 4 | Not correct after all cued attempts | - | - | - |

is cluster analysis. Cluster analysis is commonly used to evaluate constructs that identify homogeneous subcategories within broader clusters, including civic language disorders, autism spectrum disorders, and SSDs [32]. Moreover, they are used to detect non-speech and co-occurring speech characteristics in childhood apraxia of speech [79]. The test validity is evident if the results of the examination mirror those conducted using different diagnostic tools.

Automatic speech analysis tools Children with difficulties producing intelligible speech are categorised as having paediatric SSD [75]. Speech impairment can occur during speech production's motor planning, linguistic, or motor execution phases [77]. Technological advancements in automatic speech analysis have reinforced the idea that artificial intelligence can use for speech assessment and intervention for children with SSD [3, 53]. Clients and parents have shown interest in the cost-friendly alternative measure since the existing speech assessment and intervention techniques are costly for children who need intensive and long-term speech therapy, placing multiple barriers in the way of effective service delivery. Computer-driven approaches incorporating online gaming are the long-term solution to removing the aforementioned barriers [81]. Tabby Talks is one of the automated tools for assessing childhood apraxia of speech. Devices are composed of clinician interface, mobile application, and speech processing engine and identify grouping errors, articulation errors, and prosodic errors [73]. Tabby Talks tool offers the capability to reduce the enormous amount of speech therapists' work and the time and finance for families.

The earliest forms of automatic speech analysis and recognition (ASA) tools developed in the 1960s and 70s could process isolated sounds from minute to medium pre-defined lexicon [44]. Linear predictive coding (LPC) was developed to account for variations arising from vocal tract differences. Technological advancements in the 1980s based on statistical probability modelling that a specific set of language symbols matched the incoming utterance signal enhanced the ASA tools.

The predominant technology utilised by most speech recognition systems is the Hidden Markov Models (HMMs), which are designed to undertake temporal pattern recognition [44]. According to McKechnie et al. (2018), In the 1990s, new pattern recognition innovations led to discriminatory training and kernel-based techniques that functioned as classifiers, such as Support Vector Machines (SVMs). Fig. 3 below shows the theme component processes model encompassed in new ASA systems [44]. Therefore, the superior technological advancements in ASA tools enable the system to sift through speech variations from different speaker.

While ASR systems have vastly improved in recent years, children's ASR remains are not as well-known as adult ASR. Children's HMM-ASR systems, like deep neural network ASR systems, require much data to train and are extremely reliant on the data they use. Clinical speech data (particularly for children's speech) is far more challenging than average speech data, and physicians cannot be expected to collect enough data for such systems. More research is needed to develop clinical evaluation systems with minimal training data. The limitation of databases that contain large languages is another element that hinders system development and performance accuracy. The speech acoustic model is the second component that impacts performance accuracy and is based on the speaker mode. The model can either rely on the speaker, be independent of the speaker, or speaker adaptive. The ASA tools system also has two other principal components that influencing its accuracy [44]. The type of speech (isolated words or continuous speech) and the lexicon's size impact the feature extraction

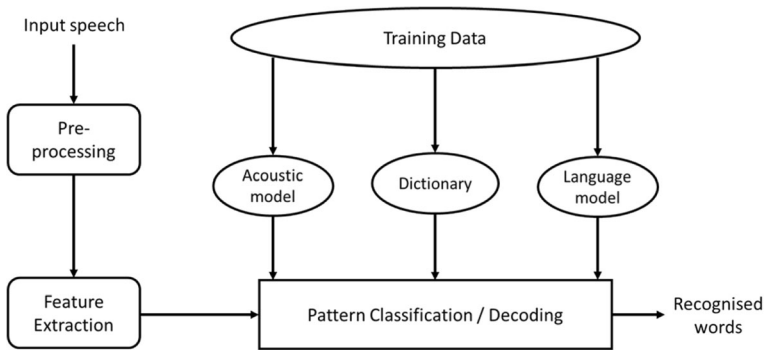


Fig. 3 Model of contemporary ASA speech recognition system

process, the first component of the ASA tools with improved performance measured through long vocabularies. Therefore, the feature extraction and speech acoustic model affect the performance accuracy of the ASA tools. Notwithstanding the significant improvements in ASA tools, computational modelling systems still experience challenges [78]. Specifically, young children undergoing developmental growth stages while committing speech errors present even more challenges for ASA tools designed to assess children's speech [44]. Therefore, the ASA tools need to consider the impact of impaired speech assessment and children intervention.

The major tool for clinical assessment of speech-language disorders, one of the most common juvenile disabilities, is auditory perceptual analysis (APA). APA outcomes, however, are subject to intra- and inter-rater variability. Manual or hand transcription-based speech problem diagnostic approaches have various drawbacks. To address these constraints, there is a growing interest in creating automated approaches for identifying speech abnormalities in children that quantify speech patterns. Landmark (LM) analysis is a method of identifying auditory events that occur as a result of sufficiently accurate articulatory motions [61] and it is suggested that LMs be used to detect speech disorders in youngsters automatically. This study offered a series of novel knowledge-based features that were not previously proposed, in addition to the LM-based features that have been proposed in previous studies. To test the usefulness of the innovative features in differentiating speech disorder patients from regular speakers, a comprehensive investigation and comparison of several linear and nonlinear machine learning classification approaches based on raw characteristics and proposed features are done.

The lest speech assessment tool Language is the medium used to exchange the abovementioned elements between people of different races, colours, and religions [84] and is defined as the sound produced by the human voice, which the ear receives and interprets by the brain [57]. The LEST scale was developed to address universal and direct language development assessment in neuro-developmental follow-up clinics. The LEST tool was used for two groups of children; the first group for 0–3 years and the second group for 3–6 years. Each category encompasses items concerning expressive and receptive language development. Therefore, the LEST is one of the various speech assessment tools clinicians use in children with SSD for diagnosis and intervention.

Battery of Western Speech and Language Assessment Tool Motor aphasia was first diagnosed by the French neurologist Paul Broca in the 1860s. The condition is associated with patients who can comprehend what is said but have difficulties exhibiting speech fluency, leading to communication breakdown. The Battery of Western Speech and Language Assessment Tools was developed to detect this speech impairment condition [17].

CHOCSLAT – Chinese Healthcare-Oriented Computerised Speech & Language Assessment Tools The CHOCSLAT relies on technology to identify speech impairments in children. This tool aims to provide a technical advance in helping children who may have speech impairment or language delay. The computer records the utterances for processing and analysis. The C-LARSP (Chinese Language Assessment, Remediation, and Screening Procedure) is used in the grammar assessment section and concentrates on the grammatical classification and meanings of children's statements.

The grammatical structures are classified by age group ("stage") and several grammatical levels (clause, phrase, and word prefix/suffix), allowing for evaluation of children's grammar at seven different age stages, ranging from 1 year to 4 years and 6 months (labelled "4; 6") and above. The marking scheme incorporates semantic and syntactic features and result in a score ranging from 0 to 5 depending on the child's response. The Phonology Assessment of Chinese (Mandarin) is used in the phonology test, and consists of 44 prompts, each of which targets a one- or two-character Chinese word. Percent Consonant Correct (PCC), present & absent consonants, and mistake patterns are three characteristics of pronunciation that are measured and analysed (mispronunciations that follow specific patterns). The average accuracy of all sample sentences is used to calculate the total accuracy. With $N = 106$ sentences, the most recent prototype iteration attained an average accuracy of 0.87. Several challenges were encountered while developing the tool, like using pinyin instead of the International Phonetic Alphabet (IPA) for the transcription, even though pinyin lacks accuracy and specificity compared to IPA. The tool was developed in close collaboration between Chinese experts in applied linguistics, computer scientists, and speech pathologists [83].

CELF-4 The Clinical Evaluation of Language Fundamentals (CELF) is a comprehensive speech impairment assessment tool to evaluate a child's speech and language skills competency in various contexts. The aim is to identify the present speech and language disorders, their category, and the necessary intervention to treat the condition [57]. For children aged 5 to 21, the CELF-4 is considered a standard gold assessment for detecting language problems or delays. CELF-4 acts as a bridge to between the speech pathologist and children, assist in determining why a child may require classroom language adaptations, improvements, or curriculum changes. Its ability to administer subtests in various ways allows for faster testing while delivering extraordinarily reliable and accurate findings. After administering the CELF-4 battery, six indices can be calculated: the core language index and five other language indices. CELF-4 is relevant and is an exciting alternative for children due to its cultural inclusiveness and visual stimuli. The CELF-4 was created to reflect the clinical decision-making process, which begins with a diagnosis and determining the severity of a language disorder, then moves on to identifying relative strengths and weaknesses, making recommendations for accommodations and intervention, and evaluating the effectiveness of the intervention.

PLS-5 English This tool was developed to assess and analyse language developmental milestones in children to identify the presence or absence of SSDs [76]. The screening test tool is designed for the children to screen their broad spectrum of language and speech skills from 0 to 7 age. Also, it helps to identify the language disorder within 6 speech and language areas in just 5 to 10 min. PLS-5 contains 2 standardized scales; one is to determine how a child communicates with others (Expressive Communication), and the second is to evaluate a child's language comprehension. The PLS-5 has a good to excellent test-retest reliability ($r = 0.86\text{--}0.95$). The auditory comprehension and expressive communication scores had an internal consistency of $r > 0.80$ and $r > 0.9$, respectively. Test content (comprehensive/skills elicited are diagnostic indicators of whether a child is developing language typically or has a language disorder), response processes (effectively elicited), the internal structure (highly homogeneous within and across scales), and evidence-based relationships with the prior version of the test ($r = 0.80$ for both subscales) and other tests that measure the same constructs are all used to support the validity of the PLS-5 (moderate to high correlations ranging from 0.70 to 0.82 with the Clinical Evaluation of Language Fundamentals Preschool 2). The PLS-5 produces norm-referenced test results, such as Standard Scores (Mean = 100, SD = 15).

GFTA-3 The Goldman-Fristoe Test of Articulation is a tool used to evaluate the articulation of consonant sounds in children to reveal the disorder's severity if present [4]. The Goldman Fristoe Articulation test is open to children over the age of 2 and under 22. The GFTA-3 is a widely used standardised speech test that assesses children's pronunciation using clinically relevant utterances. Using the GFTA-3 assessment framework, clinicians tracked the quality of each child's phoneme pronunciation; each kid was positioned in a sound booth with a double-walled sound barrier, and a student clinician administered the GFTA-3.

"Sounds in Words" and "Sounds in Sentences" are the two sections of the GFTA-3. For the sounds in words subtest, picture stimuli and target words elicited the production of 23 consonant sounds and 15 consonant clusters, whereas the storey retell task elicited connected speech for the sentence's subset. Scoring and interpretation depend on omissions, addition (phonetic transcription), and raw score (count number of incorrect responses). The raw score will be converted into standard, percentiles, and age equivalents. The scores are then used to compare individual results to gender-specific norms.

The norms were determined using a national sample of 1,500 examinees by age and gender. Test-retest and internal consistency is used to verify the tool's reliability. Evidence-based test content, response processes, the performance of a speech sound disorder group, and its relationship with the GFTA-2 are used to support the tool's validity. The GFTA-3 is appropriate to test those with suspected word production disorder. The GFTA-3 identifies the presence or absence of distinct speech sounds within the client's repertoire but is not without its disadvantages. The sentence length requirement may be too high, and some graphics may be obscure to some young children. An additional limitation is that it only for children who have trouble pronouncing consonants (b,c,d, etc.) and will not help identify whether a child has articulation issues or problems with vowels.

Bayley-III The Bayley Scale of Infant and Toddler Development (Bayley) evaluates the developmental speech milestones of children aged 1 to 42 months. This tool's primary aim is to detect any speech disorders in the child to develop the necessary intervention strategies [76]. The third edition

(Bayley-III) is a simple, straightforward, method used to measure cognitive and motor skills and its results are exceedingly reliable. It is delivered with the help of a caregiver or parent, allowing for more input from the child's natural surroundings. Furthermore, all assessment parameters are based on the child's age, allowing for more precise developmental assessments. It is a comprehensive solution for assessing the entire kid, including adaptive behaviour, cognitive, language, social-emotional, and motor abilities. The Bayley-III produces composite and subscale scores for fine and gross motor development and composite and subscale scores for cognition and motor ability. For composite scales, raw scores are converted to norm-referenced standard scores (mean = 100, SD = 15), and for motor subscales, scaled scores (mean = 10, SD = 3).

