HHS Public Access

Protocol for the Connected Speech Transcription of Children with Speech Disorders: An Example from Childhood Apraxia of Speech

Catherine Barrett[a], **Patricia McCabe**[a], **Sarah Masso**[a], **Jonathan L. Preston**[b]

[a]Speech Pathology, The University of Sydney, Sydney, NSW, Australia

[b]Department of Communication Sciences and Disorders, Syracuse University, Syracuse, NY, USA

Abstract

Objective: While it is known that connected speech has different features to single-word speech, there are currently few recommendations regarding connected speech transcription. This research therefore aimed to develop a clinically feasible protocol for connected speech transcription. The protocol was then used to assist with description of the connected speech of children with childhood apraxia of speech (CAS), as little is known about their connected speech characteristics.

Participants and Methods: Following a literature review, the Connected Speech Transcription Protocol (CoST-P) was iteratively developed and trialled. The CoST-P was then used to transcribe 50 connected utterances produced by 12 children (aged 6–13 years) with CAS. The characteristics of participants' connected speech were analysed to capture independent and relational analyses.

Results: The CoST-P was developed, trialled, and determined to have adequate reliability and fidelity. The frequency of inter-word segregation (mean = 29) was higher than intra-word segregation (mean = 4). Juncture accuracy was correlated with intelligibility metrics such as percentage of consonants correct.

Conclusion: Connected speech transcription is challenging. The CoST-P may be a useful resource for speech-language pathologists and clinical researchers. Use of the CoST-P assisted in displaying CAS speech characteristics unique to connected speech (e.g., inter-word segregation and juncture).

Introduction

Children's communicative competence in connected speech tasks relates directly to their performance in everyday social interactions [1]. However, connected speech is sampled and transcribed less frequently than single words [2, 3]. Broad phonemic transcription of single-word naming assessments (and subsequent segmental analysis) is the most common procedure in speech assessment [2, 3]. While direct connected speech assessment is not commonly included in assessment batteries, there are some well-established connected speech sampling techniques. For example, connected speech is most frequently elicited through spontaneous conversational utterances [4] with a minimum of 50 utterances recommended to obtain a valid sample (e.g., [4–7]).

Collecting speech samples in different elicitation contexts (i.e., connected speech and single words) allows speech-language pathologists to gain a more complete understanding of the speech of children with speech disorders. The characteristics of speech that are present in single words may differ from those found in connected speech samples of the same child (e.g., [8–10]). Morrison and Shriberg [8] identified that connected speech provides information regarding error patterns present in inter-word transitions (alongside measures of single-word accuracy [e.g., percentage of consonants correct – PCC]). Coarticulatory differences present in spontaneous conversational speech should be documented when transcribed, illustrating the importance of effective transcription.

Morrison and Shriberg [8] (as well as Masterson et al. [9] and Wolk and Meisler [10]) compared the accuracy of single-word and connected speech samples of children with speech disorders using speech analysis procedures based on variations of the Independent and Relational Phonological Analysis framework [11]. Independent analyses frequently reported include inventories of singleton phonemes, consonant clusters, syllable shapes, word length, and the lexical stress. Relational analyses include the calculation of the PCC, percentage of vowels correct [12], syllable shapes correct, word lengths correct, and syllable stress patterns correct. Additionally, phonological processes and loss of phonemic contrasts were noted. Although the combinations of these analyses reported in the single-word and connected speech comparison studies [8–10] were comprehensive, they failed to consider the unique characteristics of connected speech.

More recent literature has highlighted the presence of characteristics which are unique to connected speech [13]. To speak in a connected string of sounds, as both children and adults do, there must be coordination of the motoric requirements of words, as well as of the prosodic and word juncture behaviours to link those words together [14]. Juncture between words is one of the unique characteristics of connected speech. Close juncture such as the linking /ɹ, w/, and /j/ is present between words where a word final vowel and word initial vowel are adjacent [15]. For example, consider the phrases "two eggs," "three eggs," and "four eggs." Here, most English speakers use a linking /w/ in /tu wɛgz/, a linking /j/ in /θɹi jɛgz/, and a linking /ɹ/ (a type of liaison) in /fɔ ɹɛgz/. Note that linking /ɹ/ is only relevant for analysis in non-rhotic dialects of English [16]. Open juncture includes assimilatory and elision behaviours that appear at word junctures in connected speech (cf. [13, 17, 18]). Assimilation across syllables occurs such that a neighbouring phoneme influences the sound

(e.g., "handbag" /hændbæg/ → [hæmbæg]). Elision refers to the omission of one or more sounds in a word at the boundary with an adjacent word (e.g., "curved round" /kɜvd raʊnd/ → [kɜv raʊnd]). Additionally, there is evidence to suggest that coarticulatory factors may be present in connected speech. Coarticulation (a form of assimilation) describes the impact of preceding and following phonemes on the articulatory characteristics of a segment. These changes may occur across syllable and word boundaries [19]. Coarticulatory characteristics of connected speech were highlighted by Daniloff and Moll [20], who proposed that the articulatory requirements for a given segment affected production of up to four preceding phonemes irrespective of word position.

