AJSLP

Research Article

# Reliability of Perceptual Judgments of Phonetic Accuracy and Hypernasality Among Speech-Language Pathologists for Children With Dysarthria

Kristen M. Allison,[a] Mackenzie Russell,[a] and Katherine C. Hustad[b,c]

**Purpose:** The objectives of this study were to: (a) compare interrater reliability of practicing speech-language pathologists' (SLPs) perceptual judgments of phonetic accuracy and hypernasality between children with dysarthria and those with typical development, and (b) to identify speech factors that influence reliability of these perceptual judgments for children with dysarthria.
**Method:** Ten SLPs provided ratings of speech samples from twenty 5-year-old children with dysarthria and twenty 5-year-old children with typical development on two tasks via a web-based platform: a hypernasality judgment task and a phonetic accuracy judgment task. Interrater reliability of SLPs' ratings on both tasks was compared between children with dysarthria and children with typical development. For children with dysarthria, four acoustic speech measures, intelligibility, and a measure of phonetic accuracy (percent stops correct) were examined as predictors of reliability of SLPs' perceptual judgments.
**Results:** Reliability of SLPs' phonetic accuracy judgments and hypernasality ratings was significantly lower for children with dysarthria than for children with typical development. Among children with dysarthria, interrater reliability of perceptual judgments ranged from strong to weak. Percent stops correct was the strongest predictor of interrater reliability for both phonetic accuracy judgments and hypernasality ratings.
**Conclusions:** Reliability of perceptual phonetic accuracy judgments and hypernasality ratings among practicing SLPs for children with dysarthria is reduced compared to ratings for children with typical development. Findings underscore the need for more reliable methods to assess phonetic accuracy and hypernasality for children with dysarthria.

Children with dysarthria exhibit a variety of atypical speech features that can be associated with deficits in one or more of the speech subsystems of articulation, resonance, phonation, and respiration. Articulation is particularly problematic, with studies showing clear evidence that children with dysarthria secondary to cerebral palsy (CP) have reduced speech sound accuracy relative to age expectations (Nordberg et al., 2014; Workinger & Kent, 1991). Studies suggest that segmental errors continue to be present in adults with developmental dysarthria secondary to CP (H. Kim et al., 2010; Platt, Andrews, & Howie, 1980; Platt, Andrews, Young, & Quinn, 1980). Although the ways that speech sounds develop and change over time in children with dysarthria have not been systematically studied, it stands to reason that production capabilities are malleable to some extent for children by virtue of their age and the fact that children with dysarthria are still developing at least through 9 years of age (Hustad et al., 2020, 2019; Schölderle et al., 2020). Children with dysarthria also present with speech characteristics that globally distort the speech signal (e.g., hypernasality, voice quality disturbance, slow rate) due to impairments in other speech subsystems (Allison & Hustad, 2018a; Lee et al., 2014; Workinger & Kent, 1991). Deviations in these global speech features lead to reduced distinctiveness of speech sounds and word boundaries, leading to reduced intelligibility (Kent et al., 1989; Liss et al., 1998). Thus, assessment of both speech

[a]Department of Communication Sciences and Disorders, Northeastern University, Boston, MA
[b]Department of Communication Sciences and Disorders, University of Wisconsin–Madison
[c]Waisman Center, University of Wisconsin–Madison

Correspondence to Kristen M. Allison: k.allison@northeastern.edu

Editor-in-Chief: Jessica E. Huber
Editor: Theresa Schölderle

Received May 25, 2020
Revision received August 17, 2020
Accepted October 20, 2020
https://doi.org/10.1044/2020_AJSLP-20-00144

**Publisher Note:** This article is part of the Special Issue: Selected Papers From the 2020 Conference on Motor Speech—Clinical Science and Implications.

*Disclosure: The authors have declared that no competing interests existed at the time of publication.*

sound accuracy and global speech characteristics is of the utmost importance for identifying potential intervention targets to maximize intelligibility in children with dysarthria.

Clinical assessment of speech sound accuracy relies on perceptual judgment by clinicians and is important for establishing treatment goals; thus, understanding its reliability for children with dysarthria is important. Global speech features associated with dysarthria are also typically judged perceptually (i.e., clinicians make judgments about the presence and severity of dysarthria features, such as hypernasality, based on their overall impression after listening to structured speech samples; Duffy, 2005). Acoustic measures that can be obtained from speech recordings exist for quantifying many of these features (e.g., voice quality, respiration, prosody; Allison & Hustad, 2018a; R. D. Kent & Kim, 2003) and can be used to improve their objective measurement, although they are primarily used in research settings. Hypernasality is one of the most common features of dysarthria in children with CP (Workinger & Kent, 1991) but is particularly challenging to quantify acoustically (R. D. Kent & Kim, 2003). Although hypernasality is not thought to make as large a contribution to intelligibility impairment as deficits in articulatory control for children with dysarthria (Lee et al., 2014), its negative impact on intelligibility and listener attitudes is well documented (Bettens et al., 2020; Maegawa et al., 1998; Yorkston & Beukelman, 1981). Hypernasality can negatively impact listeners' ability to distinguish between oral and nasal consonants produced in the same place of articulation, thus reducing intelligibility (Kent et al., 1989). Speakers with significant hypernasality may also produce oral consonants in different places of articulation to compensate for difficulty achieving intraoral pressure, further impacting intelligibility (Kummer, 2011). Because assessment of resonance is an important component of a comprehensive motor speech exam, reliability of perceptual hypernasality ratings by clinicians is particularly important.

Currently, phonetic transcription and perceptual judgment by speech-language pathologists (SLPs), typically within the context of free speech samples or standardized articulation tests, are the clinical standard for evaluating speech characteristics; however, there are reliability challenges with these judgments (McHenry, 1999; Shriberg & Lof, 1991). Phonetic transcription for children with dysarthria presents added challenges to reliability because the distortions caused by articulatory imprecision and other speech subsystem deficits result in atypical errors that are not easily characterized by means of broad or narrow phonetic transcription. Although studies indicate that narrow phonetic transcription has poor reliability, research indicates that expert clinicians trained with the same protocol and working with one another in a research setting are able to attain adequate broad transcription reliability for children with speech sound disorders (SSDs; R. D. Kent, 1996; McSweeny et al., 1995; Shriberg & Lof, 1991).

Similarly, several standardized articulation tests that are widely used by practicing clinicians (e.g., Goldman-Fristoe Test of Articulation–Third Edition; Goldman & Fristoe, 2015) require the SLP to make binary or categorical judgments of "correct" or "incorrect" for individual target speech sounds produced in single-word utterances. These tests are commonly used to assess speech sound accuracy for children with all types of SSDs, including dysarthria. A limitation of these tests and their norms is that they do not examine accuracy of speech sounds in connected speech but, rather, are based on single-word productions featuring target phonemes. Single-word productions are a simpler task, motorically, and thus may reflect optimized productions for children with dysarthria. One solution involves the use of speech elicitation tasks that employ connected speech where the target utterances and constituent words and phonemes are known, and listeners are asked to make binary or forced-choice ratings of the accuracy of individual phonemes within target utterances. Such an approach holds promise for characterizing speech sound production in utterances using a constrained format similar to measures made from standardized articulation tests. A key question is whether such measures suffer from the same reliability problems that plague phonetic transcription approaches. Thus, a primary objective of this study was to quantify reliability of SLPs on forced-choice judgments of phonetic accuracy of speech sounds produced in multiword utterances by children with dysarthria secondary to CP. Specifically, we sought to examine whether there were reliability differences among SLP ratings for children with dysarthria versus SLP ratings for typically developing (TD) children on the same speech tasks. TD children acquire most consonants by the age of 5 years but may exhibit developmental speech sound errors until the age of 7 years (McLeod & Crowe, 2018) and are continuing to develop speech subsystem control until at least the age of 9 years (Schölderle et al., 2020). Comparing reliability of phonetic accuracy judgments for children with dysarthria to that for children with typical development is important for understanding the degree to which deviations in the speech signal associated with dysarthria impact reliability of phonetic accuracy judgments. We examined a limited set of phonemes for this task (i.e., stop consonants) in order to simplify the task for SLPs and because children with CP are known to have reduced acoustic specification of stop consonant production (Allison & Hustad, 2018a; Liu et al., 2000).