DAS-II The Differential Ability Scales assessment (DAS-II) assesses children's cognitive competencies. The device identifies mental and cognitive disorders in children aged 2 to 18 [65]. The DAS-II is a standardised cognitive assessment tool increasingly being utilised with children with autism spectrum disorders. It is also commonly used to assess students' cognitive capacity and aid in school planning. The DAS-II has a low item floor and an enlarged ceiling, allowing for adaptive testing in preschoolers or toddlers with potential deficits (especially in language). Furthermore, the DAS and DAS-II have been used to diagnose learning problems by determining processing style and doing an ability-achievement discrepancy analysis, both of which allow for more targeted intervention planning. Despite the popularity of the DAS and DAS-II as cognitive assessments for children with learning impairments or autism, their application in groups of children with hearing loss has not been independently validated.

The test assesses receptive and expressive language skills, nonverbal reasoning, and spatial abilities. The DAS-II has good test-retest reliability (> 0.73 across all index and composite scores), great internal consistency (intercorrelations of 0.84 between the index and composite scores), and good convergent validity when compared to the Weschler series tests and the Mullen Scales of Early Learning. The Nonverbal Reasoning Cluster ($r = 0.65$) and the Spatial Ability Cluster ($r = 0.67$) of the DAS have moderate associations with the WISC-III Performance IQ in students with learning difficulties. Table 3 illustrate the comparison of speech analysis tools with accuracy and other significant information.

3.1.2 Technology

In both developed and developing countries, smartphones and tablets have become increasingly accessible to children, forming a part of their daily lives. Approximately 88% and 79% of Australian households with children aged 15 and below living in major cities and rural areas, respectively, have fast and stable internet connections [44]. The statistics also show that 94%, 85%, and 62% of households access the internet via desktop or laptop computers, mobile or smartphone, or tablet, respectively. Although computer and mobile-based speech analysis techniques are not commonly used in children with SSD, they possess capabilities to access easily accessible, affordable, and objective speech assessment tools and interventions [46]. The development of such computer and mobile-based tools will likely enhance the efficiency of medical practitioners who deal with children with SSD and reduce their caseloads while also increasing accessibility and practice intensity due to reduced barriers resulting from the elimination of the face-to-face SLP [27].

Despite recognizing that early detection and treatment of communication disorders is critical for school readiness and has been shown to significantly improve communication, literacy, and mental health outcomes for young children, nearly 40% of children with speech

Table 3 Comparison of speech assessment tools

| Sl no | Tools | Target age | Features | Accuracy | Limitations/Challenges |
|-------|--|---------------|---|---|--|
| 1 | Dynamic Evaluation of Motor Speech Skill (DEMSS) | 3 and more | To assess children who have a significant speech impairment, especially reduced phonemic and/or phonetic inventories, vowel or prosodic errors, poor speech intelligibility, and/or little to no verbal communication. | 89% for test-retest reliability | DEMSS does not overdiagnose CAS but sometimes fails to identify children with CAS |
| 2 | Motor Speech Examination | 3;4–8;11-year | Identify deviant speech characteristics (vowel prolongation, contextual speech, stress testing, respiration, phonation, articulation, resonance, and prosody.). | Alternate Motion Rates (AMR) tasks among children with ranged from 0–2 (accuracy) and Sequential Motion Rates (SMR) tasks were 8 (accuracy), and the range of scores among children with DS was 0–6 | Intelligibility measurement |
| 3 | Tabby Talks | 4–16 years | Prosodic, voice quality and pronunciation to improve intelligibility classification | Pronunciation verification accuracy of 88.2% at the phoneme level and 80.7% at the utterance level. | limited application in speech therapy due to the difficulties in working with children's speech. |
| 4 | LEST | 0–6 Years | Screening tool for language delay | screening tool for language delay highly Statistically Significant, ($p < 0.005$) | In two items, delay is taken, and the prevalence decreases. |
| 5 | Battery of Western Speech and Language Assessment Tool | 5–14 years | | | |
| 6 | CHOCSLAT | 1–4 year | Grammar assessment section and concentrates on the grammatical classification and meanings of children's statements. | 87% accurate | |
| 7 | CELF-4 | 5–21 years | Overall core language; receptive and expressive language; language content; language structure; and working memory indices; various levels of the linguistic rule system; metalinguistic skills; phonological awareness | 57% accurate | More than one test to arrive at a diagnosis and also for using cross-battery testing |
| 8 | PLS-5 English | 0–7 years | Total language, auditory comprehension, expressive communication standard | 80–89% accuracy | |

Table 3 (continued)

| Sl no | Tools | Target age | Features | Accuracy | Limitations/Challenges |
|-------|------------|-------------|--|-------------------------------|---|
| 9 | GFTA-3 | 2–21 years | scores, growth scores, percentile ranks, and language age equivalents Sounds-in-Words Section, varied for Stimulability sections | At least 85% accuracy | Concurrent validity was not determined for ages 6;11–7;11. (the diagnostic accuracy of the PLS-5 is insufficient) Limits the Core Language Index |
| 10 | Bayley-III | 1–42 months | Diagnosing developmental delays in early childhood | 83.84% accuracy | Children older than 42 months diagnosis |
| 11 | DAS -II | 2–17 years | Verbal (Gv), Nonverbal Reasoning (Gf), Spatial (Gv), Working Memory (Gsm), Processing Speed (Gs), Phonological Processing (Ga), Recall of Objects (Glr), and General Conceptual Ability (GCA), which is a measure of the general factor g. | 0.79 to 0.94 test-reliability | Range limitations at each level are expected |

and language disorders do not receive appropriate intervention because their impairment goes undetected. The predominant tool for clinical assessment of aberrant speech is auditory perceptual analysis (APA); however, APA outcomes are subject to intra- and inter-rater variability. Another consideration is that some children may be hesitant to participate in lengthy testing sessions, and even if they do, transcription of big data sets of audio recordings is time-consuming and needs therapists with a high level of skill. Because of the constraints of manual or hand transcription-based diagnostic evaluation approaches, there is a growing demand for automated methods to quantify kid speech patterns rapidly and reliably, allowing them to be diagnosed whether they have impaired speech [80].

Moreover, such approaches are likely to improve the child's motivation to participate in and study exercises since they perceive them appealing, including audio prompts, reinforcers, or animation, encompass speech recording, playback responses, live manipulation of gameplay and stimuli, and prerecorded models. Nonetheless, the ASA tools that utilise diagnostic or therapeutic software are supposed to match reliability standards applied to human raters for them to be viable [44]. According to McKechnie et al. (2018), the Commonly accepted percentage agreement criteria for perceptual judgments of speech between two human raters or outcome reliability across two separate assessments of the same behavior range from 75 to 85%. Despite the extensive work on ASR, little work has been reported on developing speech therapy tools with ASR capabilities for use in paediatric speech sound disorders such as CAS. Although automated system is working with 80% accuracy, further work is needed to train automated systems with larger samples of speech to increase accuracy for assessment and therapeutic feedback. Therefore, ASA tools should meet the 80% threshold of reliability to be considered viable for speech assessment in children with SSD.

3.1.3 Protocols

Protocols are the norms and procedures for assessing speech and language using instruments. Technical specifications for data acquisition, voice and speech tasks, analysis methods, and results for instrumental evaluation of voice/speech production are all included in the protocols. Even though these types of assessments are performed regularly at many research and clinical facilities in the United States, a lack of standardised procedures/protocols currently limits the extent to which the results can be used to facilitate comparisons across clinics and research studies to improve the evidence base for the management of voice disorders. The recommended protocols aim to produce a core set of well-defined measures that can be universally interpreted and compared using instrumental approaches. These recommendations are not intended to preclude the use of additional measures or protocols that individual clinics/clinicians or researchers believe are useful in evaluating vocal function.

MSAP – Madison Speech Assessment Protocol The Madison Speech Assessment Protocol (MSAP) was developed to cater to the need to diagnose speech and language disorders in the United States. The protocol employs 17 speech-related and eight motor and language activities and tasks in a 25-measure battery with a 2-hour run time in various clinical, educational, and research programs [75].

Connected Speech Transcription Protocol (CoST-P) A clinically feasible protocol is connected speech transcription for children suffering from Apraxia. This development protocol's main reason is to assist children aged 6–13 in describing their connected speech. The connected speech can be evaluated to pick up the independent and relational analyses [8].

Trivandrum Development Screening Chart (TDSC) The TDSC (0–6 y) is a 51-item screening test created from existing developmental tools and has been validated for children up to the age of six. The TDSC is a straightforward, reliable, and valid screening tool for identifying children with developmental delays in the community. The Child Development Centre, SAT Hospital and Medical College, Trivandrum, conceived and developed it. The ranges for each test item were derived from the Bayley Scales of Infant Development standards (Baroda norms). The sensitivity and specificity of a TDSC chart with one item delay were 84.62% and 90.8%, respectively [69].

Ages and stages questionnaire test The Ages & Stages Questionnaires are a developmental screening tool that measures developmental progress in children aged one month to five and a half years. The ages and stages questionnaire was designed to help health professionals and teachers who handle young children identify speech deficits in their patients. The tool relies on parents' information about their children to detect speech deficits and other critical milestone delays [87]. Its popularity is due to its parent-centred approach and intrinsic ease of use, which has made it the most extensively used developmental screener in the world. Evidence demonstrates that the earlier a child's development is examined, the more likely they are to fulfil their full potential. Arabic, Chinese, English, French, Spanish, and Vietnamese tests are accessible. It also takes parents 10–15 minutes to complete, and professionals 2–3 minutes to grade and highlight a child's strengths and issues. The ASQ exam is used by programmes all over the country because it is highly valid, dependable, and accurate, as well as being cost-effective, easy to score in minutes, and well researched and tested with a varied sample of children. ASQ is a fun and engaging method to collaborate with parents and make the most of their expert knowledge.

The caterpillar novel reading passage The existing approaches, methods, and materials of speech assessment used by clinicians are affected by limitations in validity and reliability [37]. The importance of motor speech evaluation is that it enables the diagnosis of speech impairment and further reveals the disorder's severity [38]. The assessments' outcomes are critical in identifying the salient elements of speech production targeted for intervention to enhance communication effectively [61]. Therefore, motor speech assessment tools are critical since they reveal the degree of speech impairment among children with SSD.

Contextual speech is the most significant speech assessment activity [61]. Reading the passage provides clinicians with valuable information compared to scores assigned through syllable and word repetition exercises. The passage is designed to present a controlled and repeatable activity in speaking, gauge the speech production system and conduct a differential diagnosis. The evidence shows that pediatricians can diagnose speech and language disorders in children by reading a passage.

The My Grandfather was the most famous speech assessment passage joined by Van Riper in 1963 [61]. The passage is ill-suited to examine speech motor skills to differentiate the severity and type of motor speech disorder [61]. The author of the passage, Van Riper, concurred with the fact mentioned above when he described the tool as useful for a quick survey of the student's (client's) ability to produce correct speech sounds [61]. The seminal work of Darley et al. in 1969 is seen as the historical root of the usage of the "My Grandfather" passage in speech and motor assessment on the perceptual traits of dysarthria. Therefore, Van Riper created the "My Grandfather" passage to assess speech and sound recognition among children.

“The Caterpillar” reading novel passage was developed to systematically enhance the “My Grandfather” passage by incorporating activities that evaluate deficits within and across speech subsystems [61]. To observe the variations between connected and isolated speech performance, embedding the word and syllable repetition activities into the passage is recommended as a best practice for evaluating motor speech disorders. Additionally, the reading passage offers a chance to perceive the motor speech’s performance on exercises that cannot be evaluated in isolation, such as prosodic modulation. Therefore, researchers have an opportunity to assess various speakers’ speech performance through the use of a reading passage as a speech assessment tool.

4 Discussion

A number of reviews on speech assessment are available in the literature, of which those with a detailed discussion on the methods for the assessment are less.

