International Journal of Speech-Language Pathology

ISSN: 1754-9507 (Print) 1754-9515 (Online) Journal homepage: www.tandfonline.com/journals/iasl20

Validity of acoustic speech measures obtained through videoconferencing with children with dysarthria

Kyunghae Hwang, Frits van Brenk, Megan J. McAuliffe, Jiyoung Choi, Jan G. Švec, Young Hwa M. Chang, Bryan Keller & Erika S. Levy

To cite this article: Kyunghae Hwang, Frits van Brenk, Megan J. McAuliffe, Jiyoung Choi, Jan G. Švec, Young Hwa M. Chang, Bryan Keller & Erika S. Levy (24 Oct 2025): Validity of acoustic speech measures obtained through videoconferencing with children with dysarthria, International Journal of Speech-Language Pathology, DOI: 10.1080/17549507.2025.2563845

To link to this article: https://doi.org/10.1080/17549507.2025.2563845

International Journal of Speech-Language Pathology, 2025; Early Online: 1–17

Validity of acoustic speech measures obtained through videoconferencing with children with dysarthria

KYUNGHAE HWANG[1], FRITS VAN BRENK[2], MEGAN J. MCAULIFFE[3], JIYOUNG CHOI[1], JAN G. ŠVEC[4], YOUNG HWA M. CHANG[5], BRYAN KELLER[1] & ERIKA S. LEVY[1]

[1]*Teachers College, Columbia University, New York, NY,* [2]*University at Buffalo, Buffalo, NY,* [3]*University of Canterbury, Christchurch, New Zealand,* [4]*Palacky University Olomouc, Olomouc, Czech Republic, and* [5]*Rutgers the State University of New Jersey, Newark, NJ*

Abstract

Purpose: Children with dysarthria due to cerebral palsy often face barriers to receiving speech-language pathology services. Using online videoconferencing from home could be an appropriate solution if audio-recordings from such technology yield valid measures of the children's speech. This study assessed the validity of acoustic measures obtained from online recordings of children with dysarthria from their homes.

Method: Speech of 17 children with dysarthria was recorded from their homes simultaneously via two methods: 1) Online via Zoom and 2) offline via an audio-recording device. Nine commonly-assessed acoustic measures were obtained by each method and compared. Correlations and agreements between measures extracted from online and audio-device recordings were evaluated for whether they met predetermined criteria for validity.

Result: Second-formant range of diphthongs, fricative-affricate duration difference, word duration/articulation rate, mean fundamental frequency, and sound-pressure-level range met the criteria for validity. In contrast, fundamental frequency range, signal-to-noise ratio, and cepstral peak prominence did not meet validity criteria.

Conclusion: Findings support the validity of most commonly-analysed acoustic measures extracted from online recordings of children with dysarthria, suggesting that commercially-available videoconferencing technology could be an alternative to in-person evaluation. However, for perturbation- and noise-based measures, in-person recordings may still be necessary.

History: Received 13 November 2024; Accepted 12 September 2025

Keywords: *cerebral palsy; dysarthria; acoustics; videoconferencing; validity; speech*

Introduction

Dysarthria in children with cerebral palsy (CP) may have profoundly negative effects on the children's speech intelligibility, communicative participation, and quality of life (Mei et al., 2020; Pennington et al., 2018). Given that CP is the most frequently-occurring neuromotor disorder in children and that dysarthria is prevalent in children with CP (Mei et al., 2020), clinical services and research aimed at improving the children's speech communication are a high priority (Lee et al., 2014; Levy et al., 2017, 2021). However, families commonly experience barriers to accessing clinical services and research studies (Pennington et al., 2019). For example, mobility disabilities present in most children with CP may make travel challenging, as may families' geographic distance, time, and finances. Consequently, online videoconferencing has been increasingly relied upon in clinical practice and research as a practical solution to such challenges and disparities in access (Sevitz et al., 2021; Utianski et al., 2019). Online

Correspondence: Kyunghae Hwang, Department of Communicative Sciences and Disorders, New York University, 665 Broadway, New York, NY 10012. Email: kh3944@nyu.edu

Supplemental data for this article can be accessed online at https://doi.org/10.1080/17549507.2025.2563845.

ISSN 1754-9507 print/ISSN 1754-9515 online
Published by Taylor & Francis
DOI: 10.1080/17549507.2025.2563845

videoconferencing does, however, have some limitations in its use for assessment, treatment, and research involving children with dysarthria. Most notably, online speech recordings (henceforth, 'online recordings') are relied upon heavily for diagnostic purposes and for monitoring of treatment-related changes. However, the validity of such recordings in an everyday home context, with limited experimental control, is not clear.

Acoustic measures extracted from speech data are valuable for both clinical practice and speech research. In clinical practice, acoustic measures provide quantitative information to be considered for speech evaluations and treatment progress tracking; complementing more subjective perceptual assessments and reducing the possibility of diagnostic error (Sevitz et al., 2021). For research with children with dysarthria, acoustic measures provide objective information on speech characteristics (Lee et al., 2014), differentiate these characteristics from those of typically-developing children (Allison & Hustad, 2018), and can be used to assess the effects of speech strategies and treatments (Fox & Boliek, 2012; Levy et al., 2017, 2021; Pennington et al., 2018). Therefore, the importance of supplementing the subjective (and variable) perceptual measures collected from speech-language pathologists (SLPs) or researchers with objective acoustic measures has consistently been emphasised in the field (Ansel & Kent, 1992; Sevitz et al., 2021).

The quality of the speech signal may, however, be reduced in online recordings, in which case the acoustic measures extracted from such recordings might also be compromised. For example, uncontrolled home environments often introduce ambient noise, which may negatively affect the data. Moreover, the particular equipment used by the clients and participants, the signal processing techniques of the various videoconferencing platforms, and network issues such as bandwidth reductions can also impact the quality of online recordings. Additional factors such as packet loss leading to missing audio segments, as well as latency resulting in delayed audio transmission, may also compromise online recording quality. Such factors are inherent to online recording, with some being difficult, if not impossible, to control (Sevitz et al., 2021; Utianski et al., 2019). It is therefore important to assess the validity of acoustic measures obtained through online recordings collected using commonly-implemented methodologies, with their inherent limitations, if such measures are to be relied upon in clinical practice or research.

The impact of ambient noise, signal processing, internet speed, and network issues on online recordings may vary depending on the specific acoustic property being analysed. For example, perturbation- and noise-based acoustic measures, such as harmonic-to-noise-ratio (HNR) and cepstral peak prominence (CPP), may be relatively sensitive to noise and signal processing, whereas fundamental frequency (F0), reflecting the rate of vocal fold vibration, is often more robust (Weerathunge et al., 2021; Yoho & Borrie, 2018). Therefore, determining which measures extracted from online recordings are appropriate to include in clinical practice or research studies requires evaluation of a range of acoustic measures for the extent to which they may have been compromised by such factors. Of particular interest are acoustic measures that are commonly examined in this population, as these provide relevant information about the children's speech characteristics and intelligibility for the purposes of diagnosing and characterising dysarthria, as well as for treatment planning and tracking (Allison & Hustad, 2018).

Table I lists several acoustic measures frequently examined in the speech of English-speaking children with dysarthria and the information these provide on the children's speech characteristics and intelligibility. These include parameters that reflect articulatory and prosodic aspects of speech production, such as vocal tract shape and mobility and respiratory and phonatory function (i.e. range and slope of second formant [F2] of diphthongs, spectral differences between fricatives and affricates, articulation rate, fundamental frequency [F0], sound pressure level [SPL] range, and F0 range), as well as measures involving voice perturbation and speech noise (i.e. signal-to-noise ratio [SNR], and CPP; Allison & Hustad, 2018; Ansel & Kent, 1992; Kuschmann & van Brenk, 2019; Lee et al., 2014; Patel et al., 2018).

Given the limited research on the validity of acoustic measures collected from online recordings of children with speech disorders, studies examining the validity of online recordings of adults with speech disorders may provide insight on the topic. Constantinescu et al. (2010) assessed the validity of a custom-built computer-based videoconferencing system for online evaluation of speech disorders in adults with Parkinson's disease. The online and in-person recordings were collected simultaneously during clinical assessments of the adults' speech. A comparison of acoustic measures extracted from the online versus in-person recordings revealed that all measures tested (i.e. SPL, phonation time, and pitch range) were consistent across the recordings, providing support for the implementation of online speech assessment using their custom-built videoconferencing system.

With the rise of telehealth, recent studies have assessed the validity of online recordings obtained from publicly available platforms that can be used for telepractice while adhering to the privacy and data security standards set by the Health Insurance Portability and Accountability Act of 1996 (Weerathunge et al., 2021). For example, Weerathunge et al. (2021) investigated acoustic measures (mean F0, F0 variation, SPL variation, SNR, low-high [L/H] ratio, and smooth cepstral peak prominence [CPPS]) obtained from online

Table I. Commonly -examined acoustic measures, dysarthria characteristics, stimuli, and analysis.