Segmental and single-word analyses (such as those used by Morrison and Shriberg [8]) may not provide sufficient information regarding the characteristics unique to connected speech. Thus, it follows that single-word transcription methods may be insufficient for providing information on connected speech characteristics. Instead, transcription and analysis which considers the interaction of phonemes across word boundaries may provide a richer documentation of connected speech and allow utterances to be viewed as a series of non-linear tiers. Such analysis would involve transcription of words as connected (forming a string of phonemes). Considering speech as connected in this way reflects the perspectives proposed in the articulatory gesture literature (e.g., [21]), which suggests that the onset/conclusion of a connected string of phonemes can only be defined by a breath pause. Consistent with the literature (e.g., [22]), this paper will use the term "speech string" to describe connected phonemes which, in this case, span word boundaries. Selection of speech strings may be undertaken similarly to utterance segmentation, with non-hesitation pauses forming their boundaries [23]. A shift away from single-word-focussed transcription to a procedure involving transcription of speech strings is thus indicated. To guide this transition, a connected speech transcription protocol is required.

Considering connected speech characteristics is particularly important for clinical populations displaying difficulty with speech tasks of increasing complexity. Children with the motor planning disorder childhood apraxia of speech (CAS) are one such population [24]. Three key features of CAS were reported in the 2007 ASHA technical report [24]: (1) inconsistency of errors on phonemes (demonstrated on repeated productions of sounds, syllables, or words), (2) irregular lexical stress patterning that results in inappropriate prosody, and (3) coarticulatory transitions between syllables that are disrupted or elongated. Such disrupted coarticulatory transitions may be present between words (as characterised by Shriberg's pause marker [25]). However, there are limited data to quantify what the presence of these features, or others, may mean for connected speech in children with CAS. Perceptual analysis of the speech features of CAS is the current gold standard for diagnosis and is often undertaken primarily in a single-word context (e.g., [26–28]). Some studies have also utilised the Prosody-Voice Screening Profile Tool [29] or acoustic analysis (e.g., [30]) to describe the speech of children with CAS. The perceptual characteristics of connected speech in children with CAS have been described (e.g., [25, 31–35]). However, no quantitative connected speech profile has been developed for these children or others with more severe speech disorders. We may therefore learn more about the speech of children with CAS and others from a transcription method which allows for more detailed connected speech analysis.

Aims

This study had three inter-related aims: (1) to develop a transcription protocol for the connected speech of children with speech disorders, with potential use for typically developing populations; (2) to follow the newly formulated transcription protocol to transcribe the connected speech of children with CAS, thus determining the practicality of the protocol's use; and (3) to analyse the connected speech transcriptions to describe the speech of the children with CAS.

To address these aims, this research was conducted in two parts: Study 1 was the development of the Connected Speech Transcription Protocol (CoST-P) [36] and Study 2 was the implementation of the CoST-P with a population of children with CAS. These studies will be described in detail below.

Materials and Methods

Participants

A convenience sample of participants recruited for a previous study [37] were selected to determine the feasibility of the transcription protocol and to subsequently describe the children's connected speech. Twelve children (nine male; three female) aged 6–13 years with CAS were recruited from the Sydney metropolitan area (Table 1). All participants met the following three criteria: (1) diagnosis of CAS as determined through consensus of two speech-language pathologists (authors two and four) with expertise in the clinical diagnosis of CAS using the current gold standard [31]. For a child to be diagnosed with CAS, their speech needed to include the three core features of CAS as identified in the ASHA technical report [24], and four of the ten features of CAS described in Shriberg et al. [38]. (2) English as their primary language and at least one parent with English as their primary language (note that all children used a non-rhotic dialect of English). (3) Non-verbal intelligence and receptive language within the normal range as determined through the Test of Non-Verbal Intelligence – third edition [39], and the Clinical Evaluation of Language Fundamentals – fourth edition [40]. Children were excluded from the study if craniofacial abnormalities were noted in an oral musculature assessment or if they had other motor speech disorders as assessed through the Goldman-Fristoe Test of Articulation – second edition [41] and a connected speech sample. Children were also excluded from the study if they had hearing disabilities as determined through a hearing screener or if they had other developmental or genetic diagnoses.

All assessments were undertaken at The University of Sydney and were audio and video recorded. Informed consent was provided by the parent/caregiver of all participants before inclusion in the study. The study was approved by The University of Sydney Ethics Committee (2015/516).

Fifty spontaneous connected speech utterances were extracted from each participant's initial assessment session recording. These utterances were produced during free conversational speech with the assessing speech-language pathologist, during a deliberately collected connected speech sample based on description of the Park Play Scene [42] or during

connected speech elicited during completion of the expressive component of the Word Classes subtest of the Clinical Evaluation of Language Fundamentals – fourth edition (CELF-4) [40]. That is, they were not contiguous. To be considered as connected speech, the utterance needed to include a minimum of two words.

Study 1: Protocol Development and Validation

Materials and Methods

A literature review was undertaken by searching relevant databases (e.g., Scopus, Education Resources Information Centre, and Linguistic and Language Behaviour Abstracts) with key terms including "transcription," "connected speech," "spontaneous speech," "conversational speech," and "conversational sample". This literature review informed the development of an initial version of the CoST-P [36], which provided a list of 25 recommendations for the transcription of the connected speech of children with speech disorders. The initial version did not provide a comprehensive explanation of the preferred ways to capture connected speech and did not include examples or explanations of terminology and recommendations. A series of trial-and-revise cycles using the connected speech samples from the participants of Study 2 were then undertaken to develop the protocol iteratively. The final version of the protocol is provided in Appendix 1. Thus, following the procedures laid out in the CoST-P, utterances from each child were transcribed by the first author directly into Phon 3.0 [43], a freely available speech analysis software.