A review published in 2012 summarised the findings on speech production issues in people with Down syndrome (DS) to enhance therapeutic services and guide future research in the field [36]. In their work, the authors selected one speech impairment disorder. Another review article was published in 2013 that aids in determining the interventions for preschool children according to the circumstances utilising a practice-based model of interventions to select the intervention subgroups [1]. Though the paper included studies from January 1980 to November 2011, it focused only on the interventions.

In 2014, a literature review was published to analyse the elements contributing to the debate over describing and diagnosing CAS and examine a therapeutically relevant body of knowledge on CAS diagnosis [7]. The work entirely focused on CAS over the 10 years. Broome et al. conducted a systematic review in 2017 intending to provide a summary and assessment of speech examinations used in children with autism spectrum disorders (ASD). Later, a narrative review was reported to determine the essential components of an evidence-based paediatric speech assessment, combined with the systematic review findings, giving clinical and research guidelines for best practice [15]. The review was published with the research articles published between 1990 and 2014, assessing children’s speech only with ASD.

Another review published in 2018 by Wren et al. aimed to assess the evidence for therapies for SSD in preschool children and categorised them under a classification of interventions for SSD [90]. The intervention studies published up to 2012 were selected for the work. In 2018, a systematic search and review of the published studies on the use of automated speech analysis (ASA) tools for analysing and modifying speech of typically the developing children learning a foreign language as well as children with speech sound disorders were conducted to determine the types, attributes, and purposes of ASA tools being used. The performance of the therapeutic tools and their comparison with the human judgement was also included [44]. The research articles published between January 2007 and December 2016 were selected for the study.

Low et al. reported a systematic review in 2020 on voice for automated assessments across a more extensive range of psychiatric diseases [42]. According to the authors, speech processing technology could aid mental health assessments, but several barriers exist, including the need for extensive transdiagnostic and longitudinal investigations. The work concentrated on analysing psychiatric disorders and collected studies from the past 10 years that employ speech to identify the presence or severity of mental disorders. In 2021, another review

was published to summarise and evaluate oral sensory problems in children and adolescents with ASD [18]. A systematic search was reported in the work with the published articles from January 2000 to December 2018, concentrating entirely on ASD. Additionally, the review suggests that oral stimulation employing speech-sensory technologies may be necessary.

The present systematic literature review aimed to identify, categorize, and compare the effective speech assessment methods for analysing multiple speech disorders in children, instead of choosing only a particular disorder or speech analysis tool as observed in the existing reviews. A statistical analysis of the reported speech impairment assessment methods, protocols and case studies from the last 12 years has been included. We have also covered the state-of-the-art solutions with the level of accuracy of each tool and their contribution to the research in the field of interest.

4.1 Application of speech assessment tools for speech impairment analysis

4.1.1 CAS disorder

Different research groups have reported adopting multiple tools for the analysis of CAS. Table 4 shows the studies reported in the last decade using corresponding tools utilised.

Strand et al. used DMESS to analyse speech and prosody's motor function for children aged 3–6 years and seven months to diagnose childhood apraxia [79]. The child performed the stimuli in two ways during this protocol's application: an initial attempt and after the examiner's demonstration. The proof of construct validity and reliability presented as intra-judges' 89%, inter-judges 91%, and test-retest 89%. However, positive and negative risk ratios, sensitivity, and specificity measurements showed that CAS was not over-diagnosed by DEMSS, though children with CAS were not detected in a few cases.

In 2013, Preston et al. conducted a study on ultrasound imaging assessment and treatment on CAS [66]. The research explored the efficacy of a treatment program for children with severe speech sound errors associated with childhood speech apraxia involving ultrasound

Table 4 Studies Reported on Childhood Apraxia of Speech Disorder

| Reference | Tools | Age | Target Population | Region |
|---------------------|---|--|-------------------|-------------------------|
| Strand et al. [79] | Evaluation of Motor Speech Skill (DEMSS). | 3–6 years and seven months children | 81 | Columbia (English) |
| Preston et al. [66] | Ultrasound images. | children between 9–15 years | 6 | North America (English) |
| Barrett et al. [8] | Connected Speech Transcription Protocol (CoST-P). | children aged 6–13 years | 12 | Sydney |
| Terband et al. [82] | Acoustic models. | 3- to 6-year-old and 6- to 10-year-old | 8 and 10 | - |
| Shahin et al. [73] | Technology-based therapy tools (Tabby Talks). | 4–16 years; 4–10 years | 2; 705 | Qatar |
| Murray et al. [53] | Multivariate discriminant function analysis. | 4–12 years | 72 | Sydney |
| Abdou et al. [3] | Test battery | - | 70 | Arabic |

biofeedback. Diagnostic ultrasound imaging has, for many decades, been a popular instrument in medical practice, and it offers a healthy and productive way to visualize internal structures of the body. Children are cured of altering their gestures by using real-time ultrasound images to provide visual feedback. A multiple baseline experiment in 18 sessions was conducted in the study by six children between 9 and 15 years of age during therapies centered on developing lingual sound sequences. Even though this study achieved about 80% accuracy, cost, access and training with this technology might limit the implementation of this tool in clinics.

CoST-P utilised CAS in the case study on 12 children aged 6–13 years [8]. The participants' related speech parameters were selected to obtain independent and relational analyses. The usage of CoST-P to represent CAS speech characteristics was related to associated speech features. Children with CAS had their connected speech transcribed using the CoST-P. With appropriate reliability and fidelity scores, the CoST-P can be employed in researching children's connected speech transcription of 50 utterances and takes between 5 and 7 h per child (including orthography, target output, and actual production). Because of the time burden, the current CoST-P is used infrequently in speech-language pathology practice. Even though the tool is an adequate resource for speech-language pathologists and clinical researchers, its usage is challenging.

Terband et al. conducted a study in 2019 to assess CAS by using objective measurement techniques for 3- to 6-year-old ones [82]. The analysis has made considerable progress regarding the clinical criteria for diagnosing childhood speech apraxia (commonly described as a speech motor planning or programming disorder) in recent years. For participant selection purposes, three segmental and supra-segmental speech features, i.e., error inconsistency, lengthened and interrupted co-articulation, and improper prosody has gained broad acceptance. Few researchers have also attempted to assess the validity of these features empirically. The fact that none of these features operationalized is a fundamental challenge for analytical analysis.

In 2015, Shahin et al. did a study explaining the pipeline to detect speech processing CAS-related common errors [69] automatically. It is used for children within the age group of 4–16 years. The device achieves an accuracy of pronunciation tests of 88.2% on phoneme and 80.7% on utterance stages, with a classification of lexical stress of 83.3%. Murray et al., in 2015, did a study to establish a variety of objective measures to distinguish CAS from other speech disorders, i.e., multivariate discriminant function analysis [53]. It involves syllable segregation, matched lexical stress, proper phonemes percentages from a polysyllabic image-name task, and precise articulatory repetition. It reported that the discriminant functional analysis model had achieved 91% accuracy by expert diagnoses. Twenty-eight children met two sets of CASs diagnostic criteria; 4 other children met the CAS criteria' comorbidity. The researcher used the combination of the best-expected expert diagnoses for Multivariate Discriminating Feature Research.

Abdou et al., to identify the possible presence of CAS in Arabic-speaking children, developed a test battery, thus allowing the planning of appropriate therapy programs [3]. Seventy monolingual Arabic-speaking Egyptian children, including ten children with suspected CAS, 20 children with phonological disorders, and 40 typically developing children, were given the built-up test battery for CAS. The study concluded that the built-in test battery for CAS diagnosis is a reliable, valid, sensitive instrument that can be used to detect and differentiate between the presence of CAS in Arabic-speaking children and phonological disorders.

4.1.2 SSD and SLD

SSDs and SLDs are mostly seen in children. In some cases, their cause remains yet to be discovered or detected earlier. With the help of verbal tests, screening tests, instruments, and scales and with some tools and techniques, these disorders can be assessed and help clinicians and pathologies in the process of identifying the diseases. Table 5 lists the different styles and methods that can be used, not only for better assessment but also for therapy necessity among children with speech and language disorders.

In 2010, Shriberg et al., to identify diagnostic markers for eight subtypes of SSDs of unknown origin, developed MSAP [75]. Unlike other existing tools, the tool is not intended only to identify speech Apraxia but also for SSDs. In addition to its presentation, the protocol was also used to study different age groups and was designed to include a description of a classification system for motor speech disorders. Due to the significant prevalence of SSDs in public, Shriberg et al. did another investigation with MSAP to investigate the prevalence and phenotype of CAS in patients with lactose intolerance, albeit much information is absent from the literature. The results showed a high prevalence of the disorder in the investigated sample. Eight of the 33 respondents (24%) reported meeting the current CAS diagnostic criteria. Ataxic or hyperkinetic dysarthria criteria were seemed to be completed by two participants, 1 of whom was among the 8 with CAS. Group results for the remaining 24 respondents were consistent with a classification category called Motor Speech Disorder-Not Specified Otherwise. Here, both the evidence of validity and liability were nil.

In 2012, Carter et al. provided an approach to advancing children's speech and language evaluation methods, using the morbid results of extreme falciparum malaria research as a guideline [17]. They chose children exposed to severe malaria to test tools for children with language disabilities. Other causes of language impairment may have features that are not readily available through this adaptation process, such as the impact of social communication on language assessment. The final battery- 'speech-language assessment tool' consisted of seven assessments: (1a) receptive language (original estimate changed to an adaptation of the Grammar Reception Test), (2b) syntax (new score system adapted from the Renfrew Action Picture Test), (3) lexical semantics (minor changes to the original), (4) higher-level language (significant changes to reduce the number of different items and increase the number of questions per item), 5) test of word-finding and language-specific test (a new assessment based on the Test of Word Finding), 6) Pragmatics profile of everyday communication skills in children, 7) Peabody picture vocabulary Test.

Nelson et al. conducted a study for using transcription in assessing speech disorders in children [54]. This research analyzed transcription, facilitators, transcription use issues, and detailed transcription discrepancies with different clients' groups. Transcription charts (81%), self-practice (68%), and blogs were the three most frequently identified strategies/resources (42%). The use of two vowel notation systems, diminished transcription abilities, problems with service delivery, sampling/recording problems, and transcription to communicate were transcription challenges. This study reported that when recording children's speech with childhood speech apraxia and craniofacial impairment, participants use detailed transcription more often than transcription to record children's addresses with SSD of unknown origin.

Mehta et al., in 2015, presented an update on ongoing work using a miniature accelerometer on the neck surface below the larynx to collect a large set of outpatient data on patients with hyper-functional voice disorders (before and after treatment) and matched-control subjects [48]. Three types of analysis approach were employed to identify the best set of differentiating

Table 5 Studies that Investigated Speech Sound Disorders and Speech-Language Disorders

| Reference | Tools | Age | Target Population | Region |
|----------------------------|--|------------------------------|--|-------------------------------|
| Wallace et al. [87] | Screening test. | 5-year-old and younger | - | - |
| Shriberg et al. [75] | Madison Speech Assessment Protocol or MSAP – 15. | 3–9-year-old | - | - |
| Carter et al. [17] | Peabody Picture Vocabulary Test, Pragmatics Profile of Everyday Communication Skills in Children, Renfrew Action Picture Test, Word Finding & language specific tests of lexical semantics, Test for the Reception of Grammar, higher-level language. Phonemic transcription, Phonetic transcription, Narrow transcription, Broad transcription. | Between 6 and 9 years of age | 539 Kenyan children (males=271, females=268) | Kenya |
| Nelson et al. [54] | CHOCSLAT – Chinese healthcare-oriented computerised speech & language. Re-kindergarten NOMS, k-12 NOMS. | Children | 84 | Australia |
| Towey et al. [83] | Phonological processes | 1–4 years | - | China |
| Mullen and schooling. [52] | “The Caterpillar” passage. Face-to-face workshop transcript | Pre-schoolers | 16,000 | - |
| McLeod et al. [46] | Online discussion artifacts. | 4–5 years | 109 | china |
| Patel et al. [61] | Phonological patterns. | 4 to 71 years | 22 | - |
| McLeod et al. [45] | DAPPLE (Dynamic Assessment of Pre-schoolers’ Proficiency in Learning English) | 4 to 5 years | 1097 | Australia |
| Hasson et al. [29] | PCC and PWP | 3–5 years | 26 | (English, French and Spanish) |
| Newbold et al. [55] | Epidemiological study | 4 & 6 years | 4 | English |
| Eadie et al. [24] | Diffusion-weighted imaging | 4 years | 1494 | Australia |
| Morgan et al. [50] | Reliable and valid instrument | 9–11 years | 41 | - |
| Zarifian et al. [91] | Phonological intervention | Between 36 and 72 months | 387 | Persia |
| Jesus et al. [34] | Nordic orofacial test-screening (not-s) | 6–17 years | 22 | - |
| Mogren et al. [49] | Developmental Quotient Score | Less than 72-month-old | 61 | - |
| Chong et al. [19] | | | 91 (67 boys and 24 girls) | - |

measures between hyper-functional and standard vocal behavior patterns: (1) ambulatory voice measurements, including vocal dose and voice quality correlates; (2) aerodynamically metric measures, which are based on glottal airflow estimates derived from the specified accelerometer signal and; (3) classification of other physiological signal recordings based on machine learning and pattern-recognition approaches, which were successfully used in analyzing long-term recordings.