Speech production deficits in dysarthria extend beyond the realm of articulation. Hypernasality is a common feature of dysarthria in children with CP (Workinger & Kent, 1991) due to weakness or reduced control of the velum. It can be intermittent or constant, resulting in minor to major distortions of the speech signal. Although resonance disturbances in children can include both hyponasality and hypernasality, only the latter is a common symptom of dysarthria in children with CP. Despite its importance in comprehensively describing a child's dysarthria characteristics, hypernasality is a difficult feature to rate, particularly in speakers with dysarthria who have simultaneous involvement of multiple speech systems. Nasometry is the gold standard tool for objectively characterizing hypernasality. Measurement requires considerable instrumentation and involves simultaneous quantification of both oral intensity and nasal intensity over time, yielding nasalance percentage (Watterson, 2020). For children with velopharyngeal incompetence due to cleft palate, nasalance has been shown to be highly correlated with perceptual judgments of hypernasality (Brunnegård et al., 2012; Sweeney & Sell, 2008), and perceptual judgments of hypernasality have been shown to be reliable in trained listeners (Sweeney & Sell, 2008). Correspondence between hypernasality ratings and nasalance scores in children with dysarthria has not been reported in the literature, to our knowledge; however, studies of adults with dysarthria have found variable reliability in hypernasality judgments. Agreement between expert listeners in perceptual ratings of hypernasality and other dysarthria features on a 7-point scale was shown to be highest for speakers with the mildest and most severe mean ratings on each feature, with the lowest agreement for individuals in the middle of the severity spectrum (Bunton et al., 2007). Hypernasality ratings in adults with dysarthria have also been shown to be influenced by other disrupted speech features, such as reduced rate (McHenry, 1999). In this study, we sought to examine perceptual ratings of hypernasality by SLPs for children with dysarthria and TD peers. Although hyponasality can occur in TD children (Schölderle et al., 2020), it is not a common characteristic of dysarthria associated with CP. Our prior work has shown that 5-year-old children with dysarthria were rated by SLPs as having increased hypernasality compared to TD peers (Allison & Hustad, 2018b). Our key focus for this study was whether there were differences in the reliability of SLP ratings of hypernasality between the two groups: children with dysarthria and TD peers.

In earlier work, we examined acoustic features of speech associated with individual speech subsystems within a research framework. We found that there were distinctive acoustic features of speech that were unique to children with dysarthria and differentiated them from those in TD peers with a high degree of accuracy. Articulation rate and range of the second formant frequency (F2) for diphthongs were especially important differentiators, along with other segmental and suprasegmental features of speech (Allison & Hustad, 2018a). In addition, we found that different speech profiles existed among children with dysarthria secondary to CP and that these differences were not tied to severity (Allison & Hustad, 2018b).

In this study, we sought to extend our findings from the earlier acoustic studies to determine which acoustic features best predict reliability of perceptual ratings by SLPs for children with dysarthria secondary to CP. Specifically, we aimed to examine whether acoustic measures of articulatory precision (i.e., proportion of bursts produced, F2 range of diphthongs), voice quality (i.e., proportion of deviant voice quality), and articulation rate were related to reliability of SLPs' perceptual judgments as well as intelligibility, which is a more global measure of severity of speech motor impairment. These measures were chosen because of their known association with dysarthria in children with CP and in order to capture aspects of children's speech production spanning across speech subsystems (Allison & Hustad, 2018a).

This study examined the interrater reliability of clinicians in their perceptual judgments of phonetic accuracy and in their ratings of hypernasality for children with dysarthria secondary to CP and for TD peers. Specific questions were as follows:

1. For ratings of phonetic accuracy and hypernasality, is there a difference in SLPs' interrater reliability for children with dysarthria versus that for children with typical development?

2. For children with dysarthria, what speech factors best predict reliability of SLPs' phonetic accuracy and hypernasality judgments?

We hypothesized that reliability of SLP judgments for phonetic accuracy and for hypernasality would be higher for TD children than for children with dysarthria, due to the challenges of making perceptual judgments of speech signals that are distorted in multiple ways. In addition, we hypothesized that reliability of SLP judgments for phonetic accuracy and for hypernasality would be reduced for children with dysarthria who have more severe distortions in their speech signals. Specifically, we expected that acoustic measures of articulatory precision (i.e., F2 range of diphthongs, proportion of bursts produced) would be the strongest predictors of SLPs' reliability for both phonetic accuracy and hypernasality judgments because these measures have been shown to affect perception of phonetic contrasts, including stop–nasal contrasts that SLPs likely use to judge hypernasality (Ansel & Kent, 1992; Kent et al., 1989).

# Method

## *Participants*

### SLPs

Ten female practicing SLPs who had at least 1 year of experience working in a setting with children who have dysarthria participated in this study. The SLPs worked in pediatric medical settings from around the United States, including outpatient clinics, inpatient rehabilitation, and private practice. They were recruited by word of mouth and included professional contacts known by the investigators to have expertise in pediatric dysarthria. We also employed snowball sampling, where SLPs recommended other potential SLPs based on their professional contacts. The included SLPs had an average of 9.3 years of clinical experience ($SD$ = 4.9 years, $Mdn$ = 9 years, range: 3–20 years). They reported an average of 7.9 years of experience listening to children with dysarthria ($SD$ = 5.2 years, $Mdn$ = 6.8 years, range: 1–13 years) and an average of 4.8 years of experience listening to adults with dysarthria ($SD$ = 5.9 years, $Mdn$ = 2.5 years, range: 0.5–20 years).

### Children

Speech samples from children with dysarthria secondary to CP and children with typical development were the same as those reported in Allison and Hustad (2018a, 2018b). The CP group included 20 children with a medical diagnosis of CP who were 5 years of age ($M_{\text{age}}$ = 64.2

months, $SD$ = 3.5), composed of 13 girls and seven boys. All children with CP had a diagnosis of dysarthria based on direct clinical assessment. Details regarding speech motor assessment procedures are provided in Allison and Hustad (2018a, 2018b). Note that children with a range of dysarthria severity levels were included in order to capture the variety of motor speech presentations among children with CP. See Table 1 for demographic information on children with CP, including Gross Motor Function Classification System level (Palisano et al., 1997), speech intelligibility, and anatomic involvement.

The TD group also included 20 children who were 5 years of age ($M_{\text{age}}$ = 63.8 months, $SD$ = 3.1) and were matched to the children with CP for age and sex. All children passed speech, language, and hearing screenings, and parents reported a negative history for developmental delay or other childhood neurological conditions. Children in either group were excluded from the study if they could not repeat sentences of at least five words in length.

## Procedure

### Speech Samples

Speech samples from all children were collected with recording equipment of professional quality in a sound-attenuated suite using the same protocol. Procedures are presented in detail in our earlier studies (Allison & Hustad, 2018a, 2018b). Briefly, children repeated the same set of 60 sentences ranging from two to seven words in length and 42 single words taken from the Test of Children's Speech Plus (TOCS+; Hodge & Daniels, 2007), following a prerecorded model. Of particular interest for this study were elicited productions of 10 five-word sentences taken from the TOCS+. Recorded speech samples were edited to remove any extraneous noise, and peak amplitude was normalized to ensure consistent maximum loudness levels across samples. These samples were then presented to SLPs in the rating tasks described below.

### Online Judgment Tasks

Recordings of all children were presented to the SLPs via a secure, custom, web-based interface that they accessed from their home or workplace. SLPs provided expert perceptual judgments of children's speech samples by completing a series of three web-based tasks. The first task, a dysarthria judgment task, was not relevant to the current research questions and is not described here. SLPs all completed a hypernasality rating task, followed by a phonetic accuracy judgment task. They were instructed to use headphones when completing the tasks and asked to adjust their computer's volume to a comfortable listening level before beginning the experiment to avoid changing the volume during the experiment. The experiment took approximately 90 min to complete, and SLPs were asked to complete the tasks in one sitting. Images of the web-based user interface and the exact text of instructions given to SLPs for both tasks are included in Figures 1 and 2.