Acoustic measure	Associated dysarthria characteristics	Stimuli	Acoustic analysis
Second formant (F2) range of diphthongs	F2 range of diphthongs is greater in typically-developing (TD) children than in children with dysarthria (Allison & Hustad, 2018).	Words: doubt, shape Sentence: Find all the crayons.	F2 trajectories for the vocalic segments of diphthongs were generated in Praat (Boersma & Weenink, 2020). F2 range was calculated as the difference between the maximum and minimum F2 frequencies within each vocalic segment containing a diphthong (Allison & Hustad, 2018).
F2 slope of diphthongs	F2 slopes of diphthongs are shallower in children with dysarthria than in TD children and predict speech intelligibility in children with dysarthria (Lee et al., 2014).	Words: doubt, shape Sentence: Find all the crayons.	Onset and offset of major transition were defined by the 20/20 rule (Weismer & Berry, 2003). This marked the boundaries for extraction of the transition duration and the F2 change across that time interval. All F2 slopes were average slopes for the entire transition (Lee et al., 2014).
Fricative-affricate duration difference	Reduced fricative-affricate contrasts (i.e., smaller noise duration differences) are associated with decreased intelligibility in adults with dysarthria due to cerebral palsy (Ansel & Kent, 1992).	Words: ship-chip, shock-chalk Sentences: Don't splash any water. Break the chocolate bar.	Noise portions of fricatives and affricates were determined by means of visual inspection of spectrogram and waveform. To yield a fricative-affricate duration difference score, the noise portion of affricate was subtracted from the noise portion of fricative.
Word duration/ Articulation rate	Slower articulation rate differentiates children with dysarthria from TD children (Allison & Hustad, 2018).	All words All sentences	For words, duration was extracted by custom Praat scripts. Sentence articulation rate was measured as syllables per second excluding silent intervals greater than 200 ms (Allison & Hustad, 2018; Levy et al., 2017).
Mean fundamental frequency (F0)	Mean F0 is a significant predictor of intelligibility in children with dysarthria. Higher F0 is observed in children with CP with dysarthria than in children with CP without dysarthria (Lee et al., 2014).	Words: pot, dot Sentences: Break the chocolate bar.	Mean F0 of words and sentences (from onset to offset) was extracted by a custom Praat script (Lee et al., 2014).
Sound Pressure Level (SPL) Range	SPL range is larger in children with dysarthria than in TD children, particularly in older children (Kuschmann & van Brenk, 2019).	All words All sentences	SPL range of words and sentences (from onset to offset) was extracted by custom Praat script (Kuschmann & van Brenk, 2019).
F0 Range	F0 range is smaller in children with dysarthria than in TD children (Kuschmann & van Brenk, 2019).	Words: pot, dot Sentences: Break the chocolate bar.	F0 range of words and sentences (from onset to offset) was extracted by a custom Praat script (Kuschmann & van Brenk, 2019).
Signal-to-noise ratio (SNR)	SNR is lower in children with dysarthria than in TD children (Lee et al., 2014).	Words: pot, dot Sentences: Break the chocolate bar.	Signal-to-noise ratio of words and sentences (from onset to offset) was extracted by custom Praat script (Lee et al., 2014).
Cepstral peak prominence (CPP)	Lower CPP values are found in children with dysarthria than in TD children (Kuschmann & van Brenk, 2019; Nip & Garellek, 2021).	All words All sentences	Cepstral peak prominence of words and sentences (from onset to offset) was extracted by custom Praat script (Nip & Garellek, 2021).

recordings of adults with voice disorders collected via Zoom (Zoom Video Communications), Microsoft Teams (Microsoft Corporation), and CiscoWebEx (Cisco Systems). Effects of transmission conditions were found on all videoconferencing platforms for all acoustic measures examined except the mean F0, suggesting that most acoustic measures were altered by the platforms. The authors concluded that standard time-based measures such as mean F0 may be less affected by online transmission than perturbation- and noise-based measures. The contrasting results of Weerathunge et al. (2021) and those of Constantinescu et al. (2010), who observed comparability between online and in-person acoustic measures, suggest that the validity of acoustic measures may vary depending on factors such as the videoconferencing platforms employed, the populations studied, the specific acoustic measures examined, and the statistical methods applied. These factors require further exploration.

To our knowledge, the only investigation aimed to validate online recordings of children's speech is Waite et al. (2006) pilot study of perceptual measures obtained from recordings of six children with speech disorders. The measures were scored on a word articulation test, ratings of speech intelligibility, and judgments of oral-motor impairment, as assessed by two SLPs. High levels of agreement were found between measures based on the online versus in-person recordings. No acoustic measures were reported. The small number of children with unspecified speech disorders and the use of custom-built equipment in the study limit the generalisability of these findings.

As for research into online methods of collecting speech data from children with dysarthria, the Pennington et al. (2019) pilot randomised controlled trial investigated family perceptions of using Skype for speech treatment (versus in-person treatment) in 22 children with dysarthria. The families of the children who received speech treatment via Skype found the online speech treatment delivery to be acceptable and, according to a parent, even to promote the child's independence. The researchers concluded that online speech treatment is feasible. However, the validity of acoustic measures extracted from online recordings of children with dysarthria remains to be established.

The present study aimed to determine whether videoconferencing from the homes of children with dysarthria due to CP, using commonly-implemented methodologies, yields valid acoustic speech measures. To address this question, we evaluated the concurrent validity[1] of acoustic measures obtained from online recordings of children with dysarthria. Audio recordings were collected simultaneously online and via an offline audio-recording device in the children's homes. A relatively comprehensive set of commonly-examined acoustic measures (described in Table I) were extracted from the online recordings. These measures were compared with those extracted from the audio-device recordings to assess whether the measures were comparable, despite the inherent differences in recording methods. More specifically, correlations and agreements between the acoustic measures extracted from the two recording modalities were evaluated for whether they met predetermined criteria for validity, as described in the Method section.

It was hypothesised that the F2 range of diphthongs, F2 slope of diphthongs, mean F0, and word duration/articulation rate extracted from online recordings would meet the validity criteria, as found in the studies by Constantinescu et al. (2010), Freeman and Decker (2021), and Weerathunge et al. (2021). However, SPL range and F0 range obtained from online recordings were not expected to meet the validity criteria because of the impact of transmission of the online videoconferencing platforms, as found in Weerathunge et al. (2021). Furthermore, perturbation- and noise-based acoustic measures (i.e. SNR and CPP) were not expected to meet the validity criteria, as such measures are sensitive to ambient noise and may be impacted by the signal processing that occurs during transmission via online video conferencing platforms (Yoho & Borrie, 2018).

Method

This study was approved by the Institutional Review Board at Teachers College, Columbia University, New York City.

Participants

A total of 17 children with CP (as diagnosed by a neurologist) and dysarthria participated in the study. These children included four females and 13 males, ranging in age from 5;1 to 16;1 (years; months), with a mean age of 10;10. They were recruited through word of mouth, the internet, and a university CP database. To determine eligibility for participation, the parents first completed an online questionnaire regarding their child's language background, ability to follow directions and repeat sentences, and primary mode of communication. Next, two experienced research SLPs met with the child and a parent for a Zoom video screening session. During the screening, the child participated in sentence and word-repetition tasks, conversation, diadochokinesis tasks, and an oral peripheral exam.

From the online questionnaire and Zoom screening session, the following inclusion criteria were confirmed for participation in the study: a) Age between 4 – 17 years, b) presence of dysarthria, c) reduced speech intelligibility, d) no other speech disorders such as childhood apraxia of speech, e) English as the dominant language, f) speech as the primary mode of communication, g) ability to follow simple directions, h) ability to repeat sentences of at least five words in length, and i) parent report of child's hearing within normal limits.

The presence and severity of dysarthria were determined for each child by two SLPs after reviewing each child's online questionnaire and meeting with them for a Zoom video screening session. The SLPs were experts in dysarthria, who had more than 10 years of clinical experience with differential diagnosis of motor speech disorders. When necessary, the SLPs referred to the recording of the screening session for additional confirmation. The two SLPs independently determined the presence of dysarthria following standard clinical procedures, including history and deviant perceptual speech characteristics (e.g. imprecise articulation, atypical voice quality, slow rate of speech, and hypernasality or hyponasality) and oral mechanism exam (Allison & Hustad, 2018; Levy et al., 2021). Childhood apraxia of speech, pure phonological disorder, and other diagnoses were ruled out by the SLPs with reference to ASHA's consensus criteria and the Mayo clinic checklist (Shriberg et al., 2012), based on the children's speech characteristics and patterns on tasks involving repetition of monosyllabic and multisyllabic words (Murray et al., 2015; Shriberg et al., 2012). There was 100% agreement between the two SLPs regarding the diagnosis of dysarthria. The severity of dysarthria was also determined by both SLPs' independent assessment of the children's speech intelligibility (Kim et al., 2011). Any differences in severity judgement were reconciled by consensus (for only two of the children, there was disagreement between mild-moderate and moderate severity, which was reconciled). The participants received a $50 Amazon gift card for their participation in the study.

Table II details participant characteristics, including dysarthria severity, language comprehension, and Gross Motor Function Classification System (GMFCS; Palisano et al., 1997) levels. Language comprehension was measured based on performance on specific subtests from either the Clinical Evaluation of Language Fundamentals–Fifth Edition (CELF-5; Wiig et al., 2013) or the Test for Auditory Comprehension of Language–Third Edition (TACL-3; Carrow-Woolfolk, 1999), depending on the child's motor and visual skills and age. It should be noted that the adaptations recommended in the test manuals were often not adequate for these children with frequent motor and visual limitations; thus, the scores are offered as a gross measure of the children's language comprehension, but probably underestimate their language abilities.

Experimental procedures

Speech stimuli

The stimuli for this study included twelve contrastive words in phrases (henceforth, 'words') and three sentences from the Test of Children's Speech (TOCS+; Hodge & Daniels, 2007; henceforth, 'sentences'). The words were: *Bat, beat, boot, chalk, chip, dot, doubt, pat, pot, shape, ship,* and *shock*. For ecological validity, the words were embedded in the carrier phrase "They say CVC again" rather than in isolation (Levy et al., 2017). The sentences were: *Find all the crayons, Don't splash any water*, and *Break the chocolate bar.* These sentences sampled a range of speech sounds of interest for the acoustic measures in this study. The speech stimuli were pre-recorded by an adult female speaker of American English as a first language in preparation for the children with dysarthria to repeat (see Levy et al., 2017, for details on the adult speaker's stimuli).

Experimental setup for speech data collection

Speech recordings were completed as part of a larger study. The recordings were conducted at the child's homes for two consecutive days, with the primary researcher (the first author) communicating via Zoom from her home. The child's speech was recorded using two different methods simultaneously: 1) Online via Zoom (Zoom Video Communications Inc, 2021) and 2) offline via an audio-recording device (sent to the parents).[2]

Before the recording sessions, the researcher had sent equipment for audio-device recording and calibration to the parents, with written instructions on setting it up, as well as a prepaid label for returning it. The equipment included: a) A portable audio recorder (H4n Handy recorder), bb) a condenser lavalier microphone with a flat frequency response (Countryman EMW Omnidirectional Lavalier microphone), c) a sound level metre (Galaxy CheckMate CM140), d) a headband, e) a measuring tape, f) an extra 9 V battery for the sound level metre, and g) a surge protector. The researcher pre-set the audio recorder at a sampling rate of 44.1 kHz with 16-bit resolution on a mono channel (Levy et al., 2017, 2021, Utianski et al., 2019). These settings remained unchanged throughout the study.