In 2010, Mullen and schooling focused on the data collected from prekindergarten NOMS (National Outcomes Measurement System) and K-12 NOMS in school settings [52]. The primary objective was to serve as a data source for speech-language pathologists who were called upon to provide empirical evidence of the functional results of their clinical services to children and adult patients with different speech-language pathologies. The 2 NOMS components had reported studying more than 2,000 preschool students and 14,000 K-12 students by SLPs working in school settings. In 2013, McLeod et al. conducted a study to describe the speech of preschool children identified by parents/teachers as having difficulty “talking and making speech sounds” and to compare the speech characteristics of those who did not have access to SLP services [46]. The method of the study includes Stage 1: assessed documented parent/teacher concern about the speech skills of 1,097 children in the year 4- to 5- attending early childhood centers, Stage 2a: 143 children identified with problems, and Stage 2b: parents have returned questionnaires about service access for 109 children.

Towey et al. conducted a study in developing a diagnostic profiling tool for healthcare professionals to identify the potential problems of Chinese-speaking children with speech and language development [83]. The instrument aimed to provide a technical breakthrough to help kids with speech impairment or language delay. The case study was carried out in different stages, from 1 to 4 years. However, the exactness and specificity offered by the IPA are lacking. Due to data availability limitations, text output from the speech-to-text API is not always an accurate transcription.

The caterpillar passage study conducted by Patel et al. in 2013 describes the passage as an assessment tool or protocol to provide specific tasks aimed at informing the assessment of motor speech disorders with a contemporary, easy-to-read, contextual speech sample [61]. To demonstrate its usefulness in examining motor speech performance, twenty-two participants, 15, were recorded reading the passage “The Caterpillar” with DYS or AOS and 7 healthy controls (HC). Performance analysis across a subset of segmental and prosodic variables showed that “The Caterpillar” passage showed promise to extract individual impairment profiles that could increase current evaluation protocols and inform motor speech disorder therapy planning.

Hasson et al. conducted a DAPPLE study (Dynamic Assessment of Pre-schoolers’ Proficiency in Learning English) in 2013 [29]. To examine the ability of children to learn vocabulary, sentence structure, and phonology, the evaluation used a test-teach-test format evaluation, which takes less than 60 min to perform, given to 26 bilingual children: 12 currently on a caseload of speech and language therapy, and 14 children matched by age and socioeconomic status who never referred to speech therapy and language therapy. Qualitative analysis of individual children’s performance on the DAPPLE suggested that it can discriminate against core language deficits from the difference due to a bilingual language learning context.

In 2013, Newbold et al. compared a range of commonly used procedures for perceptual phonological and phonetic analysis of developmental speech difficulties to identify the best ways to measure speech changes in children with severe and persistent language difficulties

(SPSD) [55]. Speech output measures included the percentage of whole words correct (PWC), correct consonant percentage (PCC), total word proximity proportion (PWP), analysis of phonological patterns (process), and phonetic inventory analysis. The study was conducted on 4 SPSP children, registered at 4 years of age and again at 6 years of age, who perform naming and repetition duties.

Eadie et al. conducted a study to assess the prevalence of idiopathic sound speech, the comorbidity with language and pre-literacy difficulties of language sound disorders, and the factors contributing to the speech outcome for 4 years [24]. 1494 participants completed 4-year voice, language, and pre-literacy evaluations from an Australian longitudinal cohort. In four areas: child and family, reported parental speech, cognitive-linguistic, and motor abilities, the logistical regression examined SSD predictors. Early 4-year SSD detection should focus on family variables and 2-year language and motor skills measurement.

Morgan et al. conducted a study in 2018 to (i) test for the hypothesis that neurostructural difference in autism spectrum disorder (ASD) and CAS compared to typically developed (TD) is demonstrated by morphometric MRI measurements (ASD vs. TD and CAS vs. TD), (ii) investigating early possible diseasing-specific patterns of the two clinical groups (ASD vs. CAS) for the brain, and (iii) evaluating the machine-learning predictive strength of ASD, CAS, and TD [50]. T1-weighted brain MRI scans of 68 children (age range: 34–74 months) were analysed and divided into three cohorts: (1) 26 ASD children (mean age \pm standard deviation: 56 ± 11 months); (2) 24 CAS children (57 ± 10 months); and (3) 18 TD children (55 ± 13 months). In the ML analysis, the differences between ASD and TD children in brain characteristics were significant, while only some CAS classification trends were detected compared with TD peers.

The aim of the study conducted by Zarifian et al. was to adapt the articulation assessment, subtest the articulation, phonology diagnostic assessment, and determine its reliability and validity for Persian-speaking children [91]. The Persian version of the articulation assessment (PAA) was administered to 387 children between the ages of 36 and 72 months, with $M(SD)$: $53.7 (\pm 10.1)$ per month following the adaptation process. The study included test-retest reproducibility, score-rescore consistency, and validity evaluation through content, convergent, and discriminative validity to establish the instrument's psychometric properties. The mean scores for articulation disorders were significantly lower than those for normal children in the Persian Articulation Assessment, showing discriminative validity ($t = 7.245$, $df = 34$, $P < 0.001$). The study concluded that it is suggested in the Persian version of Articulation Assessment as a reliable and valid tool for assessing articulation skills in Persian-speaking children.

In 2019, Jesus et al. experimented on the efficacy of a modern tablet-based approach to phonological intervention and compared it to a conventional tabletop approach targeted at children with speech sound problems based on phonology (SSD) [34]. Twenty-two children with phonological SSD were randomly allocated to 1 out of 2 assessments, tabletop, phone, and evaluation based upon similar activities (11 children in each group), with delivery being the only difference. The same speech-language pathologist treated all children over two blocks of 6 weekly sessions for 12 intervention sessions. The findings provide new evidence concerning using digital materials in children with SSD to improve speech.

A study was conducted to investigate, describe, and analyze the characteristics of speech, intelligibility, orofacial function, and co-existing neurodevelopmental symptoms persisting after six years of age in children with SSD of unknown origin [49]. They concluded that the children with persistent SSD are at risk of orofacial dysfunction, general motor problems, and

other neurodevelopmental disorders, so co-occurring conditions should screen. The study included 61 children of unknown origin with SSD (6–17 years), referred for a speech and oral motor test. Parents completed context Scale Intelligibility (CIS) and a questionnaire containing heredity, health and neurodevelopment, and speech development.

In 2021, Chong et al. took a cross-sectional study in a tertiary center in Malaysia to explore the socio-demographics of children with speech delay [19]. The study was conducted at speech therapy clinics for children with speech delays less than 72 months old. Both speech and other developmental skills were assessed using the Developmental Quotient scores (DQ). There were 91 children in the study (67 boys and 24 girls), 54.9% of whom had a direct speech delay, and 45.1% had neurodevelopmental disorders. The average age was 39.9 months and 11.52 months. The average speech DQ was 54.76%, with a margin of error of 24.06%. Lower DQs in the speech was linked to lower DQs in other skills ($p < 0.01$). There was no significant relationship between screen time for children and parents and DQs of speech and other skills ($p > 0.05$).

4.1.3 Speech Articulation Disorder, Cleft Palate Disorder, Tongue-tie, Childhood Dysarthria, Oral Motor Placement Disorder

Most articulation disorders are SSDs and come under motor speech disorders. Table 6 includes Speech Articulation Disorder, Cleft Palate Disorder, Tongue-tie, Childhood Dysarthria, Oral Motor Placement Disorder studies selected for the review published between 2010 and 2021 to address speech articulation disorder in children specifically.

In 2013, Khattab et al. conducted a study to assess oral impairment levels using standardised questionnaires [37]. Thirty-four Class-I Division-1 patients with malocclusion and moderate upper teeth crowding were randomly distributed into two groups. Seventeen patients in group A were treated with fixed lingual appliances (Stealth®, AO, Sheboygan, Wisc; mean age: 20.6 years; standard deviation [SD]: 2.9 years), whereas 17 patients in group B (mean age: 21.8 years; SD: 3.3 years) treated with conventional fixed labial appliances. Using fricative/s/sound spectrographic analysis, speech performance has been tested before, immediately after (T1), 1 month after, and 3 months after bracket placement.

Wang et al., in 2013, conducted a study on articulatory speech disorder assessment via speech therapy [88]. The research objective was to compare speech therapy's efficacy with functional articulation disorders in two groups of children: those without speech Impairment disorder (SID). There were no major differences statistically between the two groups in age, gender, sibling order, parenting education, and pre-test number of pronunciation errors ($P > 0.05$). After speech therapy assessment ($F = 70.393$; $P < 0.001$) and interaction between pre/post-speech therapy assessment ($F = 11.119$; $P = 0.002$), the results showed significant changes. Speech therapy improved the articulation performance of children with functional articulation disorders, regardless of whether they have SID, but in children without SID, it results in significantly greater improvement. Thus, the assessment efficiency of speech therapy in young children with articulation disorders may be affected by SID.

In 2017, Afshan et al. introduced an automated approach to children's speech clinical evaluations using limited data [4]. Graduate clinicians have assessed the Rhotic sound pronunciation by evaluating words in the GFTA-3 with the letter 'r.' Due to their late acquisition in children; the rhotic sounds were explicitly selected. The remaining kids, used for evaluation, were aligned using the dynamic time to match the five template warping. The difference between both test child's 'r' and model child's 'r' was measured using the cosine

Table 6 Studies that Examined Speech Articulation Disorder, Cleft Palate Disorder, Tongue-tie, Childhood Dysarthria, Oral Motor Placement Disorder

| Speech impairment | Reference | Tools | Age | Target Population | Region |
|-------------------------------|---------------------------|--|------------------------|-------------------|--|
| Speech articulation disorder | Khatab et al. [37] | Standardised questionnaires | - | 34 | - |
| | Wang et al. [88] | Sensory integration dysfunction | 42–72 months | 30 | Southern Taiwan |
| | Mullen and Schooling [52] | Perceptual evaluation | Pre-K and Kindergarten | 14,852 | 37 states |
| | Afshan et al. [4] | Goldman-Fristoe Test of Articulation, 3rd Edition (GFTA-3) | Aged between 6–7 years | 10 | - |
| Cleft palate disorders | Zharkova [92] | Sets of tongue curves and cleft palate articulations. | 6–10 year | - | - |
| | Britton et al. [13] | Evidence-based standards. | 5-year-old | 1110 | United Kingdom (Great Britain and Ireland) |
| Tongue-tie | Pamplona and Yunza [58] | Whole language model (w/lm). | 4–12 years | 48 | Mexico |
| Childhood dysarthria | Ito et al. [9] | Articulation test | 5–115 days | 60 | United Kingdom |
| | Liss et al. [41] | Speech envelope modulation spectra (EMS) | - | 43 | - |
| Oral motor placement disorder | Scholderle et al. [70] | Bogenhausen Dysarthria Scales (bodys). | 3–9 years | 144 | Germany |
| | Al-Qatab and Mustafa [5] | ASR | Children | - | - |
| | Lehner et al. [40] | KomPaS | - | 107 | Germany |
| | Veenhuis et al. [85] | Radboud Dysarthria Assessment | 5–18 year | 15 | Europe |
| Oral cancer impairment | Kayiki et al. [35] | Vowels, subjective tests, objective tests | 11–18 year | 12 | Turkey |
| | Sharma and Singh [74] | Etiopathological characteristics and treatment guidelines. | 12–20 years | 9 | India |
| | [6] | German Speech Intelligibility Index | 18–22 year | 328 | German |

distance. Multiple linear regression is shown on the differential scores to generate well-correlated forecasts with Human Clinical Assessments.