*Hypernasality rating task*. In this task, SLPs were asked to rate each child's degree of hypernasality on a 7-point Likert scale. Prior to beginning the task, SLPs were provided with examples of a speech sample with "normal nasality" (rating = 1) and "severe hypernasality" (rating = 7) as a frame of reference, as shown in Figure 1. This rating scale was adapted from prior research on auditory–perceptual ratings of dysarthria features in adults (Bunton et al., 2007). Because the purpose of this study was to assess reliability of clinician judgments of hypernasality as related to dysarthria, hyponasality was not included in the scale ratings. Reference samples were speech samples that were not included in the set to be rated; the normal sample was of a 5-year-old child with typical speech and language skills, and the severe sample was of a 5-year-old child with

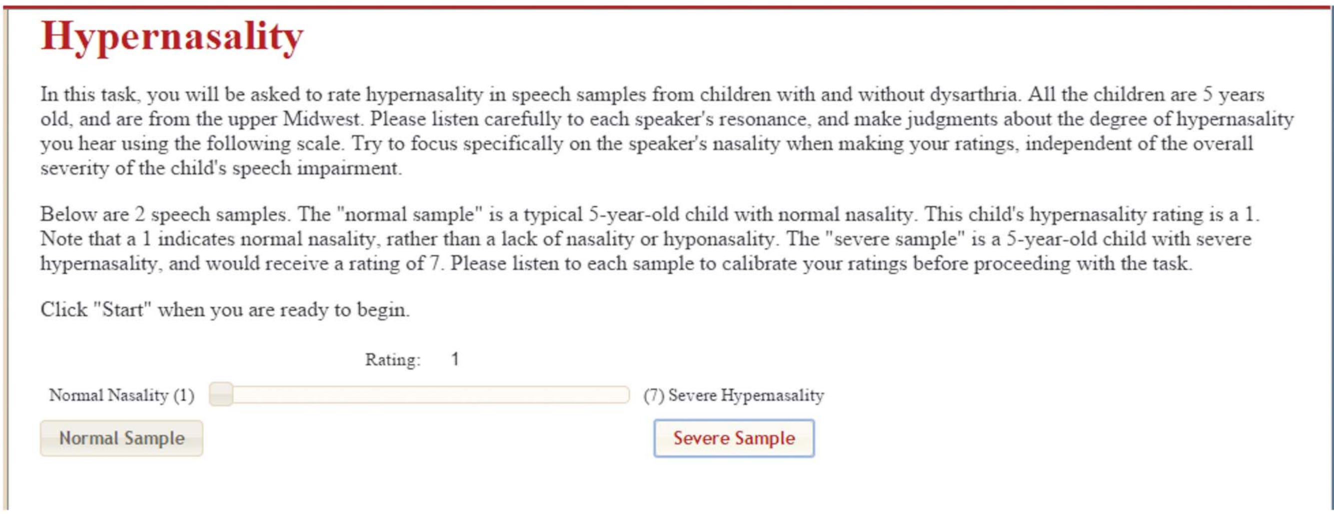
**Table 1.** Demographic characteristics of children who contributed speech samples.

| Characteristic | CP and dysarthria | Typical development |
|---|---|---|
| Sex, *n* | Female: 13<br>Male: 7 | Female: 13<br>Male: 7 |
| Age, *M* (*SD*) | 64.2 (3.5) months | 63.8 (3.1) months |
| Intelligibility,[a] *M* (*SD*) | 50.7% (20.3%) | 91.5% (11.9%) |
| GMFCS,[b] *n* | I: 7<br>II: 3<br>III: 3<br>IV: 6<br>V: 1 | |
| Anatomic involvement, *n* | Right hemiplegia: 7<br>Left hemiplegia: 2<br>Diplegia: 4<br>Quadriplegia: 6<br>Unknown: 1 | |

*Note.* CP = cerebral palsy.

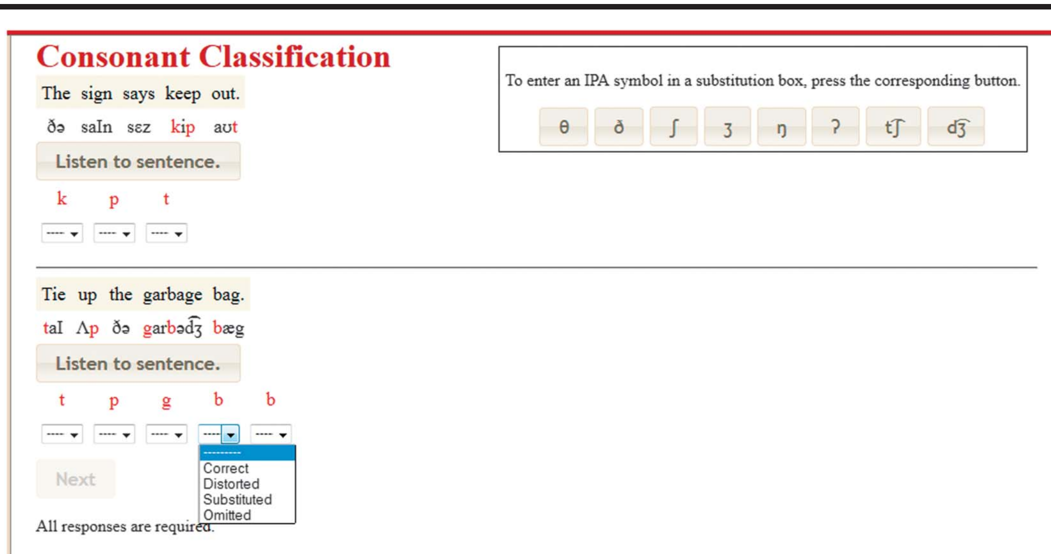
[a]Intelligibility was measured as the percentage of words correctly identified by listeners on words and sentences from the Test of Children's Speech Plus, as described in Allison and Hustad (2018a). [b]Gross Motor Function Classification System rating (I = *no/mild impairment,* V = *severe impairment*).

Figure 1. Screenshot of the nasality judgment task presented to speech-language pathologists (SLPs). After reading and listening to the instructions, SLPs clicked on the icons at the ends of the rating scale to hear the reference samples prior to beginning the experiment.




dysarthria. The reference samples were chosen as exemplars based on perceptual judgments of the principal investigator in consultation with a research SLP. During the task, each SLP was presented with 40 different speech samples, that is, one for each of the children in the study in a randomized order. Each child sample was composed of the same set of 10 five-word TOCS+ sentences, also randomized in order. Transcripts of the sentences were displayed on the screen along with speech samples to minimize the impact of reduced intelligibility on hypernasality ratings. After each speech sample was played once, the SLP assigned a rating between 1 and 7, representing a collective impression of hypernasality across all 10 utterances. After every five speech samples, the "normal" reference sample was presented again to help SLPs anchor their ratings. Anchor samples have been shown to improve reliability of auditory–perceptual ratings on equal-interval scales (Gerratt et al., 1993). Because each speech sample in this study was fairly long (i.e., approximately

Figure 2. Screenshot of the phonetic accuracy judgment task presented to speech-language pathologists (SLPs). SLPs were presented with the following instructions along with practice items prior to beginning the task: "In this task, you will be asked to make judgments regarding the accuracy of children's productions of specific consonant sounds. You will be presented with two sentences spoken by the same child. For each sentence, you will be asked to determine whether the consonants highlighted in red are correct, substituted, distorted, or omitted. You can listen to each sentence up to 5 times, by clicking the 'Listen to sentence' button." IPA = International Phonetic Alphabet.



20–30 s in duration), we chose to present the SLPs with only the "normal" reference as a repeated anchor during the experiment rather than with both the "normal" and "severe" reference samples provided in the orientation to the task. SLPs were instructed to listen specifically to each child's resonance characteristics and judge hypernasality independent of overall severity of speech involvement.

Each SLP provided a hypernasality rating for each child (10 SLPs × 40 children = 400 ratings). For each child, ratings from the 10 SLPs were used to obtain a mean and a standard deviation of hypernasality ratings.

*Phonetic accuracy judgment task.* In the second rating task, SLPs were asked to judge the accuracy of eight stop consonant sounds within the same two sentences from the TOCS+: "The sign says keep out" and "Tie up the garbage bag." Each SLP rated the accuracy of the eight target consonants produced by each of the 40 children (8 consonants × 40 children = 320 total ratings per SLP). Specific sentences were chosen for their high proportion of stop consonants (/k/, /t/, /p/, /g/, and /b/), which were of particular interest in this study.

Prior to beginning the task, SLPs were provided with a set of two practice sentences to orient them to the task. An orthographic and phonetic transcription of the two TOCS+ sentences was presented on the screen, with the target consonants to be judged highlighted in red, as shown in Figure 2. SLPs clicked a button next to the written sentence to play each child's production of that sentence. After listening to the sentence, they were asked to indicate whether each target stop consonant was correct, substituted, distorted, or omitted. When they chose "substituted," SLPs were prompted to type or select the International Phonetic Alphabet symbol of the sound that was produced instead of the target consonant. They were permitted to listen to the recording of each sentence up to 5 times per child. The two sentences produced by each child were presented as a set on the screen, and the order of presentation of the children was randomized for each SLP. After judging all the target phonemes for a child, SLPs were asked to rate their overall confidence in their consonant accuracy judgments for that child.

For each child, we tallied the number of SLPs judging each phoneme as correct, distorted, substituted, or omitted, yielding a matrix of values (8 phonemes × 4 possible ratings). These matrices were used to generate an interrater reliability value for each child, using Randolph's free-marginal kappa (Randolph, 2005).