The experimental setup was driven by the broad question of whether videoconferencing as it is commonly carried out (in remote clinical practice and online research) yields valid acoustic speech measures. Therefore, while recording methods vary across clinical and research sites, we selected frequently-implemented methods described below. For example, the online method involved the internal microphone of families' online devices, with the speech signal transmitted through the commonly-used Zoom platform (Campbell & Goldstein, 2022; Sevitz et al., 2021). Because audio settings selected for Zoom recordings vary across practices and research studies, we utilised settings (described below) that are often implemented and recommended for capturing voice recordings clinically and in research, as they permit assessment of voice and speech with minimal distortion (Weerathunge et al., 2021) and can be easily adjusted by families, SLPs, and researchers. Similarly, replicating current dysarthria research study methods, the offline recording method entailed connecting a recording device to an external microphone (e.g. Levy et al., 2017, 2021; Moya-Galé et al., 2023). As detailed in the Discussion, while the methodological differences reduced experimental control of the speech signal, the differences between the signal in the online versus offline modalities were intrinsic to the research question and were assessed in our analysis.

Setup for online recording (Zoom)

At the beginning of the recording sessions, a parent connected to Zoom on their own laptop or iPad and

Table II. Participant characteristics of children with dysarthria due to cerebral palsy (CP).

Child	Age	Sex	Type of CP	GMFCS[a]	Dysarthria severity	Language comprehension
CP01	5.;1	M	spastic	IV	mild	63rd[b]
CP02	5;2	M	spastic	IV	moderate	<1st[b]
CP03	6;6	F	spastic	IV	moderate	9th[b]
CP04	7;4	F	mixed[d]	III	mild	<1st[b]
CP05	9;3	M	spastic	V	severe	50th[b]
CP06	9;7	M	mixed[e]	II	mild	37th[b]
CP07	10;9	M	spastic	II	mild	25th[b]
CP08	11;1	M	mixed[d]	IV	moderate-severe	25th[c]
CP09	11;7	M	ataxic	III	mild-moderate	<1st[c]
CP10	12;0	M	mixed[e]	V	severe	2nd[c]
CP11	12;5	F	spastic	IV	severe	37th[c]
CP12	12;8	M	spastic	V	moderate	37th[c]
CP13	13;2	F	spastic	IV	moderate-severe	<1st[c]
CP14	14;1	M	ataxic	II	mild-moderate	<1st[c]
CP15	14;7	M	spastic	II	mild	<1st[c]
CP16	15;7	M	mixed[d]	V	mild	9th[c]
CP17	16;1	M	spastic	I	mild	5th[c]

Note. F = female; M = male; GMFCS = Gross Motor Function Classification System (Palisano et al., 1997); WNL = within normal limits.
[a]GMFCS rating: I = no/mild impairment, V = severe impairment.
[b]Percentile rank obtained from the Test for Auditory Comprehension of Language–Third Edition, Elaborated Phrases and Sentences subtest.
[c]Percentile rank comprehension obtained from the Clinical Evaluation of Language Fundamentals–Fifth Edition, Word Classes subtest..
[d]Mixed spastic-ataxic, [e]Mixed spastic-dyskinetic.

was guided remotely by the experimenter via Zoom. The Zoom platform was selected as this is the most widely utilised online videoconferencing platform by SLPs (Campbell & Goldstein, 2022). Of the 17 participants, 14 used laptops and three used iPads. The child was seated 30 cm from the microphone of the laptop or iPad,[3] as measured by the parent and confirmed by the experimenter via Zoom (Theodoros et al., 2016). The iPads used by the participants were confirmed to have their microphones at the top, and were set up to stand upright on the table, allowing the cameras and screens to face the children directly (Sevitz et al., 2021). No external microphone was used for the online recording because telepractice typically employs the internal microphone of a family's online device (Sevitz et al., 2021) and as children with CP often do not tolerate external microphones well. Both the parent and the experimenter selected the 'original sound' function to disable Zoom's artificial noise suppression, high-pass filtering, and automatic gain control. The experimenter's and participant's microphone recording levels were set at 100% on their Zoom audio settings (Weerathunge et al., 2021).

Setup for offline recording (audio recorder)

Next the experimenter guided the parent via Zoom to set up the audio-device recording. She instructed them to connect the lavalier microphone to the audio recorder. To ensure that the recording of the adult speech stimuli for repetition by the children would be played at consistent SPLs across all of the children, each recording session began with a calibration process. The experimenter guided the parents to use their laptop or iPad on Zoom to access an online tone generator website (https://www.szynalski.com/tone-generator/). They were then instructed to place the lavalier microphone and sound level metre next to each other, 8 cm away from the midpoint of the loudspeaker of their device. The sound level metre was set to default values (i.e. A-frequency weighting, slow-time weighting). Next, the parents played a pure tone from the online tone generator website at the default setting (440 Hz and 75% recording level). They were instructed to adjust the recording level of their device so that the SPL of the steady tone at 8 cm would measure 75 dB SPL on the sound level metre (Švec & Granqvist, 2018).

Subsequently, the experimenter instructed the parents to secure the microphone to the child's forehead with a headband (Fox & Boliek, 2012), with a mouth-to-microphone distance of 8 cm (Levy et al., 2017, 2021). This H4n Handy audio recorder and lavalier microphone setup represents commonly implemented setups in recent dysarthria literature, particularly in studies that consider SPL in their intelligibility assessment (e.g. Levy et al., 2017, 2021; Moya-Galé et al., 2023). Because many children with CP have difficulty controlling their movements and wearing headset microphone (see Levy, 2014) a lavalier microphone taped to the forehead is common in childhood dysarthria research (Fox & Boliek, 2012; Levy, 2014; Levy et al., 2017, 2021) and was employed in this study.

The parents measured the mouth-to-microphone distance with the measuring tape, and the distance was confirmed by the experimenter via Zoom. Figure 1 shows the experiment setup, with the child seated 30 cm from the laptop microphone, wearing a lavalier microphone (8 cm from his mouth) connected to an audio recorder.

Speech data collection

Once the online and audio device recording setup and calibration were completed, the child participated in the speech production task, as guided by the

Figure 1. Experiment setup of the study.

experimenter, with assistance from the parent. Throughout the recording session, the child's mouth-to-microphone distance, as well as the distance between the child and the microphone of the laptop or iPad, to the extent possible, was kept constant (Fox & Boliek, 2012; Levy et al., 2017, 2021; Theodoros et al., 2016). For the recording, the child was seated in a chair at home, in a quiet room. To elicit the child's speech, the experimenter played the adult speaker's utterances through Zoom and instructed the child to repeat the utterances. The child's speech was recorded simultaneously using Zoom for the online recording and the portable audio recorder for the audio-device recording. If the child failed to repeat the utterances, the experimenter provided them with reminders. All responses that were perceived to be attempts at repeating the utterance were accepted, except when productions were incomplete, extraneous noise occurred or the child was off-task, in which case the experimenter instructed them to repeat the production. Models and encouragement were provided if the child demonstrated disinterest or fatigue and breaks were offered as needed.

Following the final recording session, the parent returned the equipment, with the audio files on the audio recorder, by mail. The researcher transferred the audio file data from the secure digital (SD) card of the audio recorder onto a computer for analysis.

Data analysis

Because the recording feature of Zoom online video-conferencing produces audio files as MP4 files, the MP4 recordings were converted to WAV format by means of iTunes to be compatible with Praat (Boersma & Weenink, 2020). Nine acoustic measurements (F2 range of diphthongs, F2 slope of diphthongs, fricative-affricate duration difference, word duration/articulation rate, mean F0, SPL range, F0 range, SNR, and CPP) were taken from 1,622 word-level tokens and 571 sentence-level tokens, totalling 2,193 measurements from the speech of each of the 17 children or altogether 37,281 measurements. Acoustic analysis was performed by the primary researcher, one doctoral student, and four research assistants, following a detailed protocol and training led by the primary researcher. The onset and offset of speech sounds of interest were determined based on standard acoustic criteria (Levy et al., 2017). The F2 range of diphthongs, F2 slope of diphthongs, fricative-affricate duration difference, and articulation rate were manually obtained through inspection of a combination of the waveform and wideband spectrogram in Praat (Boersma & Weenink, 2020). Mean F0, word duration, SPL range, F0 range, SNR, and CPP were obtained with a Praat script. Each researcher conducted acoustic analysis independently and was blinded to the other researchers' work. Table I displays the details on the stimuli and how each acoustic measure was analysed.

In order to ensure the reliability of the manually obtained acoustic findings from both online and audio-device recordings, a random selection of 20% of words and sentences were remeasured in two ways: 1) By the same researcher without reference to their initial measurement (researcher intrajudge reliability) and 2) by a second research assistant without reference to the first researcher's measurement (researcher interjudge reliability). A Pearson product-moment correlation was used to measure the intrajudge and interjudge reliability. The researcher intrajudge reliability demonstrated strong correlations for all acoustic measures at both the word and sentence level (words: $r = 0.93 - 0.98$, sentences: $r = 0.93 - 0.97$). Furthermore, a strong correlation was found for researcher interjudge reliability across all acoustic measures at both word and sentence levels (words: $r = 0.92 - 0.96$, sentences: $r = 0.91 - 0.95$). Overall, the results indicate strong reliability of the acoustic findings. Intrajudge and interjudge reliability measures were not calculated for the acoustic measures obtained using a Praat script.

Statistical analysis

Concurrent validity was determined by comparing measures extracted from the online recordings to measures extracted from the audio-device recordings. Two approaches were implemented: Repeated measures correlations and mixed effects Bland and Altman (1986) Limits of Agreement (LoA). The validity of the acoustic measures was assessed separately at both word and sentence levels. These measures were considered valid if they met the specific validity criteria set a priori for both approaches, as follows:

(1) *Repeated measures correlations* assess the linear relationships between two continuous variables when there are multiple observations on each child. To minimise the risk of Type I errors associated with

multiple comparisons, the Benjamini and Hochberg (1995) correction method was applied. A correlation cut-off value of $r_{rm} \geq 0.7$ between the two data sets was set a priori for determining whether the measures extracted from the online recording were considered valid (Curtis et al., 2022).