The risk of speech disorder is more for children born with cleft palate. Cleft lip or cleft palate are congenital disabilities that result in the incorrect formation of the fetal lip or mouth during pregnancy. Together, these congenital disabilities are usually known as “orofacial clefts.” Speaking and feeding are difficult in such situations and surgical interventions are required to restore normal scar-free function. Language therapy helps to correct speech problems, if necessary. Zharkova, in 2013, conducted a study to describe ultrasound tongue imagery as a potential tool in cleft palate speakers for quantitative tongue function analysis [92]. The other three steps compare tongue curve sets to quantify tongue displacement dynamics, token-to-token variability in the tongue’s position, and the extent of separation between tongue curves for different sounds of speech.

Britton et al. conducted a study to develop national standards for speech results and care treatment processes for children with cleft palate ± lip [13]. In this large, multi-center, prospective cohort study, 12 cleft centres in Great Britain and Ireland collected speech recordings of 1,110 five-year-old with cleft palate who were involved (born 2001 to 2003). Results were compared against the evidence-based method, speech outcome requirements, and statistical analysis performed. The development of standards facilitated increased reporting of speech and treatment results. To Study whether Tele Practice (TP) intervention/assessment in SLP could efficiently improve the speech performance in children with cleft palate (CCP), Pamplona and Ysunza conducted a study in 2020 during COVID – 19 [58]. There was a significant CA severity improvement at the end of the TP period ($p < 0.001$). The researcher indicates that TP can be a safe and reliable tool for CA improvement. The COVID-19 pandemic would radically alter healthcare services delivery long-term, so studying and implementing alternative service delivery modes.

Ankyloglossia is a congenital condition in which an abnormally short, thickened, or tight lingual frenulum is born to a neonate, limiting the tongues mobility. In 2015, Ito et al. conducted a study to determine the efficacy of tongue-tie division (frenuloplasty/frenulotomy) in children with ankyloglossia for speech articulation disorder (tongue-tie) articulation test [33]. Articulation testing was performed in five children (3-8years) with speech problems with tongue-tie division. A speech therapist interviewed the patients and asked them to pronounce what the picture card showed. Substitution and deletion improved relatively early after the tongue-tie division and progressed to distortion, a form of articulation disorder that is less impaired. Thus, distortion required more time for improvement, and in some patients, it remained a lousy speaking habit.

In 2010, Liss et al. investigated automated analysis of speech envelope modulation spectra (EMS), which quantified speech rhythmicity within specified frequency bands and examined whether comparable results could be obtained [41]. EMS was conducted on sentences produced by 43 speakers with 1 of 4 types of dysarthria and healthy controls. EMS consisted of full-signal slow-rate (up to 10 Hz) amplitude modulations and 7-octave bands ranging from 125 to 8000 Hz in centre frequency. Discriminant function analysis (DFA) determined which sets of predictor variables between groups best discriminated against. For group membership, these variables achieved 84% 100% lassification precision. Dysarthria could be described in acoustic output by quantifiable temporal patterns. EMS shows promise as a clinical and research tool because the analysis is automated and requires no editing or linguistic assumptions.

Paediatric dysarthria is a sound disorder of motor speech that results from neuromuscular weakness, paralysis, or incoordination of the muscles needed for speech production. The child's speech may be slurred or distorted, and speech may vary in intelligibility based on the extent of neurological weakness. There are some well-established therapy and tools for assessing and treating childhood dysarthria. Scholderle et al. conducted a study in 2020 to collect auditory-perceptual data from typically developing children between 3 and 9 years of age on established symptom categories of dysarthria to create age standards for assessing dysarthria [70]. We are used to analysing speech recordings of the Bogenhausen Dysarthria Scales' auditory-perceptual criteria, a standardised German assessment tool for dysarthria in adults. The Bogenhausen Dysarthria Scales (scales and characteristics) cover clinically relevant speech dimensions and assess well-established categories of dysarthria symptoms. Several speech characteristics overlapped with established symptom categories of dysarthria in typically developing children. The results published in the study are a first step towards establishing auditory-perceptual standards for dysarthria in kindergarten and elementary school children.

Al-Qatab and M. Mustafa investigated the acoustic features and feature selection approaches utilised to improve dysarthric speech classification in ASR based on the severity of impairment in 2021 [5]. They used four acoustic features in their study: prosody, spectral, cepstral, and voice quality, as well as seven feature selection methods: Interaction Capping (ICAP), Conditional Information Feature Extraction (CIFE), Conditional Mutual Information Maximization (CMIM), Double Input Symmetrical Relevance (DISR), Joint Mutual Information (JMI), Conditional redundancy (Condred), and Relief. In addition to that, they used Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Artificial Neural Network (ANN), Classification and Regression Tree (CART), Naive Bayes (NB), and Random Forest (RF) as classification techniques in the experiment. They stated their experiment has several merits that add knowledge to the classification of dysarthric speech according to the level of severity like, the research has identified the features that can work in most of the classifiers, looked at the importance of feature selection in the classification of dysarthric speech and it looked at the best combination that gives the best classification accuracy in the classification. But their disadvantages were that they used a small database – Nemour and the other was that they did not adopt the state-of-the-art classifiers such as deep learning.

This study by Lehner et al. in 2021 covers the development of KommPaS, a web-based instrument for assessing communication impairment in dysarthria patients [40] KommPaS (Communication-related Factors in Speech Disorders) allows doctors to crowdsource laypeople to evaluate dysarthric speech samples for communication-related parameters such as intelligibility, naturalness, perceived listener effort, and efficiency (intelligible speech units per unit time). Significant problems about test efficiency, reliability, and validity would be addressed in addition to material influencing variables and the link between the four KommPaS characteristics.

Researchers used the Radboud Dysarthria Assessment in adults (over 18 years old) and the Radboud Dysarthria Assessment in children (5–18 years old) to assess dysarthria, which included observational tasks such as “conversation” and “reading,” as well as speech-related maximum performance tasks such as “repetition rate,” “phonation time,” “fundamental frequency range,” and “phonation volume” in 2021. Twenty-two people (15 children [5–17 years], seven adults [19–47 years], 14 men and eight females; mean age 19 years, SD 15 years 2 months) took part in the study. All subjects had dysarthria, defined by ataxic components in adults and similar uncontrollable movements in youngsters. Dysarthria in ataxia-telangiectasia

is defined by uncontrolled, ataxic, and involuntary movements, which result in monotonous, unsteady, sluggish, hypernasal, and chanted speech, according to Veenhuis et al. They concluded by stating that the Radboud Dysarthria Assessment and the paediatric Radboud Dysarthria Assessment can be used to assess dysarthria in ataxia-telangiectasia.

In 2012, Kayikci et al. conducted a study to evaluate (1) whether Hawley retainers cause speech disturbance and (2) objective and subjective tests the duration of speech adaptation to Hawley retainers [35]. This study included 12 young people aged 11.11 to 18.03 years. Before and after the Hawley retainer application, speech sounds were assessed subjectively using an articulation test and objectively using acoustic analysis. After wearing Hawley retainers, patients showed statistically significant speech disturbances with consonants [ʃ] and [z]. Statistically significant changes were reported to the vowels. In 2018, Mugada et al. conducted a study to evaluate the quality of life for Head and neck cancer patients who received the therapy [51]. The study was conducted for 9 months. The EORTC QLQ-C30 Items (European Organization for Cancer Research and Treatment Quality of Life Questionnaire Core 30) were used, including the H&N-35 module, to evaluate QOL. The contrast of Specific socio-demographic and clinical features with EORTCC domains created between Questionnaire QLQ-C30 and the H&N35 QLQ EORTC. At $p < 0.05$, the significance level was taken.

Sharma and Singh, in 2016, conducted an observational study on squamous cell carcinoma of the pediatric head and neck, which is rare [74]. For assessing clinicopathological characteristics, treatment, and outcome of this emerging problem, obtained data on pediatric head and neck cancer in the younger age group (20 years of age) was used. Nine patients aged 20 years or younger were identified for analysis in this study during the said period. Various parameters were recorded and analyzed for the outcome, such as age, clinical features, clinical stage, and patients' treatment. Further clinical studies need to be conducted to establish etiopathological characteristics and treatment guidelines in this issue.

In 2021, Bachmann et al. conducted a study to adapt the well-known Speech Handicap Index (SHI) to German, test its suitability for assessing the speech-related quality of life, and compare it to the German Voice-Handicap-Index (VHI) to aid in the treatment of oral cancer patients who experience post-treatment speech difficulties. Participants conducted a web-based survey with a 2 (experienced problem: speech/articulation-related vs. voice-related) x 2 (SHI vs. VHI) between-subject experimental design to distinguish between voice and intelligibility deficits and determine the discriminatory ability of the two instruments. They concluded that the German SHI is a more reliable and responsive measure of speech intelligibility and articulation-related quality of life than the VHI.

4.1.4 Cerebral Palsy, Autism Spectrum Speech Disorder, Hearing Loss, Phonology and Articulation, Friedreich Ataxia (FRDA), Aphasia, Epilepsy, Craniofacial Microsomia

Table 7 shows the papers included in the review used in studies investigating cerebral palsy, autism spectrum disorder, hearing loss, phonology and articulation, Friedreich ataxia (FRDA), Aphasia Epilepsy and Craniofacial Microsomia that cause speech impairment in children.

A preliminary language classification system for cerebral paralysis children was suggested and tested in 2010 by Hustad et al. In the laboratory, 34 children with cerebral paralysis were assembled and collected their speaking and language assessment data (CP; 18 males, 16 female) with an average age of 54 months (SD = 1.8) [32]. The study provided preliminary support for classifying CP children's speech and language skills into 4 initial profile groups. To validate the entire classification system, further research is necessary.

Table 7 Studies that Investigated Cerebral Palsy and Autism Spectrum Disorder, Hearing Loss, Phonology and Articulation, Friedrich Ataxia (FRDA), Aphasia, Epilepsy, and Craniofacial Microsomia

| Disorders | Reference Author | Tools | Age | Target Population | Region |
|----------------------------|------------------------|--|------------------------|-------------------|---|
| Autism spectrum disorder | [32] Hustad et al. | Speech and language assessment system | 4-year-old | 34 | - |
| Down syndrome | [62] Pejovic et al. | Visual orientation task and Audiovisual speech processing task | 5–7 month | 31 | Portugal |
| Cerebral palsy | [63] Pennington et al. | Viking speech scale. | Mean age of 6.03 years | 139 | Lisbon and Oporto, Madrid, Estem sweden |
| Hearing loss | [25] Erimer et al. | Speech intelligibility assessment | 2–15 year | 44 | - |
| | [2] Abaskohi et al. | ASR | Preschool | | Tehran |
| | [27] Fulcher et al. | AV intervention for language development. | 3–5 year | 94 | Australia |
| | [30] Hochmuth et al. | Spanish matrix sentence test. | 21–27 year | 68 | Germany |
| | [65] Phillips et al. | Leiter International Performance Scale-Revised (Leiter-R Brief IQ) and Differential Ability Scales – Second Edition (DAS-II Nonverbal Reasoning Index) | 3–6 year | 54 | - |
| Phonology and articulation | [56] Ng et al. | CUCHILD | 3–6 year | 1986 | Hong Kong |
| Friedreich ataxia (FRDA) | [12] Brendel et al. | Oral diadochokinetic Sustained vowel productions. | 15–22 year | 20 | Germany |
| Aphasia | [64] Pettilo et al. | Progressive Aphasia Severity Scale (PASS) | 10 years | 18 | Italy |
| Epilepsy | [57] Overvliet et al. | Clinical evaluation of Language Fundamentals, CELF | 8–14 years | 25 | Netherlands |
| Craniofacial microsomia | [76] Speltz et al. | Bayley-III - Bayley Scales of Infant and Toddler Development-Third Edition or Preschool Language Scales-Fifth Edition | 12–24 months | 108 cases | Los Angeles, Chicago |

This study compared Down syndrome (DS) and TD infants between the ages of 5 and 7 months in a visual orientation test as well as an audiovisual speech processing task, which examined infants' gazing patterns to communicative signals (i.e., face, eyes, mouth, and waving arm) by Pejovic et al. in 2021 [62]. The study found that DS infants' early visual attention and audiovisual speech processing may be disrupted, with implications for their communication development, suggesting new options for early intervention in this clinical population. According to the findings, DS newborns orient their visual attention slower than TD infants. Both groups focused on the eyes rather than the mouth and the face rather than the waving arm. Furthermore, the findings of this research imply that DS children may require more time to detect/attend to communicative cues in face-to-face communication and that caregivers should emphasize face-to-face communication as a way of training attention to communicative cues from an early age.