### Speech Measures

Six speech measures were investigated as potential predictors of SLPs' reliability for phonetic accuracy and hypernasality judgments. Four acoustic measures and intelligibility levels were previously obtained from the same set of TOCS+ sentences for all child speakers. Acoustic and intelligibility measurement procedures are described in detail in prior publications (Allison & Hustad, 2018a, 2018b) and were chosen to reflect children's speech characteristics across speech subsystems. These four acoustic measures were reexamined in this study. Measures were *F2 range of diphthongs*, *proportion of bursts produced, proportion of deviant voice quality,* and *articulation rate*. *Intelligibility* was measured for each child as the average percentage of words accurately identified by five naive listeners based on an orthographic transcription of the TOCS+ sentences. Each listener transcribed speech samples from only one child, and five different listeners provided orthographic transcriptions for each child. Intelligibility data were collected as part of a larger ongoing longitudinal study (Hustad et al., 2020).

For this study, we also sought to examine each child's segmental accuracy as a predictor of SLPs' reliability. To do this, we determined a final accuracy status for each target stop consonant in order to calculate *percent stops correct* for each child. Because only the accuracy of stop consonants was judged by SLPs in this study, percent stops correct was calculated as an index of segmental accuracy in lieu of the more common metric, namely, percent consonants correct (Shriberg et al., 1997). In cases where at least eight of the 10 SLPs agreed on the status of a phoneme (i.e., correct, distorted, substituted, or omitted), it was considered the "true" status of the phoneme. In cases where three or more SLPs disagreed on the phoneme status, the first and the second author independently judged the accuracy of each phoneme, using spectrograms to guide judgments. This was done for 57 out of 320 total phonemes (six from children with typical development, 51 from children with dysarthria). The first and the second author agreed on 68% of analyzed tokens (39/57), yielding a final status for those phonemes. The remaining 18 disagreements were resolved through consensus coding by the first and the second author. The final status of stop consonants produced by children in the dysarthria and TD groups is shown in Table 2.

## Statistical Analysis

### Interrater Reliability

*Phonetic accuracy judgments.* Interrater reliability of the 10 SLPs for phonetic accuracy judgments was calculated using Randolph's free-marginal multirater kappa for each child speaker (Randolph, 2005; Warrens, 2010). Using free-marginal kappa is recommended for assessing interrater reliability of categorical data when raters do not have to assign a specified number of items to each category

**Table 2.** Distribution of errors in production of target stop consonants.

| Phoneme status | CP | TD |
|---|---|---|
| Correct | 73.75% | 99.38% |
| Distorted | 15.63% | 0.63% |
| Omitted | 5% | 0% |
| Substituted | 5.63% | 0% |

*Note.* Percentages are calculated out of the 160 total phonemes produced by children in the group (8 phonemes × 20 children per group). CP = children with cerebral palsy and dysarthria; TD = children with typical development/typically developing children.

(Brennan & Prediger, 1981). For each child speaker, the 10 SLPs' categorical accuracy judgments of the child's eight target phonemes were used to generate a kappa value for that child. An independent-samples *t* test was used to determine whether kappa values for the children with dysarthria differed from those for children with typical development.

*Hypernasality ratings.* For each child, interrater reliability of hypernasality judgments was calculated as the standard deviation of the 10 SLPs' hypernasality ratings (i.e., lower standard deviations reflect less variability in ratings across SLPs). The standard deviation of ratings was used to index reliability instead of the more traditional reliability measures (e.g., intraclass correlation) because it allowed us to quantify reliability for each individual child. An independent-samples *t* test was used to determine whether standard deviations of hypernasality ratings significantly differed between children with dysarthria and those with typical development.

#### Predictors of Interrater Reliability for Children With Dysarthria

A multiple linear regression approach was used to determine how well the speech features of children with dysarthria predicted SLPs' phonetic and hypernasality judgment reliability. First, the six speech features (i.e., F2 range of diphthongs, proportion of bursts produced, proportion of deviant voice quality, articulation rate, intelligibility, and percent stops correct) were correlated with the interrater reliability values for phonetic accuracy judgments (Randolph's free-marginal kappas) and hypernasality judgments (standard deviation of hypernasality ratings) to descriptively examine the relationship between each speech variable and reliability measures. Then, for each reliability measure, speech features were entered into a linear regression model in order of decreasing strength of correlation. Likelihood ratio tests were used to compare successively complex models and evaluate changes in model fit. Factors were added to the model one at a time until including an additional factor no longer significantly improved model fit in order to determine the most parsimonious model that best fit the data.

## Results

### Interrater Reliability of Phonetic Accuracy and Hypernasality Judgments

Group differences in interrater reliability of phonetic accuracy and hypernasality judgments by SLPs are shown in Figures 3a and 3b, respectively. Descriptively, Randolph's free-marginal kappa values indicated that, on average, interrater reliability of phonetic accuracy judgments among the 10 SLPs was moderate for children with dysarthria, mean $K = .61$ ($SE = .03$), and high for children with typical development, mean $K = .95$ ($SE = .01$). Results of an independent-samples *t* test showed that interrater reliability of phonetic accuracy judgments was significantly lower for children with dysarthria than for children with typical development, $t(38) = -9.24$, $p < .001$.

SLPs' hypernasality judgments were more variable for children with dysarthria than for children with typical development; on average, standard deviations of hypernasality ratings for children with dysarthria were 1.7 points on a 7-point rating scale, whereas standard deviations of hypernasality ratings for children with typical development were 0.7 points. Results of an independent-samples *t* test showed that standard deviations of hypernasality ratings were significantly larger for children with dysarthria than for children with typical development, $t(38) = 8.801$, $p < .001$, indicating lower interrater reliability for the dysarthria group. In order to descriptively summarize the hypernasality rating data, the 7-point scale ratings were grouped as follows: 1 = *normal,* 2–3 = *mild hypernasality,* 4–5 = *moderate hypernasality,* and 6–7 = *severe hypernasality*. SLPs' hypernasality ratings for individual children in the TD and dysarthria groups are shown in Figure 4. Among the children with CP and dysarthria, 11 out of the 20 children had hypernasality ratings ranging from normal to severe hypernasality across SLPs. Among the children with typical development, all 20 were rated by at least one SLP as having mild hypernasality, and three had ratings ranging from normal to moderate hypernasality.

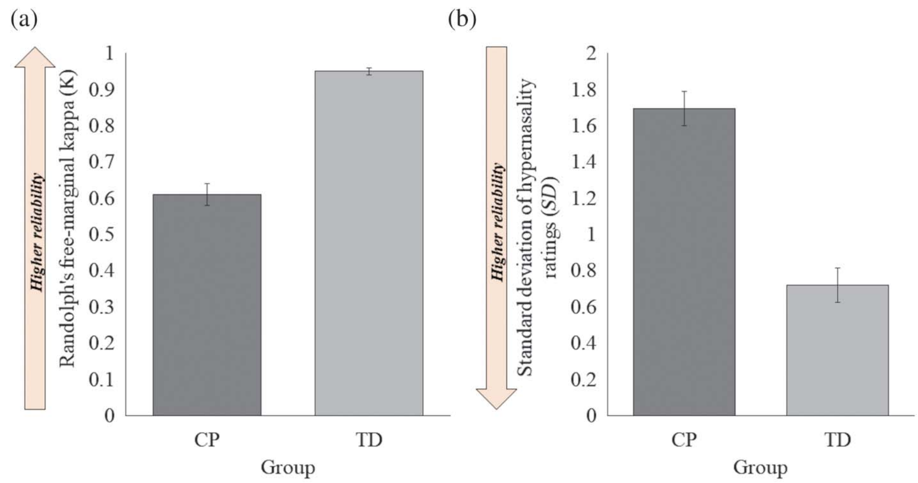
### Predictors of Interrater Reliability for Children With Dysarthria

Correlations between the six speech measures and measures of interrater reliability for the 20 children with CP and dysarthria are presented in Table 3. Three speech measures showed moderate to strong positive correlations with interrater reliability of SLPs' phonetic accuracy judgments (Randolph's *K*): percent stops correct, intelligibility, and F2 range of diphthongs. Two speech measures, namely, percent stops correct and proportion of bursts produced, had a moderate positive correlation with standard deviation of hypernasality ratings. It is important to note that this positive correlation indicates that larger standard deviations in hypernasality ratings, and thus lower reliability, were associated with higher stop consonant accuracy in children with dysarthria.

Results of multiple linear regression models for phonetic accuracy judgment reliability showed that the best-fitting model contained only percent stops correct as a predictor of Randolph's kappa ($\beta = .51$, $p < .001$, $R^2 = .64$). Results of likelihood ratio tests showed that including additional speech variables in the model did not significantly improve model fit.