Transcription of connected speech samples involved broad phonemic transcription with use of additional diacritics. Sennheiser HD280 Pro headphones were used for all transcriptions. A limit of five listening repetitions of each audio segment was imposed for transcription purposes (to eliminate the effects of auditory-visual illusions [44]). All transcriptions were reviewed once the transcription was complete to ensure consistency to the final protocol.

A protocol fidelity measure was developed to evaluate implementation of the CoST-P. The fidelity (i.e., a transcriber's adherence to the protocol) was assessed using a transcriber who had been trained in phonetic and phonemic transcription but had not been involved in the development of the protocol. For the purposes of determining the protocol fidelity, participant samples were purposively selected to capture a range of speech disorder severity. Thus, the protocol fidelity measure was used to determine an independent transcriber's application of the CoST-P with ten utterances from three participants: (1) one child with a low PCC (i.e., considered too unintelligible for relational analysis), (2) one child with a moderate PCC (between 70–80% of consonants correct), and (3) one child with a relatively high PCC (above 90%). The fidelity rater chose to transcribe by hand, and the entire transcription process was filmed using an iPhone camera. Following the completion of the three sets of sample transcriptions, the first author viewed the footage and completed the protocol fidelity measure for each sample transcribed. No changes to the protocol were made following completion of the fidelity assessment.

Intra- and inter-rater reliability of connected speech transcriptions were conducted on a randomly selected 20% of the samples. Intra-rater reliability transcription was conducted 4 weeks after the initial transcription. The second author transcribed a randomly selected 20%

of the samples for inter-rater reliability. The transcription protocol remained constant, with each utterance listened to a maximum of five times. Both point-point reliability of phonemes and total point-point reliability (including diacritics and notes) were calculated.

Results

A protocol containing six broad recommendations, ten sub-recommendations, and 32 considerations was developed after eight trial-and-revise cycles. The recommendations and considerations were then operationalised into the tasks included in the CoST-P (Appendix 1). Multiple versions of the CoST-P were developed to account for considerations which arose during the transcription process (Table 2). Although the CoST-P may be used for either paper-based or computer-based transcription, some of the revisions were prompted by challenges completing the computer-based transcription using Phon [43]. The CoST-P provides recommendations regarding the transcription of the children's target and actual productions. Broad phonemic transcription should be used, with additional diacritics to be included when transcribing distortions and errors in the children's actual productions. A key feature of the CoST-P is the recommendation for words to be transcribed as connected where they run together naturally in connected speech (as this allows for the identification and transcription of inter-word features such as inappropriate pause). Other important considerations relate to management of contextually and/or dialectally appropriate reduced forms, assimilation, consonant deletion, between-syllable epenthesis, and juncture (linking / ɹ, w, j/). There is also a section of the CoST-P which relates to revision of the target transcription based on the context of the actual production to allow for correct alignment (matching) between the orthographic, target, and actual transcriptions.

Total protocol fidelity based on the transcription of ten utterances from three samples yielded a score of 88%. Transcription of features such as dysfluencies had high fidelity (100%). However, transcribing the utterances as speech strings had low fidelity (55%). The fidelity rater commented that she had difficulty remembering to transcribe the speech as connected, and frequently forgot to include the juncture phonemes. Intra- and inter-rater reliability scores of 92 and 86%, respectively, were achieved regarding broad phonemic transcription (excluding diacritics) on 3,205 data points. Scores of 86% (intra-rater reliability) and 79% (inter-rater reliability) were achieved when diacritics were included. There was no substantial pattern to the disagreement between raters. However, one potential source of difference was the frequency of one transcriber using a glottal stop phoneme where the other transcriber marked an inappropriate pause. Following the CoST-P, transcription of a child's production of 50 utterances of connected speech into the three tiers ("Orthography," "IPA Target," and "IPA Actual") took researchers between 5 and 7 hours.

Study 2: CAS Connected Speech Characteristics: Protocol Implementation

Materials and Methods

All transcriptions were completed using the CoST-P and entered into Phon 3.0 [43]. Within Phon, data entry consisted of the orthography of the phrase, target production of the phrase (IPA Target tier), and the realised or actual production of the participant (IPA Actual tier). It is also possible to document notes for each utterance within the Notes tier in Phon [45].

Once completed, these connected speech samples were analysed to address the secondary aim of the study: description of the connected speech of children with CAS.

Speech Analysis

Speech characteristics identified in the independent and relational phonological analyses originally described by Stoel-Gammon and Dunn [11] and modified by Baker [46] were analysed using Phon [43]. The independent and relational analyses which were conducted are outlined below.

Independent Analyses.—Independent analyses included calculations of the number of phonemes, syllables, and connected speech strings (as opposed to words), as well as analysis of the types of stress patterns and syllable shapes. The output consisted of counts of primary, secondary, and unstressed syllables, as well as the frequency of occurrences of inter- and intra-word segregation. Counts of instances of juncture (i.e., the presence of linking /ɹ, w/, and /j/), dysfluency, and voice (i.e., counts of resonance/voice diacritics) were also undertaken.