(2) The *Bland-Altman LoA* method was included as a complementary approach to address a limitation of correlation measures in that the LoA method takes into account the magnitude of the difference between the two variables. A mixed effects LoA analysis was employed to account for the repeated measurements taken from the same child (Bland & Altman, 1986). The LoA method estimates an interval in which 95% of the differences between two measures are expected to fall. Maximum acceptable differences are established a priori based on clinically or analytically relevant criteria. If the LoA are within these maximum acceptable differences, the two measures are considered to be in agreement and interchangeable.

To our knowledge, maximum acceptable differences have not been previously reported for the measures investigated in this study for children with dysarthria, although they have been for studies of adults with dysarthria (Constantinescu et al., 2010). The standard error of measurement (SEM) is frequently used to set the maximum acceptable differences (Constantinescu et al., 2010; Storm et al., 2020). The SEM is calculated in terms of the intraclass correlation coefficient (ICC) from test-retest variability at two-time points: SEM = SD$\sqrt{}$(1-ICC). Thus, for the present study, the SEM calculated based on test-retest data was considered a suitable criterion for maximum acceptable differences. Table III displays the maximum acceptable differences based on the SEM of the test-retest variability in the audio device data across the first and second days. To be considered valid, the LoA for each acoustic measure was required to be within the defined maximum acceptable difference (Constantinescu et al., 2010).

Repeated measures correlations were conducted for all acoustic measures at both the word and sentence levels, followed by the mixed effects Bland-Altman analysis. Normality of Bland–Altman difference scores was assessed using Shapiro–Wilk tests. For most variables, the assumption of normality was not met; thus, we applied non-parametric bootstrapping (1,000 resamples) to estimate 95% confidence intervals for mean bias. Method reporting was guided by Abu-Arafeh et al. (2016). A corresponding checklist is provided in Supplementary Table 1.

Result

Correlations between online and audio-device recordings

Word level

At word level, strong positive correlations were found (range: $r_{rm} = 0.76 - 0.95$, $p < 0.001$) between online and audio-device recordings for F2 range of diphthongs, F2 slope of diphthongs, fricative-affricate duration difference, word duration, mean F0, SPL range, and CPP. These correlations met the validity criterion of 0.7. A weak positive correlation that did not meet the validity criterion were found for F0 range ($r_{rm} = 0.35$, $p = 0.02$). Additionally, a weak negative correlation that did not meet the validity criterion was observed for SNR ($r_{rm} = -0.31$, $p = 0.04$). Outliers in F0 range are not uncommon in research on individuals with CP, as F0 range can vary widely in this population (Kim et al., 2011). After the outliers were identified and removed, the repeated measures correlations still indicated a moderate positive correlation ($r_{rm} = 0.48$, $p = 0.002$), not meeting the validity criterion for correlation. Repeated measures correlations between online and audio-device recordings for all acoustic measures at the word level are shown in Figure 2.

Sentence level

At the sentence level, strong positive correlations ($r_{rm} = 0.78 - 0.99$, $p < 0.001$) were revealed for F2 range of diphthongs, F2 slope of diphthongs, fricative-affricate duration difference, articulation rate, mean F0, and SNR, meeting the validity criterion of 0.7. However, while SPL range exhibited a moderate positive correlation (SPL range: $r_{rm} = 0.58$, $p < 0.001$), it did not meet the validity criterion. The weakest correlations between online and audio-device recordings, not meeting the validity criterion, were observed for F0 range and CPP ($r_{rm} = 0.27 - 0.3$). Figure 3 depicts the repeated measures correlations between online and audio-device recordings of all acoustic measures at the sentence level.

Limits of agreements between online and audio-device recordings

Word level

Figure 4 presents the mixed effects Bland-Altman LoA and the predetermined maximum acceptable differences for all acoustic measures at word level.

The LoA between online and audio-device recordings for F2 range of diphthongs, fricative-affricate duration difference, word duration, mean F0, and SPL range fell within the respective maximum acceptable differences. The upper LoA of CPP fell within the maximum acceptable difference; however, the lower LoA fell outside of the maximum acceptable difference. The LoA of F2 slope of diphthongs, F0 range, and SNR fell outside the respective upper and lower maximum acceptable differences, with F2 slope of diphthongs missing the criterion by only 0.23 Hz/ms.

Table III. Maximum acceptable differences for all acoustic measures.

	Maximum acceptable difference	
Acoustic measures	Words	Sentences
F2 range of diphthongs (Hz)	210.65	212.06
F2 slope of diphthongs (Hz/ms)	2.31	2.32
Fricative-affricate duration difference (ms)	77.88	107.43
Word duration (s) or Articulation rate (Syllable/s)	0.13	0.17
Mean F0 (Hz)	49.49	32.92
SPL range (dB)	11.38	9.03
F0 variation (Hz)	17.63	12.36
SNR (dB)	3.21	2
CPP (dB)	1.53	1.46

The maximum acceptable difference was determined by calculating the standard error of measurement (SEM) of the test-retest variability in the audio-device recording data across the first and second days of data collection.

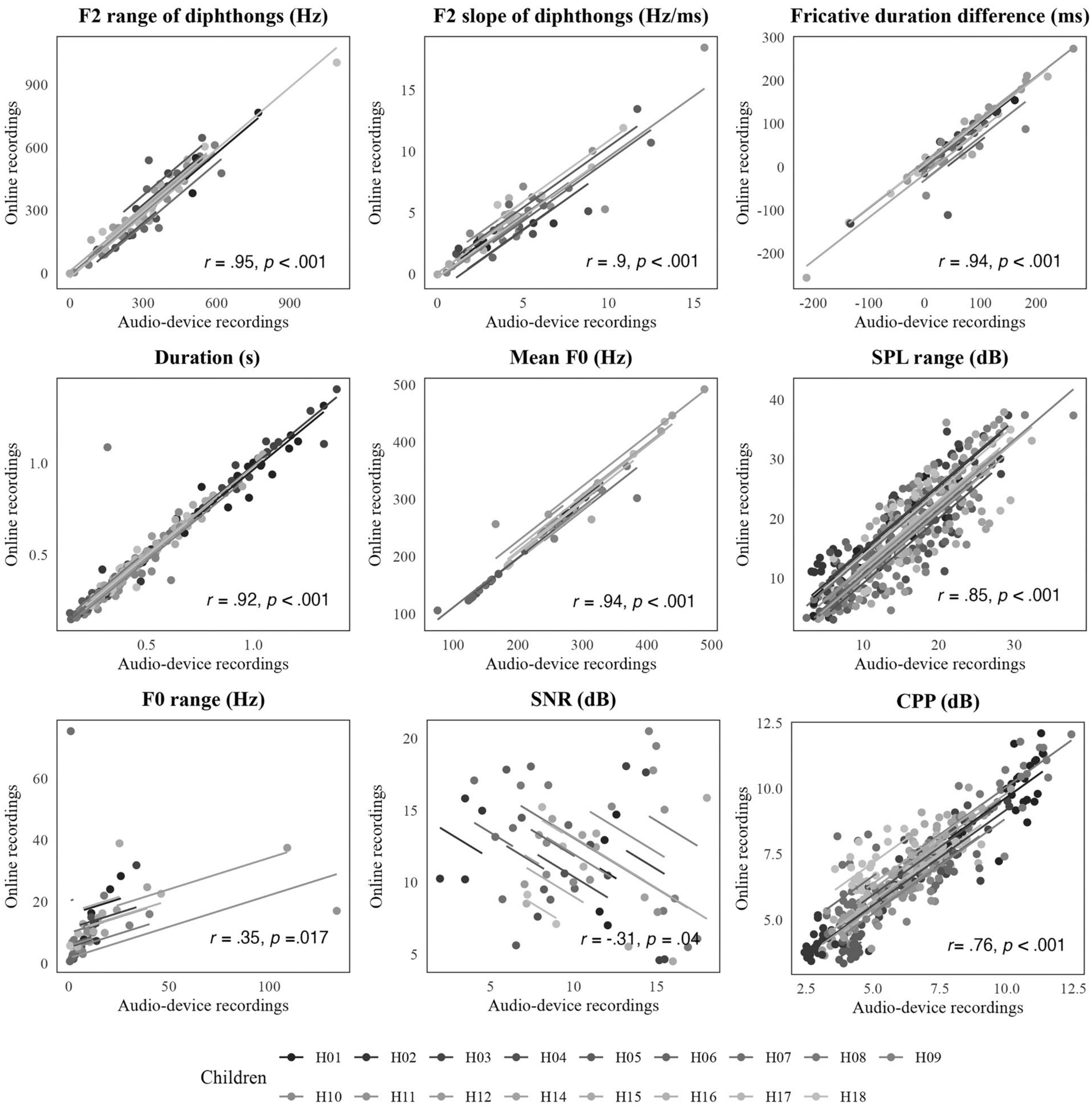

Figure 2. Repeated measures correlations between online and audio-device recordings at word level. F2 = Second formant, F0 = fundamental frequency, SPL = sound pressure level, SNR = signal-to-noise ratio, CPP = cepstral peak prominence. Observations from the same participant are depicted in the same gray scale, with corresponding lines indicating the rmcorr fit for each participant. The displayed p-values were adjusted using the Benjamini and Hochberg (1995) method.