The evolution of a scale would classify children's speech performance for use in brain paralysis monitoring registers by Pennington et al. Its reliability across raters and over time analyzed [63]. Cerebral paralysis speech of 139 children (85 boys, 54 girls; mean age 6.03 years, SD 1.09) were classified from the observation and prior knowledge of the children from their language therapist and speech therapists, parents, and other health professionals. Another group of health professionals also rated children's speech from the data in their medical notes. Instead, it asked to assess the scale's simplicity to use, and the scale used Likert scales to describe the child's speech production. More than 74% of raters reported the scale easy or relatively easy to use; 66% of parents and more than 70% of health care professionals judged the scale to describe children's speech well or very well. The Viking Speech Scale was a reliable tool for describing the speech performance of children with cerebral paralysis by observing children or reviewing case notes.

Ertmer et al. investigated children with hearing loss to determine whether scores from a commonly used word-based articulation test are closely associated with speech intelligibility [25]. GFTA – II and 10 short sentences produced words from 44 children with hearing losses. Correlations between 7 word-based predictor variables and percentage-intelligible scores derived from the hearer judgment of stimulus phrases performed. However, regression analysis revealed that the variability in intelligibility scores accounted for no single variable or multivariable model predictor for over 25%.

In 2010, Florian Stelzle et al. conducted a study to introduce and validate a computer-based speech recognition system (ASR) for automatic speech evaluation after dental rehabilitation in edentulous patients with complete dentures [78]. 28 patients twice recorded reading a standardised text - with and without their complete dentures in situ—the speech quality measured by the percentage of the word accuracy (WA) by a polyphone-based ASR. The wearing of complete dentures, on the other hand, considerably increased the WA of the edentulous patients. The reconstitution of speech production quality is essential for dental rehabilitation and can be improved by complete dentures for edentulous patients. The ASR proved a helpful, practical, and easily applicable tool for an automatic speech evaluation in a standardised way.

Fulcher et al. conducted a study in 2012 to check whether a homogeneous cohort of early identified children (approximately 12 months) with all severities of hearing loss and no other concomitant diagnoses could not only significantly outperform a similarly homogeneous cohort of later identified children (> 12 months and < 5 years), but also achieve and maintain age-appropriate speech/language outcomes by 3, 4 and 5 years of age [27]. The children had attended the same program of oral auditory-verbal early intervention. Standardized speech/language assessments performed at 3, 4, and 5 years of age typically developing hearing

children. The previous children identified have significantly outperformed the late children identified at all ages. 93% of all early identified participants scored for speech within normal limits (WNL) by 3 years of age; 90% were WNL for vocabulary understanding, and 95% were WNL for speech production.

Hochmuth et al. carried out a case study on a new Spanish noise sentence test to develop, optimise, and evaluate [30]. The trial included a fundamental matrix of 10 names, verbs, numerals, names, and adjectives. This matrix is used for test lists of 10 sentences of the same syntactic structure, containing the entire language material. The speech material was the distribution of phonemes in Spanish. Independent measures to examine the training effects, comparability of test lists, open-set vs. closed-set test format, and listeners' performance from various Spanish varieties were conducted and assessed. In total, 68 normal-hearing native Spanish-speaking listeners were selected. No significant differences indicate that the test applies to Spanish and Latin American listeners for listeners of different Spanish varieties.

A study was conducted by Phillips et al., in a group of children who are deaf or hard-of-hearing to test the concurrent validity of the Leiter International Success Scale-Revised (Leiter-R Brief IQ) and Differential Ability Scales-Second Edition (DAS-II Nonverbal Reasoning Index) [65]. The participants included 54 children between the ages of 3 and 6 with permanent bilateral hearing loss. The mean values in the two assessments did not vary significantly. Hearing loss severity is not linked to the nonverbal IQ of either the Leiter-R or the DAS-II. Almost a quarter of the assessed children had significant intra-individual differences.

In 2020, Ng et al. described the design and development of CUCHILD, a Cantonese corpus of child speech evaluation tool, on a large scale [56]. The corpus includes words from 1,986 children between the ages of 3 and 6 years. 130 words with 1 to 4 syllables in length had in the speech materials. Speakers cover children with speech disorders, TD, and those with other speech disorders. The aim is to provide corpus support for scientific, clinical, and technological research relating to child speech evaluation. The corpus' design is described in detail, including word selection, recruitment of participants, data acquisition process, and data pre-processing.

A cardinal feature of FRDA is dysarthria, which often leads to severe impairments in daily functioning. However, its precise characteristics are only poorly understood to date. In 2013, Brendel et al. carried out a comprehensive evaluation of the severity of dysarthria and the profile of speech motor deficits in 20 patients with a genetic diagnosis of FRDA, based on a carefully selected battery of speech tasks and two commonly used Paraspeech studies, i.e., oral diadochokinesis and sustained vowel production [12]. Breathing, voice quality, voice instability, articulation, and tempo were identified as the most affected speech dimensions by perceptual ratings of the speech samples. The outcome indicated that FRDA pathology is differentially susceptible to speech production components and trunk/limb motor functions. Evidence has also emerged that part speech tasks do not permit an adequate scaling of FRDA speech deficits.

Functional neuroimaging studies and investigations have shown increased activation of the unaffected hemisphere in aphasia patients, which hypothetically reflects a maladaptive brain reorganisation strategy [72]. Seniow et al. investigated whether, when combined with speech/language therapy, repetitive magnetic transcription (rTMS) stimulation inhibiting the homologue in the right hemisphere in Broca improves the repair of the language. 40 aphasia patients were randomised to a 3-week aphasia rehabilitation protocol combined with real rTMS by using the Boston Diagnostic Aphasia baseline test. They reported that severe aphasic rTMS showed significantly more improvement than patients receiving repeated sham stimulation.

Petrillo et al. experimented in 2021 for the Italian version of the progressive aphasia severity scale (Italian PASS), which was built according to guidelines for cross-cultural

adaptation of self-report measures to aid researchers and clinicians in the diagnosis and follow-up of a primary progressive aphasia (PPA) in Italian populations [64]. This tool would allow researchers to gather data on patients with PPA's communicative functioning in everyday contexts, considering standardised tests employed in the clinical setting and the perspectives of their caregivers. Furthermore, it could be particularly beneficial for long-term disease monitoring to track its advancement, and it could be an ideal way to check the success of speech/language treatment in delaying disease progression.

Laganaro et al. released a screening version of a speech assessment protocol (MonPaGe-2.0. s) in 2021 as a response to the demand for objective screening tools for motor speech disorders. It is based on semi-automated acoustic and perceptual assessments of many speech characteristics in French (MSD) [39]. They tested the screening tool's sensitivity and specificity and compared the results to external standard evaluation methods. Data from 80 patients with mild to moderate MSD and 62 healthy test controls were compared to normative data from 404 neurotypical speakers, with Deviance Scores calculated on seven speech dimensions (articulation, prosody, pneumophonatory control, voice, speech rate, diadochokinetic rate, intelligibility) using acoustic and perceptual measures. The MonPaGe, TotDevS, and an external MSD composite perceptual score provided by six experts had a good connection. The sensitivity and specificity of the MonPaGe screening technique for diagnosing the existence and severity of MSD have been demonstrated. They concluded that to distinguish MSD subtypes, more implementations are needed to complement the definition of compromised dimensions.

Rolandic epilepsy is associated with developmental language impairment. Literature does not show exactly which domains are affected. In 2013, Overvliet et al. studied performance among children with Rolandic epilepsy and healthy controls in the language domains [57]. That is a focal study compared to healthy controls of children with Rolandic epilepsy. A CELF language test was carried out on 25 children with Rolandic epilepsy (mean 136.6 months, SD 23.0) and 25 years with healthy inspections matched with age (Clinical Evaluation of Language Fundamentals, Dutch edition). The core language score was significantly lower in children with epilepsy than healthy controls.

Speltz et al., in 2018, assessed whether infant cases with craniofacial microsomia (CFM) show lower neurodevelopmental status than demographically comparable infants without a craniofacial diagnosis ('controls') and examined the neurodevelopmental outcomes of cases by facial phenotype and hearing status [76]. Observational study on 108 cases and 84 controls aged 12–24 months was carried out. The third edition of Bayley scales for children and Toddlers and the fifth edition of the preschool linguist scales have been evaluated by participants (PLS-5). With the Craniofacial Microsomy Phenotypic Assessment Tool, facial features are categorised. Among women and those with higher socioeconomic status, outcomes were better. Facial phenotype and hearing status among cases showed little to no association with results. Although learning problems in older children with CFM have been observed, no evidence of developmental or language delay has been reported among infants.

5 Challenges, limitations and future research possibilities

With an increasing number of children with speech impairment, improving and devising methods for early detection is paramount to preventing disease progression. The development of this field may help adults and children receive better assessment and treatments from

clinical trials and hospitals. Therefore, several tool methods have been proposed to detect and predict this speech impairment; however, these techniques have fundamental limitations. This part discusses some of the challenges and future research directions to help more researchers address them.

One of the challenges against universal screening is that identifying and correctly diagnosing infants with speech impairment at 24 months of age, unless it is a cleft palate, is very difficult. There is still a pressing need to identify the appropriate mix of assessment tool modalities that would improve detection rates and reduce false-positive results. The development of such diagnostic tools can lead to a precise and conclusive diagnosis of speech impairment and the early detection of the condition. Two more challenges that need to be addressed include cost and dataset availability. Sustained efforts into developing a proper universal speech assessment tool will positively impact children's self-esteem and self-confidence with SSD [89]. The challenges faced during the study included a lack of databases that are dedicated to assessment tools for speech-impaired children. The absence of comprehensive datasets is a major setback to future development, as most publicly available datasets contain missing values for numerous detection algorithms. Data analysis is also complicated due to a lack of sufficient data. Techniques for early detection of speech problems in children are too costly for families and society to handle. In terms of screening children at an early age, progress is being made in improving screening techniques that can be cost-friendly, eco-friendly, and reliably identify at-risk status. Given the large amount of positive results, more effort is needed to duplicate, expand, and individualise available therapies and screening and diagnostic tools.

Additionally, the available literature is contained in databases that require either subscription or specific institutional credentials to have access. This phenomenon is quite frustrating since scientists should have unlimited access to the available data to conduct their studies seamlessly [11]. The researcher must perform numerous searches in various databases to capture all the relevant peer-reviewed studies for inclusion in the systematic review. Moreover, several papers were contained in multiple databases, which drastically reduced the number of eligible articles for inclusion in the systematic review.

Furthermore, Due to the limitation of manual or hand transcription-based diagnostic evaluation approaches, there is a growing demand for automated methods to quantify child speech patterns and aid in the rapid and reliable diagnosis of speech impairment [80]. Automatic assessment models are promising tools for detecting speech impairment. Artificial intelligence approaches, such as deep learning, effectively model exceedingly complex data accurately. These models are more resilient and interpretable than other similar techniques, yet they are computational models that try to find the relationship between a collection of datasets and their results. These models rely on many hyperparameters, all of which must be fine-tuned. Datasets are also crucial to the effectiveness of deep learning models; they must be impartial to achieve the best outcomes. Features in the datasets must also be thoroughly studied and unrelated. Another significant problem is predicting speech impairment in newborns and infants between 0 and 24 months.