Relational Analyses.—The relational analysis included calculation of the percentage of correct phonemes (consonants and vowels), syllables, connected speech strings, stress patterns, and instances of juncture. Note that close juncture behaviours will be referred to as juncture in this paper. Additionally, the PCC-revised (PCC-R), which does not account for distortions [47], was calculated. Also, the proportion of whole-word proximity [48] – a measure of the actual phonological mean length of utterance (PMLU) divided by the target PMLU – was calculated.

Both built-in analyses (e.g., "Phonemic Inventory" and "Word Level Analysis of Polysyllables" [49]) and custom queries (e.g., frequency counts of stress patterns) in Phon 3.0 [43] were used to extract the required data from the connected speech samples. The data produced by Phon were summarised using Microsoft Excel. This involved summing values (e.g., number of phonemes correct) to produce totals for each participant. Additionally, totals across participants, means, and standard deviations were calculated. Note that some analyses at word level were not undertaken as speech strings rather than words were transcribed.

In addition to the data extracted from children's connected speech, single-word data for the participants were sourced from the original study file. The PCC for single words was calculated based on the participants' raw scores on the Goldman-Fristoe Test of Articulation – second edition [41]. PCC was obtained by first subtracting the raw score (number of errors) from the total number of target phonemes, then dividing by the total number of targets, and finally multiplying by 100.

Statistical Analysis

Statistical analysis was conducted using Microsoft Excel. For the descriptive component of the study, the analysis involved calculations including means and standard deviations. Pearson's correlation coefficient (*r*) was used to evaluate the relationship between the children's PCC in single words and PCC in connected speech. Paired two-tailed *t* tests were

performed to determine whether the differences between single-word and connected speech PCCs were significant. A default statistical significance level of $p = 0.05$ was selected.

Results

Independent Analysis

Independent analyses were conducted on all twelve participants. The samples consisted of multiple connected speech tasks. All samples contained free connected speech. Six samples contained connected utterances based on the Park Play Scene [42] and six contained connected speech elicited during the completion of the CELF-4 [40]. The participants produced a mean of 1,028 (SD = 491) phonemes across the 12 50-utterance speech samples analysed. Participants used more syllables with primary stress (m = 258, SD = 77) than secondary stress (m = 33, SD = 15) or unstressed (m = 28, SD = 39) syllables. Proportionally, syllables with primary stress were an average of 81% of total syllables, secondary stress were 10%, and unstressed syllables comprised 9%. Participants used a mean of 14 (SD = 6) orthographic words containing three or more syllables. There was a mean of 29 (SD = 21) instances of inter-word segregation in the samples, compared to the mean of 4 instances of intra-word segregation. Table 3 provides a summary of the independent analyses completed.

Relational Analysis

Relational analysis was conducted on the speech of ten participants. The speech of two participants was considered too unintelligible to reliably transcribe the target productions (consistent with the protocol instructions).

The mean PCC was 80 (SD = 10) and the mean percentage of vowels correct was 88 (SD = 9). The accuracy of the speech strings (i.e., the percentage of correct phonemes and stress patterns) was 35% (SD = 9). The accuracy of juncture use (i.e., the percentage of appropriate use of the linking /ɹ, w/, and /j/) was 54% (SD = 34). The mean percentage of correct productions of pre-tonic unstressed syllables was 73% (SD = 17), whereas the mean for stressed initial syllables was 94% (SD = 5). Figure 1 provides a summary of the relational analyses. Supplementary Material A (available online at www.karger.com/doi/10.1159/000500664) provides individual participant data for a range of independent and relational measures.

The summary of phonological process analysis is provided in Table 4. Speech processes which were present on less than 5% of relevant phonemes (for each child) are not included (to account for typical errors or "slips of the tongue" during speech production). Excluded segmental processes therefore included velar fronting, coronal backing, and lateralisation, as well as the structural processes of metathesis, epenthesis, and assimilation. The process which occurred on the largest percentage of targets was deaffrication (i.e., affricates replaced with fricatives), affecting a mean of 32% of affricates (SD = 39). However, deaffrication was only present in the speech of six out of the ten children. All participants had deletion errors and deleted a mean number of 55 phonemes (SD = 34). The most common deletion occurred on final consonants (defined in this study as deletion of some or all coda consonants), with a

mean of 32 instances per child (SD = 22), resulting in a mean of 17% (SD = 15) of total word-final consonants being deleted.

Relationships between Speech Production Variables

Relationships between variables were examined using Pearson's correlations. Most correlations produced an insignificant result and a small r value. However, several of the more interesting results (which may warrant further study) are reported here for the purposes of discussion. The correlation between juncture accuracy and PCC was not significant, with an r value of 0.62 ($p = 0.056$). However, there was a significant positive correlation between juncture accuracy and percentage of phonemes correct (PPC) ($r = 0.66$, $p = 0.038$). The correlations between instances of inter-word segregation and PCC or PPC were not significant, characterised by r values of 0.07 ($p = 0.85$) and 0.1 ($p = 0.78$), respectively. The correlation between PCC and final consonant deletion was characterised by an r value of 0.58 ($p = 0.08$).