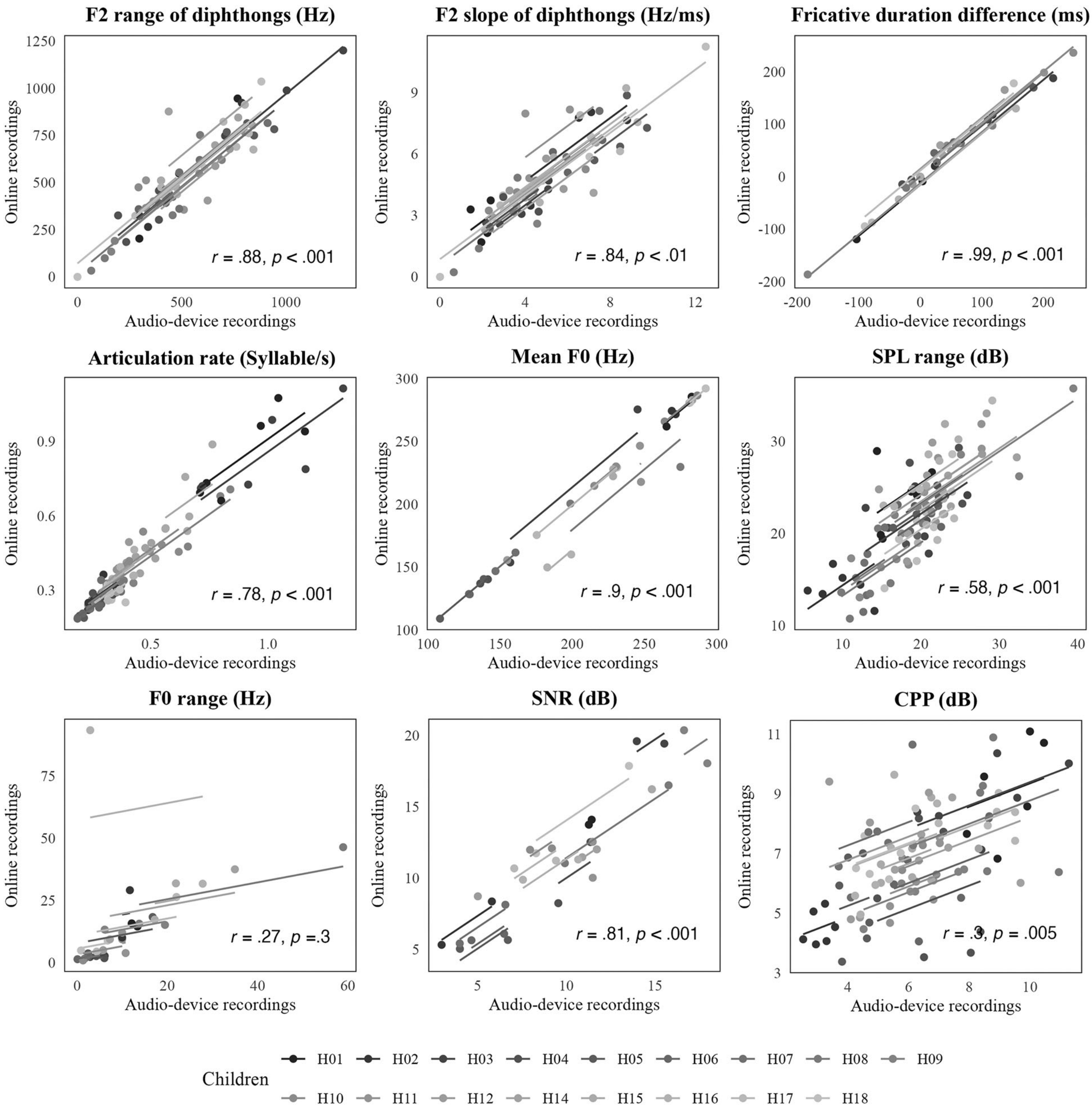

Figure 3. Repeated measures correlations between online and audio-device recordings at sentence level. F2 = Second formant, F0 = fundamental frequency, SPL = sound pressure level, SNR = signal-to-noise ratio, CPP = cepstral peak prominence. Observations from the same participant are depicted in the same gray scale, with corresponding lines indicating the rmcorr fit for each participant. The displayed p-values were adjusted using the Benjamini and Hochberg (1995) method.

Sentence level

The mixed effects Bland-Altman LoA of all acoustic measures at sentence level between online and audio-device recordings are shown in Figure 5.

At the sentence level, LoA's of F2 range of diphthongs, fricative-affricate duration difference, articulation rate, and mean F0 fell within the predetermined maximum acceptable differences. The lower LoA of F2 slope of diphthongs fell within the maximum acceptable difference, but the upper LoA did not, exceeding the criterion by .09 Hz/ms. The analyses for SPL range and SNR demonstrated a similar pattern, with the upper LoA falling within the maximum acceptable difference, but the lower LoA exceeding the maximum acceptable difference. The LoA of the F0 range and CPP did not fall within the predetermined maximum acceptable differences.

Table IV provides an overview of the correlations and agreements between the acoustic measures extracted from the online recordings and those from the audio-device recordings and indicates (in bold print) whether they met the two validity criteria. The F2 range of diphthongs, fricative-affricate duration difference, word duration/articulation rate, mean F0, and SPL range met validity criteria (for the correlations and LoA) at word and/or sentence level. However, F0 range did not meet either validity criterion at word or sentence level. While SNR met the validity criterion for the correlations only at the sentence level, CPP met the correlation criterion only at the word level.

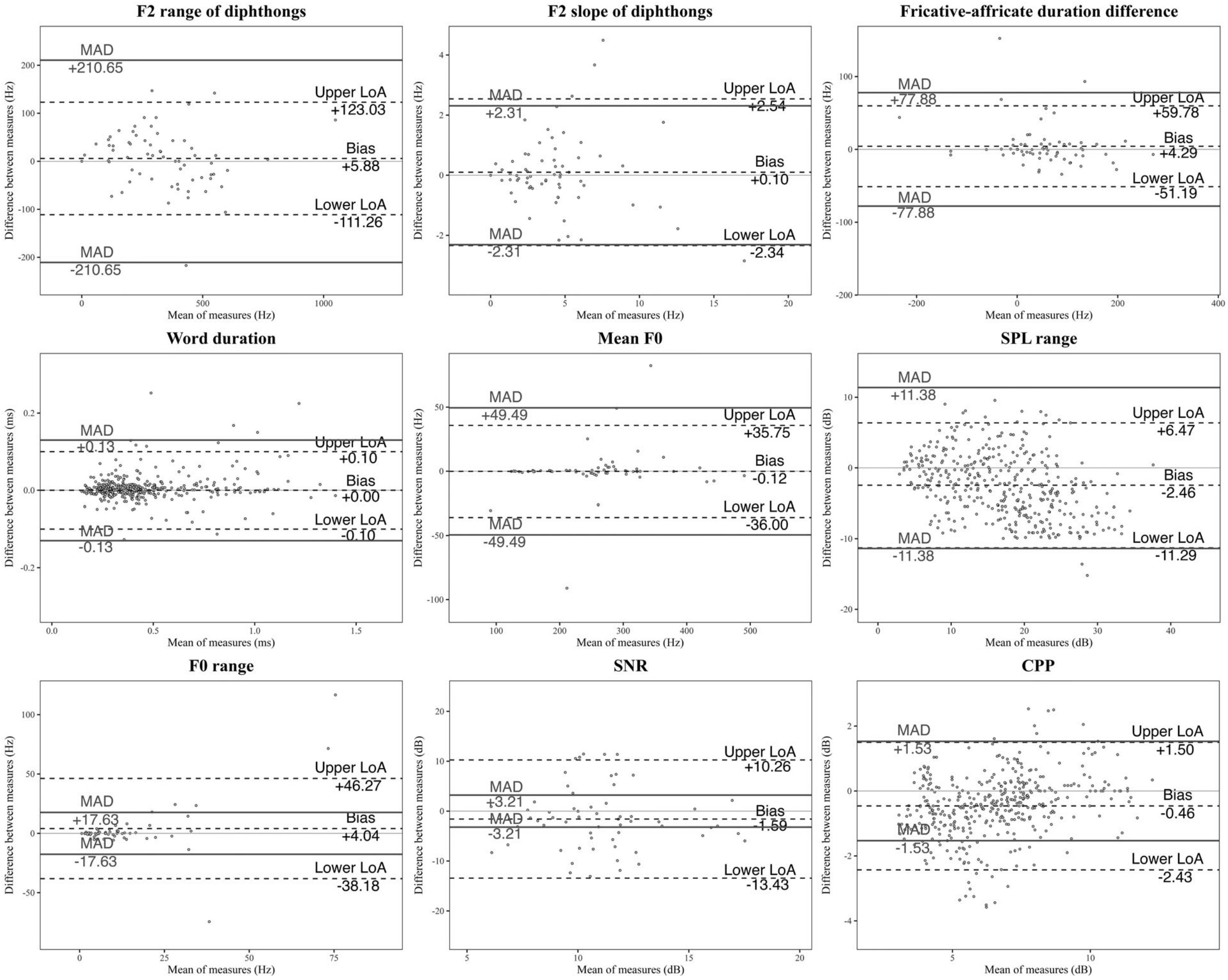

Figure 4. Mixed effects Bland-Altman plots of agreement between measures extracted from online and audio-device recordings at word level. F2 = Second formant, F0 = fundamental frequency, SPL = sound pressure level, SNR = signal-to-noise ratio, CPP = cepstral peak prominence. LoA and bias are shown as dotted gray lines with 95% confidence intervals. MAD is presented as solid gray lines.

Neither SNR nor CPP met the validity criterion for the LoA at word or at sentence level.

In the participant sample, three children (CP08, CP09, CP14) exhibited dyskinesias during the recording session. To examine whether their head and body movements may have impacted the findings—particularly given the potential effects on mouth-to-microphone distance in the virtual recordings—a sensitivity analysis was conducted excluding these participants. This exclusion resulted in minor changes to some correlation coefficients and limits of agreement. For example, a small reduction in correlation for sentence-level F0 range (from r = .93 to .92) and slightly reduced limits of agreement for CPP (from ±3.59 to ±3.50) and SNR (from ±12.11 to ±11.25) resulted at the word level. However, these changes did not alter the overall classification of validity; i.e. measures that met both criteria remained valid and those that did not remained below threshold.

To further evaluate agreement between recording methods, bootstrapped Bland–Altman bias estimates were calculated for each acoustic measure. These analyses revealed small to moderate mean differences across several measures. For example, fricative duration at the word level was lower in the online recordings, with a mean bias of −4.29 ms (95% CI [−11.58, 1.94]), whereas F2 range at the sentence level was higher in the online recordings, showing a mean bias of 9.18 Hz (95% CI [−15.9, 36.11]). Most confidence intervals included zero, indicating that systematic differences between modalities were generally small.