Results of multiple linear regression models for hypernasality judgment reliability also showed that the best-fitting model contained only percent stops correct as a predictor of the standard deviation in SLPs' hypernasality ratings ($\beta = .61$, $p = .02$, $R^2 = .23$). Results of likelihood ratio tests showed that including additional speech variables in the

**Figure 3.** Group differences in interrater reliability of speech-language pathologists' (a) phonetic accuracy judgments and (b) hypernasality ratings between children with cerebral palsy and dysarthria (CP group) and children with typical development (TD group). Bars indicate means, and error bars indicate *SE*. Please note that the direction of higher reliability differs between the two measures.




model did not significantly improve model fit. Relationships between percent stops correct and interrater reliability measures for the children with CP and dysarthria are shown in Figure 5.

## Discussion

The objective of this study was to investigate interrater reliability among practicing SLPs for perceptual judgments of phoneme accuracy and hypernasality in connected speech from children with dysarthria. Results of this study revealed two primary findings: (a) SLPs are less reliable in their phonetic accuracy and hypernasality judgments of children with dysarthria than in their phonetic accuracy and hypernasality judgments of children with typical speech development, and (b) among children with dysarthria secondary to CP, reduced speech sound accuracy may interfere with reliability of SLPs' perceptual judgments.

**Figure 4.** Distribution of speech-language pathologists' (SLPs') nasality ratings for (a) children with cerebral palsy and dysarthria (CP group) and (b) children with typical development (TD group). Light-gray bars indicate the number of SLPs who rated the child as having normal nasality (i.e., a rating of 1), medium-gray bars indicate the number of SLPs who rated the child as having mild hypernasality (i.e., a rating of 2–3), dark-gray bars indicate the number of SLPs who rated the child as having moderate hypernasality (i.e., a rating of 4–5), and black bars indicate the number of SLPs who rated the child as having severe hypernasality (i.e., a rating of 6–7). Children in both groups are ordered from lowest to highest average nasality rating.

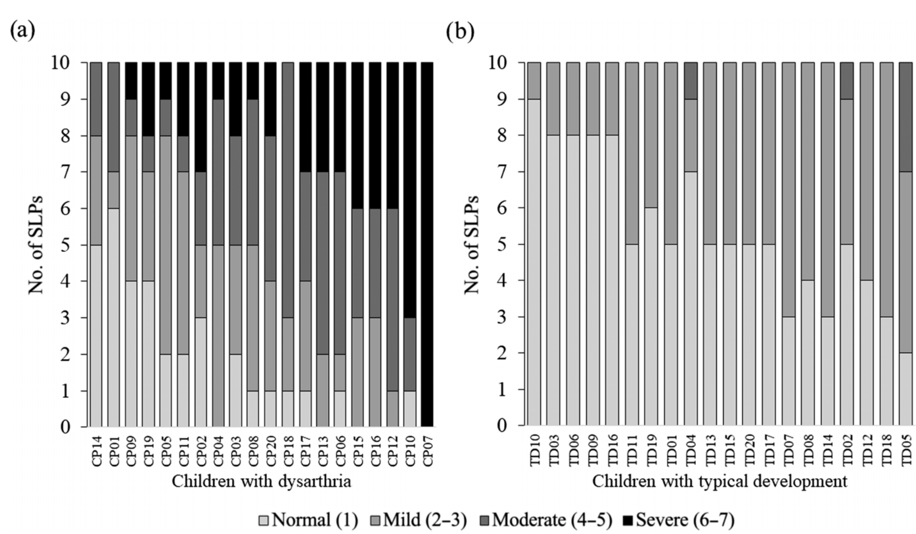


**Table 3.** Correlations between speech variables and measures of interrater reliability for children with dysarthria.

| Variable | Randolph's *K* *r* | Hypernasality *SD* *r* |
|---|---|---|
| F2 range | .55 | .10 |
| Prop. bursts | .28 | .40 |
| Prop. deviant voice | −.25 | −.20 |
| Articulation rate | .34 | .04 |
| Intelligibility | .48 | < .01 |
| % stops correct | .80 | .49 |

*Note.* *SD* = standard deviation; F2 = second formant frequency; Prop. = proportion (of); % = percent.

## *SLPs' Perceptual Judgments Are Less Reliable for Children With Dysarthria Than for Children With Typical Speech Development*

### Phonetic Accuracy Judgments

Results showed that phonetic accuracy judgments by SLPs were significantly less reliable for 5-year-old children with dysarthria than for age-matched TD peers. For the children with typical development, reliability of phonetic accuracy judgments was high, and children made very few errors on production of the target consonants. The high accuracy is consistent with normative information on speech sound development (McLeod & Crowe, 2018) and likely accounts for the high degree of reliability among SLPs. When production accuracy is high or within age-level expectations, there is little room for disagreement among professionals.

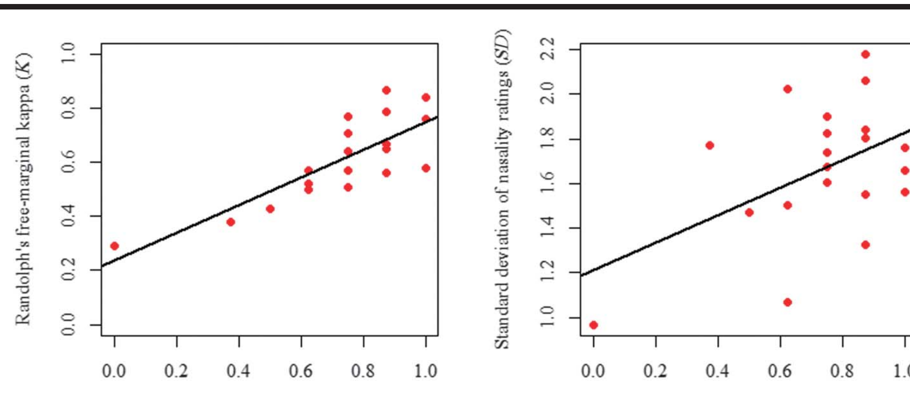
The children with dysarthria in this study had more errors in producing the target stop consonants overall compared to the TD group, although the majority of stops were produced accurately. Distortions were the most common type of error observed. Prior research has shown that narrow transcription used to characterize distortions is particularly challenging for reliability (Shriberg & Lof, 1991). In this study, SLPs did not have to transcribe the distorted speech sounds but, rather, only to make a categorical determination regarding whether sounds were correct, distorted, substituted, or omitted. Although this task is considerably simpler than a narrow transcription task, distortions produced by children with dysarthria are complex due to the involvement of multiple speech subsystems (e.g., may deviate in resonance, voice quality, and articulatory precision). SLPs likely differ in their perceptual thresholds for categorizing these deviant productions that fall between categorical boundaries. For example, a partially devoiced /d/ could be judged as a distortion or a substitution depending on the rater. In addition, the multiple disordered-speech features that globally affect the acoustic signal produced by children with dysarthria may have been challenging for SLPs to perceptually separate from phonetic accuracy when making these judgments, particularly in the context of connected speech.

It is important to note that the phonetic accuracy judgment task used in this study was less complex than what clinicians typically do when administering standardized articulation tests because it involved judging a very limited set of phonemes in only two sentences. In addition, SLPs were able to listen to the same productions repeatedly, a luxury that is not always possible in the clinical administration of an articulation test. Thus, we might expect this type of task to yield a best-case scenario for reliability; however, even with the highly focused and somewhat idealized nature of the task, SLPs' reliability was moderate for the children with dysarthria. In clinical practice, reliability is likely to be even lower because SLPs have to judge the accuracy of all phonemes to administer tests and obtain a full representation of the child's speech sound inventory.

### Hypernasality Ratings

Hypernasality ratings by SLPs were also significantly less reliable for children in the dysarthria group than for children in the TD group. Interestingly, all of the TD children were rated by at least one SLP as having mild hypernasality.

**Figure 5.** Relations between proportion of stops correct and (a) reliability of speech-language pathologists' (SLPs') phonetic accuracy judgments (Randolph's free-marginal kappa) and (b) reliability of SLPs' nasality ratings for children with cerebral palsy and dysarthria.