There was a relationship between the children's PCC in single words and PCC in connected speech ($r = 0.75$, $p = 0.01$). Results of a two-tailed paired t test indicated no significant difference between participants' single-word and connected speech PCC scores (t [df 9] = 0.8, $p = 0.4$).

Discussion

This study has resulted in the development of the CoST-P [36], which provides an evidence-based method for consistent connected speech transcription. During the CoST-P development phase, multiple versions were trialled, with the final protocol providing six broad recommendations and ten sub-recommendations. The sub-recommendations indicate the process which should be undertaken to complete the transcription in the most efficient and effective manner. Additionally, 32 considerations are documented which provide instruction regarding specific contingencies that may be required when transcribing the connected speech of children (e.g., how to manage the occurrence of contextually/linguistically appropriate reduced forms).

The CoST-P was applied to the transcription of the connected speech of children with CAS. The CoST-P was used with adequate reliability and fidelity scores (consistent with Landis and Koch [50]) and therefore, can be considered for use in the study of children's connected speech. Transcription of 50 utterances using the CoST-P (including the orthography, the target output, and the actual production) took between 5 and 7 h per child. This is a significant time commitment, which will limit use of the current CoST-P in daily speech-language pathology practice. Further study, focussed on the formulation of a quick version of the protocol, may help address this issue.

CAS Characteristics

Transcription of the speech of children with CAS using the CoST-P allowed for the quantitative description of connected speech characteristics of this population. These descriptive results provide clinicians and researchers with an initial profile of the connected speech of children with CAS. Overall, the independent and relational analysis data appear to

be consistent with the CAS speech characteristics reported in the single-word literature. For example, inappropriate prosody (in relation to lexical/phrasal stress) is reported to be a core feature of the speech of children with CAS [24]. The connected speech results here supported this finding, with a mean of only 62% of the speech string stress patterns produced correctly. Additionally, there was a relatively large standard deviation to this accuracy measure, reflecting reported heterogeneity in children with CAS [24]. These findings are of particular note in light of typically developing children's innate receptivity to the lexical stress patterning of the ambient language [51] (e.g., natural production of trochaic disyllabic nouns and iambic productions of disyllabic verbs [52]).

Although not specific to inter-word juncture, another core CAS feature is lengthened/ disrupted coarticulatory transitions between sounds and syllables [24]. This was present in connected speech, represented by the high mean frequency of inappropriate inter-word segregation (pauses). However, there was significant variation in the count of these instances across the participants. One participant's sample had only five instances of inappropriate inter-word segregation, whilst another participant had 67 inappropriate inter-word segregations. The accuracy of use of juncture (linking) phonemes was 54%; however, the variation was again large for this measure (SD = 34). These results suggest that inappropriate inter-word segregation (and analysis of juncture) may not serve as the most effective means of diagnosis of CAS due to the significant variation in participant presentation of this feature in free connected speech. Also, there was no correlation found between the instances of inappropriate inter-word segregation (pauses) and PPC or PCC. This indicates that there is no link between the occurrence of this potential CAS marker and the traditional metrics of determining speech disorder severity. This lack of correlation replicates the results of Shriberg et al. [25], who reported that the pause marker scores (which provide a percentage of the instances of inter-word pauses) were not associated with other measures of speech precision or competence for children with CAS (e.g., phoneme accuracy, lexical stress, and speech rate). However, the present study identified a moderate correlation between accuracy of juncture and both PPC and PCC, indicating that juncture accuracy may be a beneficial metric for determining disorder severity amongst children with CAS.

Interestingly, whilst all children displayed inappropriate inter-word segregation (with a mean of 29 instances), only seven participants had inappropriate intra-word segregation, with a mean of four instances per child. One potential source of this discrepancy is the phenomenon of avoidance reported by Morrison and Shriberg [8] whereby children do not produce certain challenging phonemes, syllable shapes, and word shapes. This may be a factor as children with CAS are known to have increasing difficulties with increasing word length [24]. The results support this, with the children producing a mean of only 14 words containing three or more syllables across all 50 utterances (4% of all words). This result contrasts with Stoel-Gammon's study of typically developing children aged 30 months, which found that polysyllabic words constituted 8% of total words produced [53]. Furthermore, the count of polysyllabic words in this study may be inflated, as six of the participants discussed the "Park Play Scene" [42], which includes multiple polysyllabic items. The overall lower frequency of polysyllabic words could explain the presence of fewer instances of intra-word segregation and would also partially account for the relatively

high proportion of syllables with primary stress. However, further research would be needed to ascertain whether a phenomenon such as avoidance is involved in word choice for these children.

Although CAS is a disorder of motor planning, it may be of benefit to place the results in phonological context, considering that these children are in a period of global development (e.g., [54, 55]). Furthermore, the intersection between phonology and motor speech has been discussed in the literature (e.g., [56–58]), with children with reduced motor speech success demonstrating difficulty learning expressive phonological contrast (e.g., [59, 60]). All participants demonstrated the structural phonological process of deletion, suggesting that this may be a common feature amongst children with CAS. The presence of deletion may have resulted from difficulties with coarticulatory transitions (a core diagnostic feature of CAS) [24]. Deletion of word codas (partial or complete) was the most common process, reflecting the literature which indicates that this is common amongst less intelligible children (e.g., [61]). This was further supported by the moderate correlation found between PCC and instances of word coda deletion. The segmental phonological process of deaffrication was present on 32% of the total affricates. However, as only six of the ten children produced this error and two of the children had 100% deaffrication, the results were skewed. While structural processes such as metathesis, epenthesis, and assimilation have been described as common in the speech of children with CAS (due to phoneme and syllable sequencing difficulties [62]), it is interesting to note these structural processes were present on less than 5% of total phonemes. It is possible that avoidance of complex words affected these results [8].