6 Conclusion

The number of children with SSD is expected to rise in the future, along with the cost of treatment and intervention. Various speech assessment tools have been developed to diagnose

and treat SSD, such as “The Caterpillar” and “My Grandfather” automatic tools, DEMSS, and MSE. However, their success is limited due to varied cultural practices and orientations, and lack of universality due to limited validity and reliability. Detecting SSD accurately at the child’s preschool years ensures that the condition is eliminated and does not persist into adolescence. Future studies will have to incorporate studies dedicated to testing speech-impaired children’s speech assessment tools’ validity, reliability, and universality. It is essential to ensure that researchers develop a universally accepted speech assessment tool that transcends all cultural barriers to help speech-language pathologists. For example, future studies should include more research on developing a speech assessment tool ideal for multilingual and bilingual children. Furthermore, studies should consist of more than 150 peer-reviewed papers to improve reliability and validity. In total, there still exists a need to develop speech assessment tools independent of human judgment to help diagnose and intervene to aid in the early detection and intervention of SSD in children.

Data availability Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Declarations

Competing interests The authors have no competing interests to declare relevant to this article’s content.

References

1. (PDF) A systematic review of the interventions used with preschool children with primary speech and language impairment. https://www.researchgate.net/publication/315077019_A_systematic_review_of_the_interventions_used_with_preschool_children_with_primary_speech_and_language_impairment. Accessed 19 Jun 2022
2. Abaskohi A, Mortazavi F, Moradi H (2022) Automatic speech recognition for speech assessment of Persian preschool children. arXiv
3. Abdou D, Afsah O, Baz H, Abou-Elsaad T (2020) Developing a test battery for diagnosis of childhood apraxia of speech in arabic speakers. *Egypt J Otolaryngol* 36:20. <https://doi.org/10.1186/s43163-020-00021-5>
4. Afshan A, Lulich S, Alwan A et al (2017) Predicting clinical evaluations of children’s speech with limited data using exemplar word template references. <https://doi.org/10.21437/SLaTE.2017-28>
5. Al-Qatab BA, Mustafa MB (2021) Classification of dysarthric speech according to the severity of impairment: an analysis of acoustic features. *IEEE Access* 9:18183–18194. <https://doi.org/10.1109/ACCESS.2021.3053335>
6. Bachmann AS, Wiltfang J, Hertrampf K (2021) Development of the german speech intelligibility index for the treatment of oral cancer patients. *J Cranio-Maxillofacial Surg* 49:52–58. <https://doi.org/10.1016/J.JCMS.2020.11.009>
7. Backer V (2014) Literature review of childhood apraxia of speech: approach on controversy of labeling, diagnosing, and intervention. *Research papers*
8. Barrett C, McCabe P, Masso S, Preston J (2020) Protocol for the connected speech transcription of children with speech disorders: an example from childhood apraxia of speech. *Folia Phoniatr Logop* 72:152–166. <https://doi.org/10.1159/000500664>
9. Berry J, Griffiths M, Westcott C (2012) A double-blind, randomized, controlled trial of tongue-tie division and its immediate effect on breastfeeding. *Breastfeed Med* 7:189–193. <https://doi.org/10.1089/bfm.2011.0030>
10. Bessell A, Sell D, Whiting P et al (2013) Speech and language therapy interventions for children with cleft palate: a systematic review. *Cleft Palate Craniofac J* 50:1–17
11. Booth E, Carns J, Kennington C, Rafla N (2020) Evaluating and improving child-directed automatic speech recognition, pp 11–16

12. Brendel B, Ackermann H, Berg D et al (2013) Friedreich ataxia: dysarthria profile and clinical data. *Cerebellum* 12:475–484. <https://doi.org/10.1007/s12311-012-0440-0>
13. Britton L, Albery L, Bowden M et al (2014) A cross-sectional cohort study of speech in five-year-olds with cleft palate ± lip to support development of National Audit Standards: Benchmarking Speech Standards in the United Kingdom. *Cleft Palate Craniofac J* 51:431–451. <https://doi.org/10.1597/13-121>
14. Brookes A, Bowley DM (2014) Tongue tie: the evidence for frenotomy. *Early Hum Dev* 90:765–768
15. Broome K, McCabe P, Docking K, Doble M (2017) A systematic review of speech assessments for children with autism spectrum disorder: recommendations for best practice. *Am J Speech Lang Pathol* 26:1011–1029. https://doi.org/10.1044/2017_AJSLP-16-0014
16. Brumbaugh KM, Smit AB (2013) Treating children ages 3–6 who have speech sound disorder: a survey. *Lang Speech Hear Serv Sch* 44:306–319. [https://doi.org/10.1044/0161-1461\(2013\)12-0029](https://doi.org/10.1044/0161-1461(2013)12-0029)
17. Carter JA, Murira G, Gona J et al (2012) Speech and language disorders in Kenyan children: adapting tools for regions with few assessment resources. *J Psychol Afr* 22:155–169. <https://doi.org/10.1080/14330237.2012.10820514>
18. Chaware S, Dubey S, Kakatkar V et al (2021) The systematic review and meta-analysis of oral sensory challenges in children and adolescents with autism spectrum disorder. *J Int Soc Prev Community Dent* 11: 469. https://doi.org/10.4103/JISPCD.JISPCD_135_21
19. Chong WW, Abd Rahman FN, Harun NA (2022) Screen time of children with speech delay: a cross-sectional study in a tertiary center in Kuantan, Malaysia. *Pediatr Int* 64:e15105. <https://doi.org/10.1111/ped.15105>
20. Cummins N, Scherer S, Krajewski J et al (2015) A review of depression and suicide risk assessment using speech analysis. *Speech Commun* 71:10–49. <https://doi.org/10.1016/j.specom.2015.03.004>
21. Cunningham BJ, Washington KN, Binns A et al (2017) Current methods of evaluating speech-language outcomes for preschoolers with communication disorders: a scoping review using the ICF-CY. *J Speech Lang Hear Res* 60:447–464. https://doi.org/10.1044/2016_JSLHR-L-15-0329
22. Demenko G, Wagner A, Cylwik N (2010) The use of speech technology in foreign language pronunciation training. *Arch Acoust* 35:309–329. <https://doi.org/10.2478/v10168-010-0027-z>
23. Dodd B (2014) Differential diagnosis of pediatric speech sound disorder. *Curr Dev Disord Rep* 1:189–196. <https://doi.org/10.1007/s40474-014-0017-3>
24. Eadie P, Morgan A, Ukoumunne OC et al (2015) Speech sound disorder at 4 years: prevalence, comorbidities, and predictors in a community cohort of children. *Dev Med Child Neurol* 57:578–584. <https://doi.org/10.1111/dmcn.12635>
25. Ertmer DJ (2010) Relationships between speech intelligibility and word articulation scores in children with hearing loss. *J Speech Lang Hear Res*. [https://doi.org/10.1044/1092-4388\(2010\)09-0250](https://doi.org/10.1044/1092-4388(2010)09-0250)
26. Esch BE, LaLonde KB, Esch JW (2010) Speech and language assessment: a verbal behavior analysis. *J Speech Lang Pathol Appl Behav Anal* 5:166–191. <https://doi.org/10.1037/h0100270>
27. Fulcher A, Purcell AA, Baker E, Munro N (2012) Listen up: children with early identified hearing loss achieve age-appropriate speech/language outcomes by 3 years-of-age. *Int J Pediatr Otorhinolaryngol* 76: 1785–1794. <https://doi.org/10.1016/j.ijporl.2012.09.001>
28. Guiberson M, Atkins J (2012) Speech-language pathologists' preparation, practices, and perspectives on serving culturally and linguistically diverse children. *Commun Disord Q* 33:169–180. <https://doi.org/10.1177/1525740110384132>
29. Hasson N, Camilleri B, Jones C et al (2013) Discriminating disorder from difference using dynamic assessment with bilingual children. *Child Lang Teach Ther* 29:57–75. <https://doi.org/10.1177/026659012459526>
30. Hochmuth S, Brand T, Zokoll MA et al (2012) A Spanish matrix sentence test for assessing speech reception thresholds in noise. *Int J Audiol* 51:536–544. <https://doi.org/10.3109/14992027.2012.670731>
31. How much does speech therapy cost? <https://connectedspeechpathology.com/blog/how-much-does-speech-therapy-cost>. Accessed 20 Dec 2022
32. Hustad KC, Gorton K, Lee J (2010) Classification of speech and language profiles in 4-year-old children with cerebral palsy: a prospective preliminary study. *J Speech Lang Hear Res* 53:1496–1513. [https://doi.org/10.1044/1092-4388\(2010\)09-0176](https://doi.org/10.1044/1092-4388(2010)09-0176)
33. Ito Y, Shimizu T, Nakamura T, Takatama C (2015) Effectiveness of tongue-tie division for speech disorder in children. *Pediatr Int* 57:222–226. <https://doi.org/10.1111/ped.12474>
34. Jesus LMT, Martinez J, Santos J et al (2019) Comparing traditional and tablet-based intervention for children with speech sound disorders: a randomized controlled trial. *J Speech Lang Hear Res* 62:4045–4061. https://doi.org/10.1044/2019_JSLHR-S-18-0301
35. Kayikci MEK, Akan S, Ciger S, Özkan S (2012) Effects of Hawley retainers on consonants and formant frequencies of vowels. *Angle Orthod* 82:14–21. <https://doi.org/10.2319/032911-226.1>

36. Kent RD, Vorperian HK (2013) Speech impairment in down syndrome: a review. *J Speech Lang Hear Res* 56:178–210. [https://doi.org/10.1044/1092-4388\(2012/12-0148\)](https://doi.org/10.1044/1092-4388(2012/12-0148))
37. Khatib TZ, Farah H, Al-Sabbagh R et al (2013) Speech performance and oral impairments with lingual and labial orthodontic appliances in the first stage of fixed treatment: a randomized controlled trial. *Angle Orthod* 83:519–526. <https://doi.org/10.2319/073112-619.1>
38. Krishnan S, Alcock KJ, Mercure E et al (2013) Articulating novel words: children's oromotor skills predict nonword repetition abilities. *J Speech Lang Hear Res* 56:1800–1812. [https://doi.org/10.1044/1092-4388\(2013/12-0206\)](https://doi.org/10.1044/1092-4388(2013/12-0206))
39. Laganaro M, Fougeron C, Pernon M et al (2021) Sensitivity and specificity of an acoustic- and perceptual-based tool for assessing motor speech disorders in French: the MonPaGe-screening protocol. *Clin Linguist Phon* 35:1060–1075. https://doi.org/10.1080/02699206.2020.1865460/SUPPL_FILE/ICLP_A_1865460_SM1210.DOCX
40. Lehner K, Pfäb J, Ziegler W (2022) Web-based assessment of communication-related parameters in dysarthria: development and implementation of the KommPaS web app. *Clin Linguist Phon* 36(12): 1093–1111. <https://doi.org/10.1080/02699206.2021.1989490>
41. Liss JM, LeGendre S, Lotto AJ (2010) Discriminating dysarthria type from envelope modulation spectra research note. *J Speech Lang Hear Res* 53:1246–1255. [https://doi.org/10.1044/1092-4388\(2010/09-0121\)](https://doi.org/10.1044/1092-4388(2010/09-0121))
42. Low DM, Bentley KH, Ghosh SS (2020) Automated assessment of psychiatric disorders using speech: a systematic review. *Laryngoscope Investig Otolaryngol* 5:96. <https://doi.org/10.1002/LIO2.354>
43. McFaul H, Mulgrew L, Smyth J, Titterton J (2022) Applying evidence to practice by increasing intensity of intervention for children with severe speech sound disorder: a quality improvement project. *BMJ Open Qual* 11:e001761. <https://doi.org/10.1136/bmj-2021-001761>
44. Mckechnie J, Ahmed B, Gutierrez-Osuna R et al (2018) International journal of speech-language pathology automated speech analysis tools for children's speech production: a systematic literature review automated speech analysis tools for children's speech production: a systematic literature review. *Int J Speech Lang Pathol* 20:583–598. <https://doi.org/10.1080/17549507.2018.1477991>
45. McLeod S, Harrison LJ, McAllister L, McCormack J (2013) Speech sound disorders in a community study of preschool children. *Am J Speech Lang Pathol* 22:503–522. [https://doi.org/10.1044/1058-0360\(2012/11-0123\)](https://doi.org/10.1044/1058-0360(2012/11-0123))
46. McLeod S, Verdon S, Bowen C et al (2013) International aspirations for speech-language pathologists' practice with multilingual children with speech sound disorders: development of a position paper. *J Commun Disord* 46:375–387. <https://doi.org/10.1016/j.jcomdis.2013.04.003>
47. McLeod S, Verdon S, Baker E et al (2017) Tutorial: speech assessment for multilingual children who do not speak the same language(s) as the speech-language pathologist. *Am J Speech Lang Pathol* 26:691–708. https://doi.org/10.1044/2017_AJSLP-15-0161
48. Mehta DD, Van Stan JH, Zañartu M, Ghassemi M, Guttag JV, Espinoza VM, Cortés JP, Cheyne HA 2nd, Hillman RE (2015) Using ambulatory voice monitoring to investigate common voice disorders: research update. *Front Bioeng Biotechnol* 16(3):155. <https://doi.org/10.3389/fbioe.2015.00155>
49. Mogren Å, Sjögreen L, Barr Agholme M, McAllister A (2020) Orofacial function in children with speech sound disorders persisting after the age of six years. *Int J Speech Lang Pathol* 22:526–536. <https://doi.org/10.1080/17549507.2019.1701081>
50. Morgan AT, Su M, Reilly S et al (2018) A brain marker for Developmental Speech Disorders. *J Pediatr* 198: 234–239.e1. <https://doi.org/10.1016/j.jpeds.2018.02.043>
51. Mugada V, Kiran Kolakota R, Karri SR et al (2018) Evaluation of quality of life of head and neck cancer patients: a descriptive cross-sectional study. *Int J Res Rev (www.ijrjournal.com)* 5:241
52. Mullen R, Schooling T (2010) The national outcomes measurement system for pediatric speech-language pathology. *Lang Speech Hear Serv Sch* 41:44–60. [https://doi.org/10.1044/0161-1461\(2009/08-0051\)](https://doi.org/10.1044/0161-1461(2009/08-0051))
53. Murray E, McCabe P, Heard R, Ballard KJ (2015) Differential diagnosis of children with suspected childhood apraxia of speech. *J Speech Lang Hear Res* 58:43–60. https://doi.org/10.1044/2014_JSLHR-S-12-0358
54. Nelson TL, Mok Z, Ttofari Eecen K (2020) Use of transcription when assessing children's speech: Australian speech-language pathologists' practices, challenges, and facilitators. *Folia Phoniatr Logop* 72: 131–142. <https://doi.org/10.1159/000503131>
55. Newbold EJ, Stackhouse J, Wells B (2013) Tracking change in children with severe and persisting speech difficulties. *Clin Linguist Phon* 27:521–539. <https://doi.org/10.3109/02699206.2013.790479>
56. Ng S-I, Ng CW-Y, Wang J et al (2020) CUCHILD: a large-scale cantonese corpus of child speech for phonology and articulation assessment. *Interspeech 2020*. ISCA, ISCA, pp 424–428
57. Overvliet GM, van der Besseling RMH et al (2013) Clinical evaluation of language fundamentals in Rolandic epilepsy, an assessment with CELF-4. *Eur J Pediatr Neurol* 17:390–396. <https://doi.org/10.1016/j.ejpn.2013.01.001>