This raises an important issue of how global speech characteristics such as resonance, which are the gold standard for describing dysarthric speech, develop throughout childhood. A recent study by Schölderle et al. (2020) examined developmental trajectories in perceptual speech features typically associated with dysarthria among TD German-speaking children between the ages of 3 and 9 years. Results showed that, at young ages, many TD children present with speech features typically associated with dysarthria. Normalization toward an adultlike standard with age was found for all auditory–perceptual features studied, including resonance, although resonance disturbances were less prevalent in young children than other features (Schölderle et al., 2020). In addition, craniofacial development affects resonance, and hyponasality in young children has been explained by a relatively small nasal cavity (Awan, 2001; Brunnegård & van Doorn, 2009). Although SLPs were not asked to rate hyponasality in this task, it is possible that hyponasality in the children with typical development may have contributed to SLPs' ratings of atypical resonance in this group. The lack of a "normal resonance" perceptual measurement standard for children at different ages and the large variation in resonance among young TD children exacerbate the difficulties with perceptual rating of hypernasality in children with dysarthria. Dialect differences are known to impact measures of nasalance (Awan et al., 2015; Gildersleeve-Neumann & Dalston, 2001) and may also impact auditory–perceptual ratings of hypernasality. Children in this study were from the Upper Midwest (Inland North and North Central dialectal regions; Labov et al., 1997), whereas SLPs were geographically distributed throughout the United States and likely had varying dialects. Findings from this study underscore the need for more research on how to clinically differentiate typical from atypical resonance in young children.

For the majority of the children with dysarthria, hypernasality ratings spanned from normal to severe hypernasality, indicating substantial variation in how resonance was perceived across SLPs. This is consistent with prior research on reliable perceptual judgment of hypernasality in adults with dysarthria (Bunton et al., 2007). Perception of hypernasality can be influenced by other speech characteristics, such as reduced speaking rate (McHenry, 1999). Children with dysarthria secondary to CP often have many coinciding disordered speech characteristics, including deviations in voice quality, rate, prosody, and articulatory precision. Because these characteristics all globally affect the child's speech signal, they are difficult to perceptually separate from one another. In addition, hypernasality can be either constant or intermittent for children with dysarthria, further complicating its perception. The extremely wide variation in hypernasality ratings for many of the children with dysarthria in this study provides little confidence that the mean hypernasality rating across SLPs accurately reflects the child's true velopharyngeal function. Thus, these findings suggest that a quantitative measure of "ground truth" (i.e., nasalance) may be needed to supplement perceptual ratings and make conclusions about the severity of hypernasality for children with dysarthria.

### *Factors Affecting Reliability of Perceptual Judgments for Children With Dysarthria*

Among the children with dysarthria, reliability of SLPs' phonetic accuracy and hypernasality judgments ranged from poor to strong across children. Percent stops correct was the strongest predictor of interrater reliability for both perceptual dimensions. For the phonetic accuracy judgments, this may partially be due to the fact that percent stops correct was calculated on the same set of phonemes that were transcribed by SLPs for reliability judgments but suggests that correct productions are more reliably judged than errored productions.

Surprisingly, percent stops correct was also the only speech measure that significantly predicted reliability of SLPs' hypernasality ratings, which were globally based on children's complete set of sentence productions. Furthermore, results showed that hypernasality judgments of children with *higher* stop consonant accuracy were *less* reliable. This finding is in the opposite direction to the effect of stop consonant accuracy on reliability of phonetic accuracy judgments and suggests a complex relationship between segmental accuracy and reliability of auditory–perceptual judgments by SLPs. Although the "percent stops correct" measure used in this study provides a limited representation of the child's overall phonetic accuracy, production of stop consonants is particularly affected by velopharyngeal dysfunction, which may help explain this result. Production of stop consonants requires high intraoral pressure, which cannot be attained if velopharyngeal function is severely compromised. Research on perception of hypernasality in speakers with cleft palate has shown that plosives and fricatives/affricates are perceived as more nasal than other consonants (Moore & Sommers, 1973). Thus, SLPs' hypernasality ratings may have been partially influenced by the children's production of stop consonants. One child in this study had 0% accuracy of stop consonant production and consistent ratings of severe hypernasality across SLPs, suggesting that her velopharyngeal impairment was impacting her production of stop consonants and resulting in severe hypernasality that was easily detected by SLPs. This child's data also appeared to strongly influence the correlation between percent stops correct and standard deviation of hypernasality ratings. Additional data from other children with severe dysarthria are needed to validate this finding. Unlike children with cleft palate whose anatomy precludes velopharyngeal closure, hypernasality in children with dysarthria is physiologically based and often intermittent (Hardy, 1961; R. D. Kent & Netsell, 1978). The majority of children with dysarthria in this study had mild to moderate reductions in stop consonant accuracy. For children with fewer errors, stop consonant production likely provided fewer acoustic cues to SLPs regarding the child's hypernasality, resulting in more variable hypernasality ratings across SLPs. Although many dimensions influence perception of hypernasality (Kataoka et al., 2001; Warren et al., 1994), our findings suggest that reduced stop consonant accuracy may be one factor that makes the presence of hypernasality more reliably detected. Future research systematically examining the impact of segmental errors on perception of hypernasality and other global auditory–perceptual dysarthria features has important implications for clinical rating of these features in children with developing speech sound systems.

Acoustic measures and intelligibility had less influence on reliability of SLPs' phonetic accuracy and hypernasality judgments. Although one acoustic measure (i.e., F2 range of diphthongs) and intelligibility were moderately positively associated with reliability of SLPs' phonetic accuracy judgments, they did not significantly improve prediction of SLPs' reliability after accounting for percent stops correct. Measures of F2 range and slope have been used to index acoustic specification of diphthongs in dysarthria research for decades (J. F. Kent et al., 1992; Y. Kim et al., 2009; Lee et al., 2014) and indirectly provide a metric of articulatory excursion (Allison et al., 2020; Thompson & Kim, 2019). Reductions in F2 range and slope have been associated with dysarthria in adults with many different neurologic conditions and in children (Allison & Hustad, 2018a; Y. Kim et al., 2011; Lee et al., 2014) and have been proposed as a measure for indexing severity of articulatory involvement in dysarthria (Y. Kim et al., 2011, 2009). Measures of intelligibility have been widely used to index the overall severity of dysarthria in both adults and children (e.g., Hustad et al., 2019; Y. Kim et al., 2011; Weismer & Martin, 1992). Segmental accuracy is also reduced in children with more severe dysarthria (Nordberg et al., 2014). Thus, although F2 range of diphthongs and intelligibility did not significantly contribute to prediction of reliability of SLP's phonetic accuracy judgments, results collectively suggest that phonetic accuracy was less reliably judged for children with more severe dysarthria. This finding has important clinical implications because it suggests that phonetic accuracy judgments, frequently completed as part of standardized articulation tests, may not be a reliable way to evaluate phoneme integrity for children with severe dysarthria. The acoustic measures and intelligibility were not correlated with reliability of hypernasality ratings. Prior research has shown that reliability of listener ratings for auditory–perceptual dysarthria features, including hypernasality, is highest for individuals with mild or severe deviations in the given speech feature and least reliable for those in the middle of the severity spectrum (Bunton et al., 2007). Descriptively, our findings are consistent with this prior research, as there was the greatest variation in SLPs' hypernasality ratings for children whose mean hypernasality ratings were in the moderate range.

## Limitations and Future Directions

Results from this study should be considered preliminary and interpreted in the context of its limitations. First, we focused on a limited sample of phonemes in this study, and future research is needed to determine interrater reliability of phonetic accuracy judgments for other consonants and vowels in children with dysarthria, as well as potential differences across word positions. Furthermore, there was a very low rate of errors on the target consonants among TD children, which likely contributed to the high reliability. Future research should compare reliability of these judgments between children with dysarthria and children with non–motor-based SSDs. SLPs were allowed to listen to productions multiple times for phonetic accuracy judgments, which may have increased their reliability. Only consonant productions in the context of repeated sentences were judged in this study. Future research comparing these findings to reliability of phonetic accuracy judgments in single-word productions from children with dysarthria would provide helpful information about the effects of stimulus length on SLPs' reliability. This study also involved children within a very narrow age band, and it is possible that reliability of perceptual judgments may vary with age. In particular, judgment of hypernasality may be less reliable in younger children for whom a larger "normal" range may be expected among TD children. Lastly, percent stops correct in this study was based, to some degree, on agreement among SLPs' ratings, making this measure not completely independent of phonetic accuracy judgment reliability. Future work using independent phonetic accuracy judgments is needed to validate the relationship between phonetic accuracy and reliability.

## Clinical Implications

Findings from this study underscore the challenges with making reliable perceptual judgments of speech features and phonetic accuracy for children with dysarthria, even by experienced SLPs. Because these perceptual judgments remain the clinical gold standard for evaluating and describing dysarthria in children, there is a clear need for ways to improve reliability of these ratings among practicing clinicians. Development of formalized rating scales accompanied by perceptual trainings for SLPs may be one way to improve reliability. Prior research studies on various rating scales for children with dysarthria have shown that clinician training can be effective in improving reliability of perceptual judgments (Cunningham et al., 2016; Schölderle et al., 2020). There are currently no published clinical tests for evaluating auditory–perceptual dysarthria features in English-speaking children, although efforts are underway to develop such measures in other languages (Schölderle et al., 2020).