Unstressed initial syllables in iambic (wS) stress patterns (considered "un-footed" in the non-linear phonology literature) are more vulnerable to omission than trochaic patterns (i.e., words containing "footed"/stressed initial syllables) [63]. This was reflected by the greater accuracy of stressed versus unstressed initial syllables in the sample. The bias towards productions of trochaic patterns is a feature of the speech of children with general speech disorders, and thus, these results may not reflect the irregular lexical stress patterns identified as a core feature of CAS [24]. However, it is possible that whilst the patterns themselves are not irregular, the distinguishing factor is the continued presence of these errors past the age at which this issue typically resolves.

There was a strong correlation between PCC in single words versus PCC in connected speech. Five of the ten participants had a higher PCC in connected speech, while three had a higher PCC in the single-word task. The two remaining children had the same PCC across both tasks. The children's ability to control the content of their connected speech output may have led to the higher average PCC in the connected speech productions, as discussed in the literature (e.g., [8, 64]). Further, the sample size of 10 children with 50 utterances may have limited our ability to identify a more prominent correlation between the data sets and draw more definitive conclusions regarding the impact of avoidance or other factors.

We note that independent analysis of the connected speech of children with severe speech disorders (who have very poor intelligibility) provides information on inventories of phonemes, syllable shapes, and stress patterns. However, the unstructured connected speech

context meant that transcription of target utterances was not possible and therefore, relational analyses and some independent analyses (such as identification of instances of juncture or dysfluency) could not be captured. More controlled assessment tasks, such as imitated sentence tasks, with known speech targets may be of greater value for both study and therapy goal planning for children with severe speech disorders. Such tasks may also reduce transcription time substantially.

CoST-P

Although there was no significant difference in single-word versus connected speech consonant accuracy measures, transcription following the CoST-P allows for consideration of connected speech in context (rather than consonant accuracy alone). Thus, the CoST-P captures a number of speech characteristics which may otherwise not be considered by speech-language pathologists and others interested in child speech development and disorder.

Two artificial utterance samples have been provided in Table 5 to assist with description of these features. The first displays transcription following a word-by-word approach; the second displays transcription using the CoST-P. Although some similarities can be observed, the connected speech transcription yields a different description and interpretation of the child's speech.

Both methods allow for the development of a phonemic inventory (e.g., consonants: /k, l, w, d, m, n, t, f, s, h, r/, and vowels: /aɪ, ʌ, ə, a, ɛ, ɒ, ɪ, u, æ/) and analysis of phonological processes (e.g., stopping of fricative /ð/ and affricate /dʒ/). Also, calculation of other features (such as nasality) is possible using both methods. However, the context of within-speech string position may provide more information than within-word position. For example, word-by-word transcription may note inappropriate hyper-nasality on the /ɒ/ in /ɒ̃ɹəndʒ/ as there is no adjacent nasal consonant affecting its production. Contrastingly, transcription using the CoST-P highlights the proximity of the /ɒ/ to a nasal consonant (/m/), demonstrating the appropriateness of hypernasality in this instance [65]. Similarly connected speech transcription provides a context for the presence of appropriate aspiration of voiceless plosives (whereas the single word transcription can highlight this as an error due to word boundaries (e.g., hæt^h ɒn). Both methods can capture additional speech characteristics such as dysfluencies (no instances present in the sample). Also, while both methods allow for counts of intra-word segregations (one instance in this example), transcription using the CoST-P enables documentation of frequency of inter-word segregations (two instances displayed in the example). The CoST-P also allows for analysis of instances of inter-word juncture behaviours. Further research is required to understand the extent of the implications of use of the CoST-P in assessment, diagnosis, and treatment. Using the CoST-P to formulate connected speech norms (and identify disorder-specific connected speech characteristics) may result in connected speech transcription becoming more important as a contributor to accurate and detailed diagnosis. Furthermore, the additional speech characteristics which can be observed through connected speech transcription may assist with the development of more meaningful therapy approaches, goals, and therapy targets.

The reliability score was acceptable (i.e., over 85%). Of note was the difference in the raters' perceptions of pause, where one rater transcribed glottal stops and the other marked inappropriate inter-word segregation. Both replacement of word codas by glottal stops and consonant deletion (resulting in inappropriate inter-word pauses) have been reported in the literature as strategies to minimise articulation effort [18]. However, it was noted by Newton [66] that glottal stops do not occur on instances where the first consonant within the cluster is a nasal or approximant. This suggests that the protocol may benefit from further refinement to include specific recommendations regarding appropriate use of glottal stops and pause markers.