Discussion

This study evaluated the validity of acoustic speech measures of children with dysarthria due to CP collected online from their homes. Measures extracted from online recordings made via Zoom, employing commonly-implemented methods, were compared to those extracted from recordings made simultaneously by means of an audio-recording device in the children's homes. As discussed below, the majority of the tested measures extracted from the online recordings were highly correlated and exhibited good agreements with those obtained from the audio-device recordings, with only a few measures not meeting the criteria for validity. These findings support the use of the Zoom online platform for obtaining most commonly-assessed acoustic measures of the speech of children

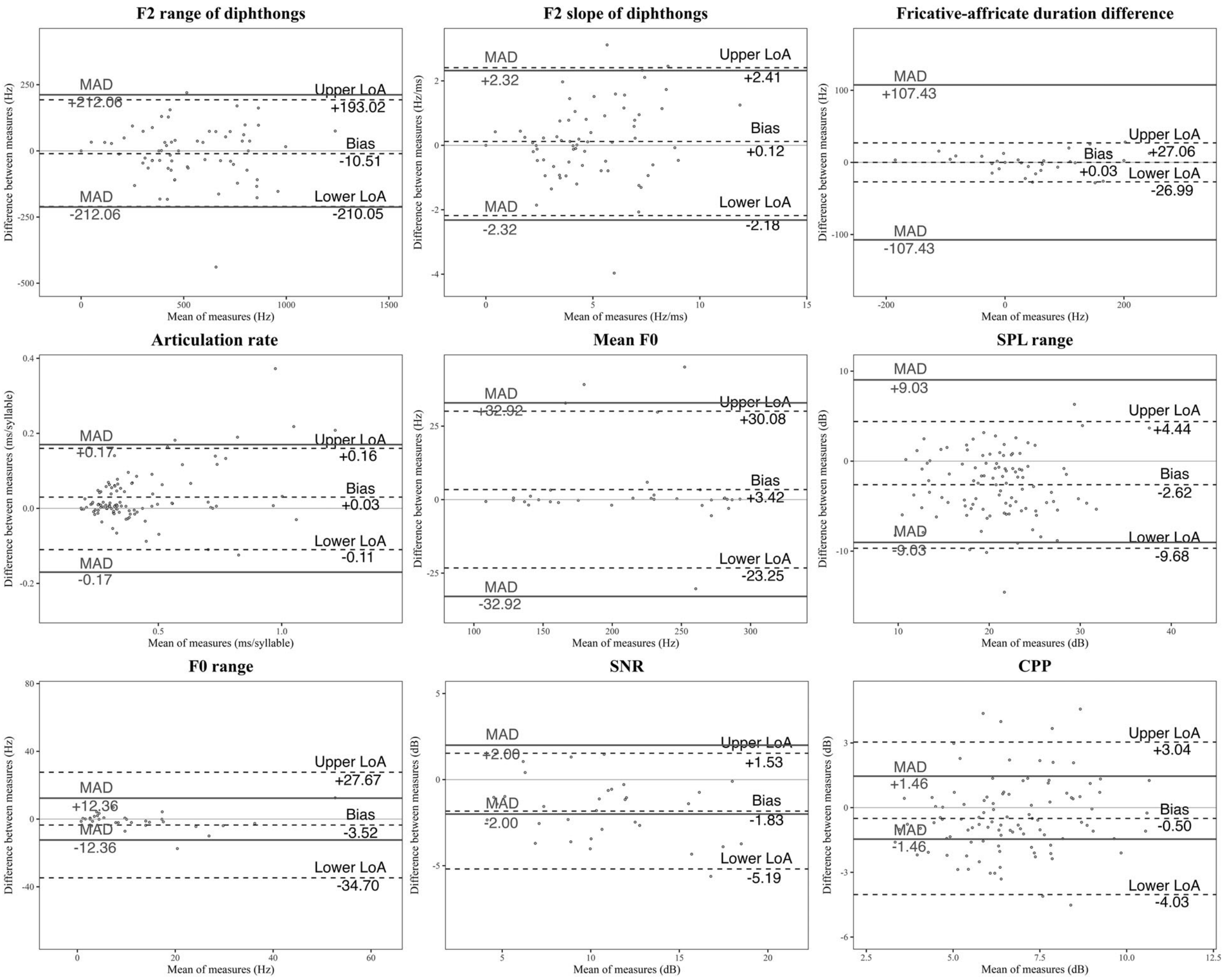

Figure 5. Mixed effects Bland-Altman plots of agreement between measures extracted from online and audio-device recordings at sentence level. F2 = Second formant, F0 = fundamental frequency, SPL = sound pressure level, SNR = signal-to-noise ratio, CPP = cepstral peak prominence. LoA and bias are shown as dotted gray lines with 95% confidence intervals. MAD is presented as solid gray lines.

Table IV. Summary of correlation, limits of agreement, and validity of acoustic measures.

	Correlation				Bland-Altman limits of agreement (LoA)	
	Words		Sentences		Words	Sentences
Acoustic measures	R_{rm}	*p*	R_{rm}	*p*	LoA	LoA
F2 range of diphthongs	**0.95**	**<0.001**	**0.88**	**<0.001**	**−111.26 – 123.03**	**−210.05 – 193.02**
F2 slope of diphthongs	**0.9**	**<0.001**	**0.84**	**<0.01**	−2.34 – 2.54	−2.18 – 2.41
Fricative-affricate duration difference	**0.94**	**<0.001**	**0.99**	**<0.001**	**−51.19 – 59.78**	**−26.99 – 27.06**
Word duration / articulation rate	**0.92**	**<0.001**	**0.78**	**<0.001**	**−0.1 – 0.1**	**−0.11 – 0.16**
Mean F0	**0.94**	**<0.001**	**0.9**	**<0.001**	**−36 – 35.75**	**−23.25 – 30.08**
Sound pressure level range	**0.85**	**<0.001**	0.58	<0.001	**−11.29 – 6.47**	−9.68 – 4.44
F0 range	0.35	0.02	0.27	0.3	−38.18 – 46.27	−34.7 – 27.67
Signal-to-noise ratio	−0.31	0.04	**0.81**	**<0.001**	−13.43 – 10.26	−5.19 – 1.53
Cepstral peak prominence	**0.76**	**<0.001**	0.3	0.005	−2.43 – 1.5	−4.03 – 3.04

Note. Bold print in the numbers indicates correlations that meet the validity criterion of $r_{rm} \geq 0.$ or LoAs within the maximum acceptable differences.

with dysarthria. The variable results, with their implications and limitations, are considered below.

Validity of articulatory and prosodic measures

Of the various acoustic measures investigated in this study, F2 range of diphthongs, fricative-affricate duration difference, word duration/articulation rate, and mean F0 revealed the strongest correlations between the recording methods, surpassing the validity criterion. Moreover, LoAs for these parameters fell well within the predetermined maximum acceptable differences. These results are consistent with our hypotheses and align with previous research generally supporting the validity of acoustic measures that reflect articulatory and prosodic aspects of adults' speech production obtained from online recordings, including Constantinescu et al.'s (2010) report of measures extracted from their custom-built videoconferencing system and Weerathunge et al. (2021) measures extracted from Zoom recordings.

The F2 range of diphthongs, reflecting articulatory transitions during diphthong production, exhibited a strong correlation between the online and audio-device recordings, and LoA within the maximum acceptable difference, consistent with Freeman and Decker (2021). In contrast, while F2 slope of diphthongs showed a strong correlation between the online and audio-device recordings, the LoA between the two recording modalities fell slightly outside the predetermined maximum acceptable difference. The differences in agreements between F2 range of diphthongs and F2 slope of diphthongs may be attributable to the relative complexity of the F2 slope of diphthongs in that it relies on both spectral and temporal characteristics. That is, cumulative errors in measurement of spectral and temporal parameters could impact the measures' validity. This finding underscores the importance of clinicians and researchers exercising caution when selecting acoustic measures to analyse from online recordings, as there may be relative differences in the measures' validity, despite reflecting similar aspects of speech production.

As hypothesised, the fricative-affricate duration difference and word duration/articulation rate obtained from online recordings met the established validity criteria, consistent with prior research showing LoA within predetermined acceptable limits for vowel duration measurements between in-person and online recordings of adults with dysarthria (Constantinescu et al., 2010). Our study further demonstrates that the consonant duration measure of fricative-affricate duration difference and more general duration measures (i.e. word duration and articulation rate) may also be appropriately obtained from online recordings of children with dysarthria. Thus, duration measures seem generally robust to the effects of online transmission.

Additionally, as anticipated, the mean F0 extracted from online recordings met our validity criteria, aligning with the findings of Constantinescu et al. (2010) and Weerathunge et al. (2021). In contrast, for the F0 range, our study showed weak correlations and agreements outside of the clinical criteria, consistent with Weerathunge et al.'s (2021) finding of the F0 range being significantly impacted by online transmission. Because F0 range would not be expected to be affected by internet transmission, this finding is likely related to the sensitivity of our study's F0 extraction algorithms to background noise (Titze & Liang, 1993), which differs in online versus audio-device recordings.

The validity of the SPL range obtained from online recordings depended on the linguistic unit measured. SPL range extracted from words in carrier phrases showed a strong correlation between online and audio-device recordings, reaching the validity criterion and exhibiting good agreement. Although SPL range extracted from sentences demonstrated a moderate correlation, this did not meet the validity criterion for correlations and the LoA fell outside the predetermined maximum acceptable difference. That SPL range did not meet validity criteria at the sentence level is consistent with Weerathunge et al. (2021) finding of SPL range in sentence-level (read) speech being significantly impacted by transmission.

Because SPL range can be affected by the varying mouth-to-microphone distance (Švec & Granqvist, 2018), a sensitivity analysis was conducted excluding the three participants who exhibited observable dyskinesia during recording. This exclusion did not change the overall validity classification, consistent with Weerathunge et al. (2021), who found that sentence-level SPL range was significantly impacted by transmission even when microphone-to-sound-source distance was fixed. These results indicate that factors beyond movement, such as ambient noise, online transmission, and platform-specific signal processing, may contribute to the reduced validity observed for longer utterances. Taken together, the findings suggest that clinicians and researchers may consider incorporating most articulatory and prosodic measures into their speech assessments of children with dysarthria performed online, whereas F0 range, and possibly SPL range, particularly for longer utterances, may be better obtained offline in clinical or laboratory settings.