For some children with dysarthria (i.e., those with more severe speech motor involvement and/or more speech sound errors), reliable auditory–perceptual judgment may be difficult to attain; thus, alternate or supplemental ways to characterize segmental errors and speech characteristics may be needed to reliably describe their speech. Many acoustic measures have been used in research that can help with quantification of specific speech dimensions (R. D. Kent & Kim, 2003; Y. Kim et al., 2011; Saxon et al., 2019); however, few are practical to implement in clinical settings. Future research should focus on the development of tools that can be used by clinicians to improve reliable, quantitative description of speech for children with severe dysarthria.

It is also important for clinicians to keep in mind that segmental accuracy in children with dysarthria is affected not only by the impaired motor execution of speech movements but also by the child's phonological development and development of motor plans for speech sound production. Although there is no clear way to distinguish the differential effects of these coinciding influences on a child's production of a particular phoneme, measures of segmental accuracy for children with dysarthria should be interpreted as a reflection of their combined output. For children with severe chronic dysarthria, "normal" articulation of all speech sounds is likely not an attainable treatment goal. Thus, assessment of the accuracy of individual phonemes may be less important for treatment planning than evaluation of cues that can maximize intelligibility (e.g., Levy et al., 2017; Sakash et al., 2020). Dynamic testing (e.g., Strand et al., 2013), which allows multiple opportunities for the clinician to judge a child's articulation and intelligibility with different levels and types of cueing, may help the clinician identify strategies with the greatest potential to improve speech function.

## Acknowledgments

This project was supported by National Institute on Deafness and Other Communication Disorders Grants R01DC009411, awarded to Katherine C. Hustad, and F31DC013925-01, awarded to Kristen M. Allison. Support was also provided by National Institute of Child Health and Human Development Core Grant P30HD03352, awarded to the Waisman Center. We would like to thank the speech-language pathologists who participated in this study.

## References

**Allison, K. M., & Hustad, K. C.** (2018a). Acoustic predictors of pediatric dysarthria in cerebral palsy. *Journal of Speech, Language, and Hearing Research, 61*(3), 462–478. https://doi.org/10.1044/2017_JSLHR-S-16-0414

**Allison, K. M., & Hustad, K. C.** (2018b). Data-driven classification of dysarthria profiles in children with cerebral palsy. *Journal of Speech, Language, and Hearing Research, 61*(12), 2837–2853. https://doi.org/10.1044/2018_JSLHR-S-17-0356

**Allison, K. M., Salehi, S., & Green, J. R.** (2020). Effect of prosodic manipulation on articulatory kinematics and second formant trajectories in children. *The Journal of the Acoustical Society of America, 147*(2), 769–776. https://doi.org/10.1121/10.0000670

**Ansel, B. M., & Kent, R. D.** (1992). Acoustic-phonetic contrasts and intelligibility in the dysarthria associated with mixed cerebral palsy. *Journal of Speech and Hearing Research, 35*(2), 296–308. https://doi.org/10.1044/jshr.3502.296

**Awan, S. N.** (2001, 1–2). Age and gender effects on measures of RMS nasalance. *Clinical Linguistics & Phonetics, 15*(122), 117. https://doi.org/10.3109/02699200109167642

**Awan, S. N., Bressmann, T., Poburka, B., Roy, N., Sharp, H., & Watts, C.** (2015). Dialectical effects on nasalance: A multicenter, cross-continental study. *Journal of Speech, Language, and Hearing Research, 58*(1), 69–77. https://doi.org/10.1044/2014_JSLHR-S-14-0077

**Bettens, K., Alighieri, C., Bruneel, L., De Meulemeester, L., & Van Lierde, K.** (2020). Peer attitudes toward children with cleft (lip and) palate related to speech intelligibility, hypernasality and articulation. *Journal of Communication Disorders, 85,* 105991. https://doi.org/10.1016/j.jcomdis.2020.105991

**Brennan, R. L., & Prediger, D. J.** (1981). Coefficient kappa: Some uses, misuses, and alternatives. *Educational and Psychological Measurement, 41*(3), 687–699. https://doi.org/10.1177/001316448104100307

**Brunnegård, K., Lohmander, A., & van Doorn, J.** (2012). Comparison between perceptual assessments of nasality and nasalance scores. *International Journal of Language & Communication Disorders, 47*(5), 556–566. https://doi.org/10.1111/j.1460-6984.2012.00165.x

**Brunnegård, K., & van Doorn, J.** (2009). Normative data on nasalance scores for Swedish as measured on the Nasometer: Influence of dialect, gender, and age. *Clinical Linguistics & Phonetics, 23*(1), 58–69. https://doi.org/10.1080/02699200802491074

**Bunton, K., Kent, R. D., Duffy, J. R., Rosenbek, J. C., & Kent, J. F.** (2007). Listener agreement for auditory–perceptual ratings of dysarthria. *Journal of Speech, Language, and Hearing Research, 50*(6), 1481–1495. https://doi.org/10.1044/1092-4388(2007/102)

**Cunningham, B. J., Rosenbaum, P., & Hidecker, M. J. C.** (2016). Promoting consistent use of the Communication Function Classification System (CFCS). *Disability and Rehabilitation, 38*(2), 195–204. https://doi.org/10.3109/09638288.2015.1027009

**Duffy, J. R.** (2005). *Motor speech disorders: Differential diagnosis and management*. Mosby.

**Gerratt, B. R., Kreiman, J., Antonanzas-Barroso, N., & Berke, G. S.** (1993). Comparing internal and external standards in voice quality judgments. *Journal of Speech and Hearing Research, 36*(1), 14–20. https://doi.org/10.1044/jshr.3601.14

**Gildersleeve-Neumann, C. E., & Dalston, R. M.** (2001). Nasalance scores in noncleft individuals: Why not zero? *The Cleft Palate–Craniofacial Journal, 38*(2), 106–111. https://doi.org/10.1597/1545-1569_2001_038_0106_nsiniw_2.0.co_2

**Goldman, R., & Fristoe, M.** (2015). *GFTA-3: Goldman-Fristoe Test of Articulation 3*. The Psychological Corporation.

**Hardy, J. C.** (1961). Intraoral breath pressure in cerebral palsy. *Journal of Speech and Hearing Disorders, 26*(4), 309–319. https://doi.org/10.1044/jshd.2604.309

**Hodge, M., & Daniels, J.** (2007). *TOCS+ intelligibility measures*. University of Alberta.

**Hustad, K. C., Mahr, T., Broman, A. T., & Rathouz, P. J.** (2020). Longitudinal growth in single-word intelligibility among children with cerebral palsy from 24 to 96 months of age: Effects of speech-language profile group membership on outcomes. *Journal of Speech, Language, and Hearing Research, 63*(1), 32–48. https://doi.org/10.1044/2019_JSLHR-19-00033

**Hustad, K. C., Sakash, A., Natzke, P., Broman, A. T., & Rathouz, P. J.** (2019). Longitudinal growth in single-word intelligibility among children with cerebral palsy from 24 to 96 months of age: Predicting later outcomes from early speech production. *Journal of Speech, Language, and Hearing Research, 62*(6), 1599–1613. https://doi.org/10.1044/2018_JSLHR-S-18-0319

**Kataoka, R., Warren, D. W., Zajac, D. J., Mayo, R., & Lutz, R. W.** (2001). The relationship between spectral characteristics and perceived hypernasality in children. *The Journal of the Acoustical Society of America, 109*(5), 2181–2189. https://doi.org/10.1121/1.1360717

**Kent, J. F., Kent, R. D., Rosenbek, J. C., Weismer, G., Martin, R., Sufit, R., & Brooks, B. R.** (1992). Quantitative description of the dysarthria in women with amyotrophic lateral sclerosis.