The fidelity was in the desirable range, indicating that the areas of difference were not of significant concern. However, the poor fidelity results regarding transcribing connected speech in strings (as opposed to word by word) highlights that a change in conceptualisation of connected speech will be required before the CoST-P may be regularly used. That is, transcription of speech strings and inclusion of features such as juncture (e.g., a linking /ɹ/) are not the current transcription default in speech-language pathology research or practice. Therefore, time and training will be required to facilitate the shift and establish more regular use of the CoST-P. However, this change may be justified in light of the preliminary results of this study, which indicate that speech string transcription assists in capturing the unique characteristics of connected speech.

Study Limitations and Future Research Directions

There are several factors which have limited the scope of this study in relation to both the CoST-P and the description of CAS. The use of a convenience sample of pre-recorded spontaneous productions may have resulted in the capturing of speech which did not contain a complete segmental or structural inventory (as the children were not required to produce specific phonemes). Studies containing structured and unstructured speech tasks transcribed using the CoST-P may provide information on the effects of avoidance, especially on unique connected speech features such as juncture. Additionally, a comparison sample from a normative, typically developing population may assist in placing the results regarding CAS speech characteristics in context. Future studies which assess the use of the CoST-P for children with other speech disorders, such as dysarthria and more severe phonological disorders, are required. This will help determine the protocol's reliability and fidelity more broadly, thus ascertaining the suitability of its use with a wide range of clinical and developmentally typical populations. Additionally, further studies are required to determine the CoST-P's applicability for children speaking different dialects of English (including non-rhotic dialects) as well as its use for other languages.

An important next step in increasing the practicality of the CoST-P is the development of a version which reduces the time required to complete a transcription. Additionally, further training resources and examples for transcribers, as well as alterations to the CoST-P based on the results of the fidelity assessment may increase the protocol's value for researchers and clinicians.

Further study is needed to determine the most appropriate tools and analyses to use to glean the most meaningful information following connected speech transcription. Direct

comparison between single words and connected speech (transcribed using the CoST-P) is another avenue for future research. This may highlight the discrepancies between the two output forms previously identified in the literature (e.g., [8]) and may display the benefit of use of the CoST-P to identify the elements that are not present in the transcription of single words. Additionally, studies to determine the extent to which connected speech features such as juncture may be of benefit in assessment and diagnosis are an important step. Further insights into the foundations of connected speech may be obtained by considering this data through other theoretical lenses.

Conclusion

The transcription and analysis of the connected speech of children with speech disorders poses multiple challenges to clinicians and researchers. This study has discussed the development of a transcription protocol for connected speech and its subsequent use with a sample of children with CAS. Conceptualising connected speech as a string of sounds rather than as individual word units assists in developing a deeper understanding of a child's speech capabilities and limitations. Given the preliminary nature of the results, limited generalisation of these results is recommended. However, it can be seen that transcription using the CoST-P provides information on features unique to connected speech, such as the presence of inappropriate inter-word segregation and juncture. The CoST-P may be a meaningful resource for speech-language pathologists, promoting increased use and awareness of the benefits of connected speech sampling and transcription.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

Appendix

Connected Speech Transcription Protocol (CoST-P) (Barrett et al., 2018)

Before phonemic transcription

1 **Gather information about the child's dialect/languages spoken at home** (other than the target language). Phonemes, syllable/word structures, and stress patterns will vary across languages/dialects. The Multilingual Children's Speech Sound Disorders website (www.csu.edu.au/research/multilingual-speech/home) is a suitable starting place.

2 Document the conversational context

3 Obtain a minimum of 50 of the child's connected utterances, and orthographically transcribe as many of the productions as possible

During phonemic transcription

1 Transcribe the target production using IPA (where possible)

(a) Use broad phonemic transcription.

(b) Transcribe words as connected (i.e., without boundaries) where words run together naturally in connected speech.

General considerations

i Transcribe contextually appropriate and/or dialectically appropriate:

- reduced forms

    For example, /gɒnə/ may be an appropriate shortening of /gɒʊɪŋ tu/.

- assimilation of consonants within/between words

    For example, "sandwich" may be produced as /sæmwitʃ/.

- consonant deletion

    For example, deletion of /t/ or /d/, such as /doʊnt noʊ/ → /doʊn noʊ/.

- between-syllable epenthesis

    For example, "hamster" is commonly produced as /hæmpstə/.

- juncture (i.e., a linking /ɹ, w/, or /j/ should be present on all occasions between two adjacent vowels within a speech string)

    For example, "car is" = /kaɹɪz/; "cow is" = /kaʊwɪz/; 'they are' = /ðeɪja/.

ii. Do *not* use diacritics to transcribe allophonic variation where variation is linguistically appropriate. For example, hypernasality on vowels preceding or following a nasal consonant (e.g., /ænt/ or /mæn/), aspiration on voiceless plosives before a stressed vowel (e.g., /ti/).

Considerations when unintelligible

Where an utterance is considered unintelligible on first inspection:

Considerations when unintelligible

i. Move to the next utterance and return to the difficult section once you are more familiar with the child's speech output.

ii. Transcribe as much information for the target as possible (consider context).

iii. If the utterance is completely unintelligible and the target is not possible to ascertain, leave the target transcription blank. Independent analysis may still be possible using these utterances.

(c) Note that you may need to return to "target" productions after transcription of "actual" productions (see considerations for revision of target form [3a]).

2 Transcribe actual productions using IPA (where possible)

(a) Use broad phonemic transcription with phonetic transcription of phonemes where required (i.e., use additional diacritics to represent distortions/errors). For example, use approximation, lateralisation, dentalisation, or vowel lengthening diacritics.