Validity of perturbation- and noise-based measures

Considerable variability was observed in the correlations between online and audio-device recordings of the perturbation- and noise-based acoustic measures; namely SNR, and CPP, with none of the measures' LoA falling within the maximum acceptable differences. The strength of the correlations of SNR and CPP varied depending on whether the measurements were taken at the word or sentence level. For instance, the correlation between online and audio-device recordings for CPP was strong at the word level, meeting the validity criterion; however, it was weak at the sentence level and failed to meet the criterion. With considerable variation in the correlations between the two recording methods for these measures, only SNR at the sentence level and CPP at the word level met the validity criterion. Furthermore, both of these measures exhibited LoA that exceeded the maximum acceptable difference for validity. These findings are consistent with previous research that has highlighted the negative impact of online transmission on perturbation and noise-based measures (Weerathunge et al., 2021). The computation of such measures involves examining cycle-to-cycle variations, which rely on fine temporal details that make them vulnerable to the effects of ambient noise, signal processing, and packet loss and latency at times of reduced internet speed and bandwidth during signal transmission. Overall, these findings suggest that, when extracted from online recordings, these perturbation- and noise-based acoustic measures that are commonly used to examine voice quality of children with dysarthria may not be valid for clinical or research purposes.

Limitations and future directions

While this study supports the use of online technology for collecting speech data for assessing most commonly-evaluated acoustic measures, some limitations should be acknowledged. First, as previously mentioned, we used an online recording protocol with the experimenter providing guidance remotely via Zoom. This differs from standard offline audio recordings, which traditionally involve the experimenter and participant being co-located in a lab, clinic, or home. This setup increased the ecological validity of the online component in that videoconferencing in clinical practice and research is typically conducted with the child at home and the clinician elsewhere. However, it is possible that the parent following the experimenter's remote guidance did not set up the equipment with the precision of an experimenter.

Secondly, the study focused exclusively on Zoom because it is the videoconferencing platform most frequently used by SLPs today (Campbell & Goldstein, 2022). However, Weerathunge et al. (2021) found not only different transmission effects across various platforms, but also noted variations within the same platform, depending on whether audio enhancement features were enabled. Therefore, our study, in which online recordings were collected via Zoom with specific recommended audio settings and placement of online devices (Sevitz et al., 2021; Weerathunge et al., 2021), should be replicated implementing various other videoconferencing platforms and setups utilised for research and clinical purposes.

We emphasise that the study was not designed to address the technical question of whether a speech signal collected by a microphone from a child with CP with dysarthria would be of the same quality if it were transmitted via Zoom as if transmitted directly to an offline recording device. This question would have entailed an alternative setup, with the speech signal from a single microphone divided by a splitter into an online and offline recording system, for example. Instead, we addressed a question more pertinent to current clinical and research practices, of whether online recording methods typically used with this population would yield valid acoustic measures.

As a result, however, our setup introduced considerable variability and relinquished some control. For example, for the online recordings, the mouth-to-microphone distance was 30 cm and utilised an internal microphone (Sevitz et al., 2021; Theodoros et al., 2016). In contrast, in the offline recordings, the mouth-to-microphone distance was 8 cm to an external microphone (Levy et al., 2017, 2021; Moya-Galé et al., 2023). While designed to represent common online and offline recording practices, this setup introduced variability not only of equipment, but likely also of ambient noise. The distance differences could have contributed to noise-related measures and SPL range at sentence level not showing validity. Although the sensitivity analysis without the children with observable dyskinesia did not change the overall validity classification, it is possible that children with more severe dyskinesia, and therefore greater and more frequent changes in mouth-to-microphone distance, would exhibit larger discrepancies between online and offline measures. Potential implications, which cannot be determined from the current study design, may be that external microphones should be utilised, if feasible, when extracting noise-related measures (and SPL range of longer utterances) from online recordings.

Analysis of measurement agreement showed small but consistent differences between the two recording methods for some acoustic measures. For example, F2 slope of diphthongs tended to be slightly longer when recorded with the audio-recording device, while SPL ranges were somewhat higher in the online recordings. Although these directional trends were modest and did not affect whether the measures met validity criteria in this study, they highlight the importance of being cautious when interpreting data from online recordings, especially in contexts where small acoustic differences could influence clinical judgments. In addition, because the statistical assumption of normality required for Bland–Altman analysis was not met for most acoustic variables, non-parametric bootstrapping was used to estimate bias confidence intervals. Future studies should re-assess these variables with larger samples where statistical assumptions are satisfied.

Moreover, the small sample of families who participated in this study was not necessarily representative of the majority of families of children with dysarthria in that, for example, the parents had responded via email to recruitment flyers and therefore represent a subset of families who likely owned online devices and had facility in using them. Additionally, the families' devices included various laptops and iPads, further sources of variability. While our measures, for the most part, still met the validity criteria, a more controlled investigation comparing validity of measures extracted from various recording technologies, collected from a more heterogeneous pool of participants, could provide insight into optimal equipment and recording methodologies for clinical and research purposes. In summary, while many possible videoconferencing methods and technologies may yield various degrees of validity for acoustic speech measures frequently assessed, our study selected only one common videoconferencing platform and one common audio device setup representative of common practice. Therefore, future research should explore the validity of other common recording methodologies, which are likely to evolve rapidly in the coming years (Sevitz et al., 2021; Utianski et al., 2019).

For a new tool to be adopted widely, it needs to be not only reliable and valid, but also feasible to implement (Curtis et al., 2022). Thus, a final limitation in this study, and of telepractice in general, is the burden placed on families to own and operate their own online devices and manage their children's behaviour during recording sessions, as well as the equipment and internet issues, and other remote-learning related issues that often arise. While the burdens of time and travel necessary to attend sessions in person are relieved through telepractice, parents' involvement in the sessions is certainly increased and equipment and knowledge of how to use it are additional requirements that may not be easily addressed. Pennington et al. (2019) found that speech treatment for children with dysarthria via Skype was feasible according to the parents. However, further implementation studies of speech assessment of children with dysarthria through telepractice should shed light on the practical feasibility, acceptability, and efficiency of collecting speech data online from this population for research and clinical purposes, as well as families' and SLPs' perspectives on the telepractice experience.

Additionally, as research and clinical practice rely heavily on both acoustic and perceptual information (Ansel & Kent, 1992; Sevitz et al., 2021), validating perceptual measures obtained from online recordings is a critical next step. For adults with Parkinson's Disease, most perceptual speech measures have been shown to be minimally affected by online transmission (Constantinescu et al., 2010). Research on similar measures for children with dysarthria is currently being conducted. Ultimately, a standardised protocol for online assessment of acoustic and perceptual speech characteristics to be used across clients with dysarthria and treatment sessions would maximise the possibility that any differences in outcome measures could be attributed to effects of the dysarthria or its treatment rather than to variations in the online assessment methods (Sevitz et al., 2021).

Conclusion

The validity of most speech measures obtained from online Zoom recordings in the present study suggests that using a commercially available online videoconferencing platform for speech research and clinical assessment at the client's or participant's home may be valid for obtaining most commonly-assessed acoustic measures; namely, F2 range of diphthongs, fricative-affricate duration difference, word duration/articulation rate, and mean F0, at both word and sentence levels, as well as vocal SPL range at word level. Even for SLPs who do not evaluate the specific measures we examined in the current study, our findings suggest the speech signal collected online using our methodology may be largely representative of the children's speech for assessment purposes. However, perturbation- and noise-based measures may not be as accurately represented in the children's online recordings. Thus, for assessment of most speech characteristics typically examined, remote speech evaluation via telepractice could be a viable alternative to in-person evaluation for individuals who face barriers to accessing in-person research and clinical services, including many children with dysarthria and potentially other populations with motor, geographical, or other challenges to accessing such services. These findings are an important step towards building an evidence base for valid online speech evaluation and treatment incorporating a subset of commonly-assessed acoustic measures.

Notes

1. Concurrent validity is the degree of agreement between two methods or assessments. Evaluation of concurrent validity involves examining the relationship between the outcomes produced by a new method and the outcomes obtained from an established standard method. A strong correlation between data collected by means of a standard method and data collected using the new method, for example, would indicate the concurrent validity of the new method (Murphy & Davidshofer, 1998). Historically, the standard for recording by means of an audio-recording device entailed the physical co-location of the experimenter and participant at the recording site. However, due to the Covid pandemic, the experimenter in the present study was not physically present with the participants. Instead, as described in the Method section, the experimenter sent the parents the recording equipment and guided them through the recording process remotely. While co-location of child and experimenter is more typical for offline recording, our setup permitted the child to be at home for the videoconferencing session, which is more typical for telepractice.
2. The audio-recorder microphone was entirely isolated from the laptop or iPad microphone used for the Zoom system, minimizing the risk of interference and signal contamination during the recordings.
3. In this study, the mouth-to-microphone distance was set at 30 cm for the online modality and 8 cm for the offline modality. These distances were selected to replicate commonly-implemented setups for each modality (Levy et al., 2017, 2021; Moya-Galé et al., 2023; Theodoros et al., 2016) and to permit calculation of SPL range. Because the range of SPL was calculated (as in Kuschmann & van Brenk, 2019), rather than absolute SPL, no further calculation was needed to adjust for the different distances. As considered in the Discussion, it is possible that the greater and more variable distance in the online condition could introduce more ambient noise, further reason to assess whether the resulting SPL range values differed between modalities.

Disclosure statement

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

ORCID

Kyunghae Hwang http://orcid.org/0009-0000-3666-2303
Frits van Brenk http://orcid.org/0000-0003-4777-919X
Megan J. McAuliffe http://orcid.org/0000-0003-3827-2215
Jiyoung Choi http://orcid.org/0009-0004-6403-387X
Jan G. Švec http://orcid.org/0000-0001-5095-7415
Erika S. Levy http://orcid.org/0000-0003-1718-3555

Data availability statement

In compliance with Institutional Review Board guidelines, the speech data from this study are not publicly accessible. However, deidentified participant data in spreadsheet format may be available upon request from the first author.