Journal of Speech and Hearing Research, 35(4), 723–733. https://doi.org/10.1044/jshr.3504.723
**Kent, R. D.** (1996). Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders. *American Journal of Speech-Language Pathology, 5*(3), 7–23. https://doi.org/10.1044/1058-0360.0503.07
**Kent, R. D., & Kim, Y.-J.** (2003). Toward an acoustic typology of motor speech disorders. *Clinical Linguistics & Phonetics, 17*(6), 427–445. https://doi.org/10.1080/0269920031000086248
**Kent, R. D., & Netsell, R.** (1978). Articulatory abnormalities in athetoid cerebral palsy. *Journal of Speech and Hearing Disorders, 43*(3), 353–373. https://doi.org/10.1044/jshd.4303.353
**Kent, R. D., Weismer, G., Kent, J. F., & Rosenbek, J. C.** (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders, 54*(4), 482–499. https://doi.org/10.1044/jshd.5404.482
**Kim, H., Martin, K., Hasegawa-Johnson, M., & Perlman, A.** (2010). Frequency of consonant articulation errors in dysarthric speech. *Clinical Linguistics & Phonetics, 24*(10), 759–770. https://doi.org/10.3109/02699206.2010.497238
**Kim, Y., Kent, R. D., & Weismer, G.** (2011). An acoustic study of the relationships among neurologic disease, dysarthria type, and severity of dysarthria. *Journal of Speech, Language, and Hearing Research, 54*(2), 417–429. https://doi.org/10.1044/1092-4388(2010/10-0020)
**Kim, Y., Weismer, G., Kent, R. D., & Duffy, J. R.** (2009). Statistical models of F2 slope in relation to severity of dysarthria. *Folia Phoniatrica et Logopaedica, 61*(6), 329–335. https://doi.org/10.1159/000252849
**Kummer, A. W.** (2011). Speech therapy for errors secondary to cleft palate and velopharyngeal dysfunction. *Seminars in Speech and Language, 32*(2), 191–198. https://doi.org/10.1055/s-0031-1277721
**Labov, W., Ash, S., & Boberg, C.** (1997). *A national map of the regional dialects of American English* (Telsur Project). Department of Linguistics, University of Pennsylvania.
**Lee, J., Hustad, K. C., & Weismer, G.** (2014). Predicting speech intelligibility with a multiple speech subsystems approach in children with cerebral palsy. *Journal of Speech, Language, and Hearing Research, 57*(5), 1666–1678. https://doi.org/10.1044/2014_JSLHR-S-13-0292
**Levy, E. S., Chang, Y. M., Ancelle, J. A., & McAuliffe, M. J.** (2017). Acoustic and perceptual consequences of speech cues for children with dysarthria. *Journal of Speech, Language, and Hearing Research, 60*(6S), 1766–1779. https://doi.org/10.1044/2017_JSLHR-S-16-0274
**Liss, J. M., Spitzer, S., Caviness, J. N., Adler, C., & Edwards, B.** (1998). Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech. *The Journal of the Acoustical Society of America, 104*(4), 2457–2466. https://doi.org/10.1121/1.423753
**Liu, H. M., Tseng, C.-H., & Tsao, F. M.** (2000). Perceptual and acoustic analysis of speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *Clinical Linguistics & Phonetics, 14*(6), 447–464. https://doi.org/10.1080/026992000415877
**Maegawa, J., Sells, R. K., & David, D. J.** (1998). Speech changes after maxillary advancement in 40 cleft lip and palate patients. *The Journal of Craniofacial Surgery, 9*(2), 177–182. https://doi.org/10.1097/00001665-199803000-00017
**McHenry, M. A.** (1999). Aerodynamic, acoustic, and perceptual measures of nasality following traumatic brain injury. *Brain Injury, 13*(4), 281–290. https://doi.org/10.1080/026990599121656
**McLeod, S., & Crowe, K.** (2018). Children's consonant acquisition in 27 languages: A cross-linguistic review. *American Journal of Speech-Language Pathology, 27*(4), 1546–1571. https://doi.org/10.1044/2018_AJSLP-17-0100
**McSweeny, J. L., Shriberg, L. D., Kwiatkowski, J., McSweeny, J., Rasmussen, C., & Widder, C.** (1995). *Segmental and suprasegmental transcription reliability* (No. 2) [Technical report].
**Moore, J., & Sommers, R. K.** (1973). Phonetic contexts: Their effects on perceived nasality in cleft palate speakers. *The Cleft Palate Journal, 10*(1), 72–83.
**Nordberg, A., Miniscalco, C., & Lohmander, A.** (2014). Consonant production and overall speech characteristics in school-aged children with cerebral palsy and speech impairment. *International Journal of Speech-Language Pathology, 16*(4), 386–395. https://doi.org/10.3109/17549507.2014.917440
**Palisano, R., Rosenbaum, P., Walter, S., Russell, D., Wood, E., & Galuppi, B.** (1997). Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Developmental Medicine & Child Neurology, 39*(2), 214–223. https://doi.org/10.1111/j.1469-8749.1997.tb07414.x
**Platt, L. J., Andrews, G., & Howie, P. M.** (1980). Dysarthria of adult cerebral palsy: II. Phonemic analysis of articulation errors. *Journal of Speech and Hearing Research, 23*(1), 41–55. https://doi.org/10.1044/jshr.2301.41
**Platt, L. J., Andrews, G., Young, M., & Quinn, P. T.** (1980). Dysarthria of adult cerebral palsy: I. Intelligibility and articulatory impairment. *Journal of Speech and Hearing Research, 23*(1), 28–40. https://doi.org/10.1044/jshr.2301.28
**Randolph, J. J.** (2005). *Free-marginal multirater kappa: An alternative to Fleiss' fixed-marginal multirater kappa* [Paper presentation]. Joensuu University Learning and Instruction Symposium 2005, Joensuu, Finland.
**Sakash, A., Mahr, T. J., Natzke, P. E., & Hustad, K. C.** (2020). Effects of rate manipulation on intelligibility in children with cerebral palsy. *American Journal of Speech-Language Pathology, 29*(1), 127–141. https://doi.org/10.1044/2019_AJSLP-19-0047
**Saxon, M., Liss, J., & Berisha, V.** (2019, May). Objective measures of plosive nasalization in hypernasal speech. In *ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 6520–6524). IEEE. https://doi.org/10.1109/ICASSP.2019.8682339
**Schölderle, T., Haas, E., & Ziegler, W.** (2020). Age norms for auditory-perceptual neurophonetic parameters: A prerequisite for the assessment of childhood dysarthria. *Journal of Speech, Language, and Hearing Research, 63*(4), 1071–1082. https://doi.org/10.1044/2020_JSLHR-19-00114
**Shriberg, L. D., Austin, D., Lewis, B. A., McSweeny, J. L., & Wilson, D. L.** (1997). The Percentage of Consonants Correct (PCC) metric: Extensions and reliability data. *Journal of Speech, Language, and Hearing Research, 40*(4), 708–722. https://doi.org/10.1044/jslhr.4004.708
**Shriberg, L. D., & Lof, G. L.** (1991). Reliability studies in broad and narrow phonetic transcription. *Clinical Linguistics & Phonetics, 5*(3), 225–279. https://doi.org/10.3109/02699209108986113
**Strand, E. A., McCauley, R. J., Weigand, S. D., Stoeckel, R. E., & Baas, B. S.** (2013). A motor speech assessment for children with severe speech disorders: Reliability and validity evidence. *Journal of Speech, Language, and Hearing Research, 56*(2), 505–520. https://doi.org/10.1044/1092-4388(2012/12-0094)
**Sweeney, T., & Sell, D.** (2008). Relationship between perceptual ratings of nasality and nasometry in children/adolescents with cleft palate and/or velopharyngeal dysfunction. *International Journal of Language & Communication Disorders, 43*(3), 265–282. https://doi.org/10.1080/13682820701438177
**Thompson, A., & Kim, Y.** (2019). Relation of second formant trajectories to tongue kinematics. *The Journal of the Acoustical*

*Society of America, 145*(4), EL323–EL328. https://doi.org/10.1121/1.5099163
**Warren, D. W., Dalston, R. M., & Mayo, R.** (1994). Hypernasality and velopharyngeal impairment. *The Cleft Palate–Craniofacial Journal, 31*(4), 257–262. https://doi.org/10.1597/1545-1569_1994_031_0257_havi_2.3.co_2
**Warrens, M. J.** (2010). Inequalities between multi-rater kappas. *Advances in Data Analysis and Classification, 4*(4), 271–286. https://doi.org/10.1007/s11634-010-0073-4
**Watterson, T.** (2020). The use of the nasometer and interpretation of nasalance scores. *Perspectives of the ASHA Special Interest Groups, 5*(1), 155–163. https://doi.org/10.1044/2019_PERSP-19-00029
**Weismer, G., & Martin, R.** (1992). Acoustic and perceptual approaches to the study of intelligibility. In R. D. Kent (Ed.), *Intelligibility in speech disorders* (Vol. 1). John Benjamins. https://doi.org/10.1075/sspcl.1.04wei
**Workinger, M. S., & Kent, R. D.** (1991). Perceptual analysis of the dysarthrias in children with athetoid and spastic cerebral palsy. In C. A. Moore, K. M. Yorkston, & D. R. Beukelman (Eds.), *Dysarthria and apraxia of speech: Perspectives on management* (pp. 109–126). Brookes.
**Yorkston, K. M., & Beukelman, D. R.** (1981). Ataxic dysarthria: Treatment sequences based on intelligibility and prosodic considerations. *Journal of Speech and Hearing Disorders, 46*(4), 398–404. https://doi.org/10.1044/jshd.4604.398