58. Pamplona M del C, Ysunza PA (2020) Speech pathology telepractice for children with cleft palate in the times of COVID-19 pandemic. *Int J Pediatr Otorhinolaryngol* 138:110318. <https://doi.org/10.1016/j.ijporl.2020.110318>
59. Parnandi A, Karappa V, Son Y et al (2013) Architecture of an automated therapy tool for childhood apraxia of speech. In: Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2013. Association for Computing Machinery
60. Pascoe M, Maphalala Z, Ebrahim A et al (2010) Children with speech difficulties: an exploratory survey of clinical practice in the Western Cape. *S Afr J Commun Disord* 57:66–75. <https://doi.org/10.4102/sajcd.v57i1.51>
61. Patel R, Connaghan K, Franco D et al (2013) “The caterpillar”: a novel reading passage for assessment of motor speech disorders. *Am J Speech Lang Pathol* 22:1–9. [https://doi.org/10.1044/1058-0360\(2012/11-0134\)](https://doi.org/10.1044/1058-0360(2012/11-0134))
62. Pejovic J, Cruz M, Severino C, Frota S (2021) Early visual attention abilities and audiovisual speech processing in 5–7 month-old down syndrome and typically developing infants. *Brain Sci* 2021 11:939. <https://doi.org/10.3390/BRAINSCI11070939>
63. Pennington L, Virella D, MjØen T et al (2013) Development of the viking Speech Scale to classify the speech of children with cerebral palsy. *Res Dev Disabil* 34:3202–3210. <https://doi.org/10.1016/j.ridd.2013.06.035>
64. Petrillo SM, Del Mauro M, Lambro BE, Occhigrossi C, Piccirilli M (2022) Italian translation and cross-cultural adaptation of the Progressive Aphasia Severity Scale. *Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology* 43(5):3065–3070. <https://doi.org/10.1007/s10072-021-05651-y>
65. Phillips J, Wiley S, Barnard H, Meitzen-Derr J (2014) Comparison of two nonverbal intelligence tests among children who are deaf or hard-of-hearing. *Res Dev Disabil* 35:463–471. <https://doi.org/10.1016/j.ridd.2013.11.020>
66. Preston JL, Brick N, Landi N (2013) Ultrasound biofeedback treatment for persisting childhood apraxia of speech. *Am J Speech Lang Pathol* 22:627–643. [https://doi.org/10.1044/1058-0360\(2013/12-0139\)](https://doi.org/10.1044/1058-0360(2013/12-0139))
67. Ravi SK, Sumanth P, Saraswathi T et al (2021) Prevalence of communication disorders among school children in Ballari, South India: a cross-sectional study. *Clin Epidemiol Glob Health* 12:100851. <https://doi.org/10.1016/J.CEGH.2021.100851>
68. Rosenbaum S, Simon P (2016) Committee on the evaluation of the supplemental security income (SSI) disability program for children with speech disorders and language disorders, board on the health of select populations, board on children, youth, and families, institute of medicine, division of behavioral and social sciences and education, & national academies of sciences, engineering, and medicine (eds). *Speech and language disorders in children: implications for the social security administration's supplemental security income program*. National Academies Press (US). <https://doi.org/10.17226/21872>
69. Shradha Salunkhe et al (2021) Assessment of speech and language delay in children upto 6 years using Trivandrumdevelopment screening chart and language evaluation scale Trivandrum. *MedPulse International Journal of Pediatrics* 17(1):06–09. <https://doi.org/10.26611/10141712>
70. Schölderle T, Haas E, Ziegler W (2020) Age norms for auditory-perceptual nonphonetic parameters: a prerequisite for the assessment of childhood dysarthria. *J Speech Lang Hear Res* 63:1071–1082. https://doi.org/10.1044/2020_JSLHR-19-00114
71. Schuster M, Stelzle F (2012) Outcome measurements after oral cancer treatment: speech and speech-related aspects—an overview. *Oral Maxillofac Surg* 16:291–298. <https://doi.org/10.1007/s10006-012-0340-y>
72. Seniów J, Waldowski K, Leśniak M et al (2013) Transcranial magnetic stimulation combined with speech and language training in early aphasia rehabilitation: a randomized double-blind controlled pilot study. *Top Stroke Rehabil* 20:250–261. <https://doi.org/10.1310/tsr2003-250>
73. Shahin M, Ahmed B, Parnandi A et al (2015) Tabby Talks: an automated tool for the assessment of childhood apraxia of speech. *Speech Commun* 70:49–64. <https://doi.org/10.1016/j.specom.2015.04.002>
74. Sharma D, Singh G (2016) Pediatric head and neck squamous cell carcinoma: a retrospective observational study. *Indian J Cancer* 53:397–398. <https://doi.org/10.4103/0019-509X.200676>
75. Shriberg LD, Fourakis M, Hall SD et al (2010) Extensions to the Speech Disorders Classification System (SDCS). *Clin Linguist Phon* 24:795–824. <https://doi.org/10.3109/02699206.2010.503006>
76. Speltz ML, Kapp-Simon KA, Johns AL et al (2018) Neurodevelopment of infants with and without craniofacial microsomia. *J Pediatr* 198:226–233e3. <https://doi.org/10.1016/j.jpeds.2018.02.076>
77. Stasak B, Epps J, Goecke R (2017) Elicitation design for acoustic depression classification: an investigation of articulation effort, linguistic complexity, and word affect. In: *Interspeech 2017*. ISCA, ISCA, pp 834–838
78. Stelzle F, Ugrinovic B, Knipfer C et al (2010) Automatic, computer-based speech assessment on edentulous patients with and without complete dentures - preliminary results. *J Oral Rehabil* 37:209–216. <https://doi.org/10.1111/j.1365-2842.2009.02047.x>

79. Strand EA, McCauley RJ, Weigand SD et al (2013) A motor speech assessment for children with severe speech disorders: reliability and validity evidence. *J Speech Lang Hear Res* 56:505–520. [https://doi.org/10.1044/1092-4388\(2012/12-0094\)](https://doi.org/10.1044/1092-4388(2012/12-0094))
80. Suthar K, Yousefi Zowj F, Speights Atkins M, He QP (2022) Feature engineering and machine learning for computer-assisted screening of children with speech disorders. *PLOS Digital Health* 1(5):e0000041. <https://doi.org/10.1371/journal.pdig.0000041>
81. Taylor OD, Armfield NR, Dodrill P, Smith AC (2014) A review of the efficacy and effectiveness of using telehealth for paediatric speech and language assessment. *J Telemed Telecare* 20:405–412. <https://doi.org/10.1177/1357633X14552388>
82. Terband H, Namasivayam A, Maas E et al (2019) Assessment of childhood apraxia of speech: a review/tutorial of objective measurement techniques. *J Speech Lang Hear Res* 62:2999–3032. https://doi.org/10.1044/2019_JSLHR-S-CSMC7-19-0214
83. Towey D, Jin L, Zhu H et al (2020) CHOCSLAT: Chinese Healthcare-Oriented Computerised Speech & Language Assessment Tools. In: Proceedings – 2020 IEEE 44th Annual Computers, Software, and Applications Conference, COMPSAC 2020. Institute of Electrical and Electronics Engineers Inc., pp 1460–1465
84. van der Jacobi I, Huiskens H et al (2010) Voice and speech outcomes of chemoradiation for advanced head and neck cancer: a systematic review. *Eur Arch Otorhinolaryngol* 267:1495–1505
85. Veenhuis SJG, van OS NJH, van Gerven MHJC et al (2021) Dysarthria in children and adults with ataxia telangiectasia. *Dev Med Child Neurol* 63:450–456. <https://doi.org/10.1111/DMCN.14811>
86. Walenski M, Love T (2018) The real-time comprehension of idioms by typical children, children with specific language impairment and children with autism. *J Speech Pathol Ther* 3:130. <https://doi.org/10.4172/2472-5005.1000130>
87. Wallace IF, Berkman ND, Watson LR et al (2015) Screening for speech and language delay in children 5 years old and younger: a systematic review. *Pediatrics* 136:e448–e462
88. Wang C-H, Lin, Hsieh et al (2013) Sensory integration dysfunction affects efficacy of speech therapy on children with functional articulation disorders. *Neuropsychiatr Dis Treat* 9:87. <https://doi.org/10.2147/NDT.S40499>
89. Wren Y, Hambly H, Roulstone S (2013) A review of the impact of bilingualism on the development of phonemic awareness skills in children with typical speech development. *Child Lang Teach Ther* 29:11–25
90. Wren Y, Harding S, Goldbart J, Roulstone S (2018) A systematic review and classification of interventions for speech-sound disorder in preschool children. *Int J Lang Commun Disord* 53:446–467. <https://doi.org/10.1111/1460-6984.12371>
91. Zarifian T, Modarresi Y, Gholami Tehrani L, Dastjerdi Kazemi M, Salavati M et al (2017) Persian articulation assessment for children aged 3–6 years: a validation study. *Iran J Pediatr* 27(4):e8217. <https://doi.org/10.5812/ijp.8217>
92. Zharkova N (2013) Using ultrasound to quantify tongue shape and movement characteristics. *Cleft Palate Craniofac J* 50:76–81. <https://doi.org/10.1597/11-196>

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