References

Abu-Arafeh, A., Jordan, H., & Drummond, G. (2016). Reporting of method comparison studies: A review of advice, an assessment of current practice, and specific suggestions for future reports. *British Journal of Anaesthesia*, *117*(5), 569–575. https://doi.org/10.1093/bja/aew320

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

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

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, *327*(8476), 307–310. https://doi.org/10.1016/s0140-6736(86)90837-8

Boersma, P., Weenink, D. (2020). *Praat: Doing phonetics by computer* (Version 6.1.14) [Computer software]. http://www.praat.org/

Campbell, D. R., & Goldstein, H. (2022). Reliability of scoring telehealth speech sound assessments administered in real-world scenarios. *American Journal of Speech-Language Pathology*, *31*(3), 1338–1353. https://doi.org/10.1044/2022_AJSLP-21-00219

Carrow-Woolfolk, E. (1999). *Test for auditory comprehension of language* (3rd ed.). Pro-Ed.

Constantinescu, G., Theodoros, D., Russell, T., Ward, E., Wilson, S., & Wootton, R. (2010). Assessing disordered speech and voice in Parkinson's disease: A telerehabilitation application. *International Journal of Language & Communication Disorders*, *45*(6), 630–644. https://doi.org/10.1044/2022_AJSLP-21-0021910.3109/13682820903470569

Curtis, J. A., Borders, J. C., & Troche, M. S. (2022). Visual analysis of swallowing efficiency and safety (VASES): Establishing criterion-referenced validity and concurrent validity. *American Journal of Speech-Language Pathology*, *31*(2), 808–818. https://doi.org/10.1044/2021_AJSLP-21-00116

Fox, C. M., & Boliek, C. A. (2012). Intensive voice treatment (LSVT® LOUD) for children with spastic cerebral palsy and dysarthria. *Journal of Speech, Language, and Hearing Research: JSLHR*, *55*(3), 930–945. https://doi.org/10.1044/1092-4388(2011/10-0235)

Freeman, V., & Decker, P. (2021). Remote sociophonetic data collection: Vowels and nasalization over video conferencing apps. *The Journal of the Acoustical Society of America*, *149*(2), 1211–1223. https://doi.org/10.1121/10.0003529

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

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: JSLHR*, *54*(2), 417–429. https://doi.org/10.1044/1092-4388(2010/10-0020)

Kuschmann, A., & van Brenk, F. (2019). Acoustic characterization of dysarthria in children with cerebral palsy: Exploring age-related effects. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences.* (pp. 3031–3035).

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: JSLHR*, *57*(5), 1666–1678. https://doi.org/10.1044/2014_JSLHR-S-13-0292

Levy, E. S. (2014). Implementing two treatment approaches to childhood dysarthria. *International Journal of Speech-Language Pathology*, *16*(4), 344–354. https://doi.org/10.3109/17549507.2014.894123

Levy, E. S., Chang, Y. M., Ancelle, J., & McAuliffe, M. (2017). Acoustic and perceptual consequences of speech cues for children with dysarthria. *Journal of Speech, Language, and Hearing Research: JSLHR*, *60*(6S), 1766–1779. https://doi.org/10.1044/2017_JSLHR-S-16-0274

Levy, E. S., Chang, Y. M., Hwang, K. H., & McAuliffe, M. J. (2021). Perceptual and acoustic effects of dual-focus speech treatment in children with dysarthria. *Journal of Speech, Language, and Hearing Research: JSLHR*, *64*(6S), 2301–2316. https://doi.org/10.1044/2020_JSLHR-20-00301

Mei, C., Reilly, S., Bickerton, M., Mensah, F., Turner, S., Kumaranayagam, D., Pennington, L., Reddihough, D., & Morgan, A. T. (2020). Speech in children with cerebral palsy. *Developmental Medicine and Child Neurology*, *62*(12), 1374–1382. https://doi.org/10.1111/dmcn.14592

Moya-Galé, G., Wisler, A. A., Walsh, S. J., McAuliffe, M. J., & Levy, E. S. (2023). Acoustic predictors of ease of understanding in Spanish Speakers with dysarthria associated with Parkinson's disease. *Journal of Speech, Language, and Hearing Research: JSLHR*, *66*(8S), 2999–3012. https://doi.org/10.1044/2022_JSLHR-22-00284

Murphy, K. R., & Davidshofer, C. O. (1998). *Psychological testing: Principles and applications* (2nd ed.). Prentice Hall.

Murray, E., McCabe, P., Heard, R., & Ballard, K. J. (2015). Differential diagnosis of children with suspected childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research*, *58*(1), 43–60. https://doi.org/10.1044/2014_JSLHR-S12-0358

Nip, I. S. B., & Garellek, M. (2021). Voice Quality of Children With Cerebral Palsy. *Journal of Speech, Language, and Hearing Research : JSLHR*, *64*(8), 3051–3059. https://doi.org/10.1044/2021_JSLHR-20-00633

Palisano, R., Rosenbaum, P. L., Walter, S. D., Russell, D., Wood, E. P., & Galuppi, B. E. (1997). Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Developmental Medicine and Child Neurology*, *39*(4), 214–223. https://doi.org/10.1111/dmcn.1997.39.issue-4

Patel, R., RAwan, S. N., Barkmeier-Kraemer, J., Courey, M., Deliyski, D., Eadie, T., Paul, D., Švec, J. G., & Hillman, R. (2018). Recommended protocols for instrumental assessment of voice: American Speech-Language-Hearing Association expert panel to develop a protocol for instrumental assessment of vocal function. *American Journal of Speech-Language Pathology*, *27*(3), 887–905. https://doi.org/10.1044/2018_AJSLP-17-0009

Pennington, L., Lombardo, E., Steen, N., & Miller, N. (2018). Acoustic changes in the speech of children with cerebral palsy following an intensive program of dysarthria therapy.

International Journal of Language & Communication Disorders, *53*(1), 182–195. https://doi.org/10.1111/1460-6984.12336

Pennington, L., Stamp, E., Smith, J., Kelly, H., Parker, N., Stockwell, K., Aluko, P., Othman, M., Brittain, K., & Vale, L. (2019). Internet delivery of intensive speech and language therapy for children with cerebral palsy: A pilot randomised controlled trial. *BMJ Open*, *9*(1), e024233. https://doi.org/10.1136/bmjopen-2018-024233

Sevitz, J. S., Kiefer, B. R., Huber, J. E., & Troche, M. S. (2021). Obtaining objective clinical measures during telehealth evaluations of dysarthria. *American Journal of Speech-Language Pathology*, *30*(2), 503–516. https://doi.org/10.1044/2020_AJSLP-20-00243

Shriberg, L. D., Lohmeier, H. L., Strand, E. A., & Jakielski, K. J. (2012). Encoding, memory, and transcoding deficits in Childhood Apraxia of Speech. *Clinical Linguistics & Phonetics*, *26*(5), 445–482. https://doi.org/10.3109/02699206.2012.655841

Storm, F. A., Petrarca, M., Beretta, E., Strazzer, S., Piccinini, L., Maghini, C., Panzeri, D., Corbetta, C., Morganti, R., Reni, G., Castelli, E., Frascarelli, F., Colazza, A., Cordone, G., & Biffi, E. (2020). Minimum clinically important difference of gross motor function and gait endurance in children with motor impairment: A comparison of distribution-based approaches. *BioMed Research International*, *2020*(1), 2794036. https://doi.org/10.1155/2020/2794036

Švec, J. G., & Granqvist, S. (2018). Tutorial and guidelines on measurement of sound pressure level in voice and speech. *Journal of Speech, Language, and Hearing Research: JSLHR*, *61*(3), 441–461. https://doi.org/10.1044/2017_JSLHR-S-17-0095

Theodoros, D. G., Hill, A. J., & Russell, T. G. (2016). Clinical and quality of life outcomes of speech treatment for Parkinson's disease delivered to the home via telerehabilitation: A noninferiority randomized controlled trial. *American Journal of Speech-Language Pathology*, *25*(2), 214–232. https://doi.org/10.1044/2015_AJSLP-15-0005

Titze, I. R., & Liang, H. (1993). Comparison of F0 extraction methods for high precision voice perturbation measurements. *Journal of Speech and Hearing Research*, *36*(6), 1120–1133. https://doi.org/10.1044/jshr.3606.1120

Utianski, R. L., Sandoval, S., Berisha, V., Lansford, K. L., & Liss, J. M. (2019). The effects of speech compression algorithms on the intelligibility of two individuals with dysarthric speech. *American Journal of Speech-Language Pathology*, *28*(1), 195–203. https://doi.org/10.1044/2018_AJSLP-18-0081

Waite, M. C., Cahill, L. M., Theodoras, D. G., Busuttin, S., & Russell, T. G. (2006). A pilot study of online assessment of childhood speech disorders. *Journal of Telemedicine and Telecare*, *12*(3_suppl), 92–94. https://doi.org/10.1258/135763306779380048

Weerathunge, H. R., Segina, R. K., Tracy, L., & Stepp, C. E. (2021). Accuracy of acoustic measures of voice via telepractice videoconferencing platforms. *Journal of Speech, Language, and Hearing Research: JSLHR*, *64*(7), 2586–2599. https://doi.org/10.1044/2021_JSLHR-20-00625

Weismer, G., & Berry, J. (2003). Effects of speaking rate on second formant trajectories of selected vocalic nuclei. *The Journal of the Acoustical Society of America*, *113*(6), 3362–3378. https://doi.org/10.1121/1.1572142

Wiig, E. H., Semel, E., & Secord, W. A. (2013). *Clinical Evaluation of Language Fundamentals.* (5th ed.). The Psychological Corporation.

Yoho, S. E., & Borrie, S. A. (2018). Combining degradations: The effect of background noise on intelligibility of disordered speech. *The Journal of the Acoustical Society of America*, *143*(1), 281–286. https://doi.org/10.1121/1.5021254

Zoom Video Communications, Inc. (2021). *Security guide.* https://irp-cdn.multiscreensite.com/ffb7ea18/files/uploaded/Zoom-Security-White-Paper