(b) If the error cannot be described with the use of a diacritic, search the full IPA chart to find a matching phoneme (i.e., do not limit transcription to use of English phonemes).

(c) Do not alter the audio file used for transcription (e.g., segmentation, pitch shifts, or speed reduction) – altering an utterance may result in loss of phonemic information.

(d) Transcribe all:

i. moments of within-word syllable segregation using a symbol or marker (for example, a pause marker if using Phon).

ii. moments of inappropriate between-word segregation as a lengthened pause (note that appropriate pauses between words can be marked with a space).

iii. moments of stereotyped speech (e.g., counting). However, it is important to note "stereotyped utterance" below.

iv. moments of dysfluency phonemically (transfer into IPA target and orthography). Note each occurrence of dysfluency with a code (e.g., "$").

v. reduced forms of words (see target transcription general considerations for an example).

vi. consonant assimilation within/between words (see target transcription general considerations for an example).

vii. consonant deletion between words (see target transcription general considerations for an example).

viii. between-sound/between-syllable epenthesis (see target transcription general considerations for an example).

ix. instances of juncture (linking /ɹ, r, w, j/) (see target transcription general considerations for an example).

x. moments of children whispering or talking in a sing-song prosody (note with an appropriate symbol).

General considerations

i. Where vowel elongation occurs, and this would alter the vowel transcription, transcribe as the new vowel (e.g., /ɪ/ → /i/). If the vowel is lengthened further, transcribe as the new phoneme with the relevant diacritic.

ii. Apply relevant diacritics for distorted consonants (e.g., approximation, lateralisation, dentalisation, or vowel/consonant lengthening diacritics).

iii. Mark voice change using diacritics (e.g., resonance or ejective diacritics) when linguistically inappropriate. Do not mark where allophonic variation is expected (see example in target transcription general consideration ii). If there is consistent inappropriate voice, mark this in notes, as this may be indicative of a voice disorder rather than phoneme-associated voice change.

iv. Where there is lack of audible release on a final consonant (/p, b, t, d, k, g/), but the consonant is marked, transcribe as target consonant (e.g., "map," "jog").

v. Where a pause is audible, make a consistent transcription decision regarding whether the child is producing a glottal stop or an inappropriate pause.

Considerations when unintelligible

i. Transcribe all transcribable phonemes (following protocol above).

ii. If the manner of consonants is audible, but the place of consonant is unidentifiable, transcribe using a generic marker from the extended IPA (e.g., X).

iii. If the place and manner of a consonant is not audible, transcribe with a generic consonant marker (e.g., C).

iv. if a vowel's place is unknown, transcribe with a generic vowel marker (e.g., V).

3 Match the orthographic transcription to the IPA Target and the IPA Actual where possible

(a) Revision of the Target form

i. If the Target is transcribed as an altered form (i.e., as a reduced form, with consonant assimilation or deletion, or with between-sound/syllable epenthesis), but is produced as the citation form by the child, determine the appropriateness based on the linguistic context/ambient language. If citation form is appropriate in the Actual, alter the Target to match.

ii. If contextually/linguistically appropriate juncture (e.g., linking /ɹ, r, w, j/) has been transcribed in the Actual but has not been included in the transcription of the Target, alter target to match.

iii. If contextually/linguistically appropriate altered forms are present in the Actual production, but are not transcribed in the Target, alter Target to match.

iv. If a between-word segregation marker is present in the Actual, but juncture has been transcribed in the Target, delete the juncture phoneme from the Target, keeping the words connected. This ensures that the error is being marked by one (rather than two) differences between the Target and Actual (essential if using transcription software).

(b) Where alignment decisions are not clear, refer to an evidence-based protocol for syllabification rules (e.g., Grunwell [1987]).

(c) Once alignment is completed, the next steps depend on the goals of transcription. You may:

i. Informally study the actual transcription and compare to the target transcription to understand the child's connected speech competence (with regards to key features/areas of difference) OR

ii. Study each disparity between the target and actual on a point-point basis to identify each error/difference, allowing for more precise study of the child's connected speech.

References

1. DuBois EM, Bernthal JE. A comparison of three methods for obtaining articulatory responses. J Speech Hear Disord. 1978 8;43(3):295–305. [PubMed: 692096]

2. Skahan SM, Watson M, Lof GL. Speech-language pathologists' assessment practices for children with suspected speech sound disorders: results of a national survey. Am J Speech Lang Pathol. 2007 8;16(3):246–59. [PubMed: 17666550]

3. Mcleod S, Baker E. Speech-language pathologists' practices regarding assessment, analysis, target selection, intervention, and service delivery for children with speech sound disorders. Clin Linguist Phon. 2014 Jul-Aug;28(7–8):508–31. [PubMed: 25000375]

4. Broome K, McCabe P, Docking K, Doble M. A systematic review of speech assessments for children with autism spectrum disorder: recommendations for best practice. Am J Speech Lang Pathol. 2017 8;26(3):1011–29. [PubMed: 28772287]

5. Binger C, Ragsdale J, Bustos A. Language sampling for preschoolers with severe speech impairments. Am J Speech Lang Pathol. 2016 11;25(4):493–507. [PubMed: 27552110]

6. Bleile K Evaluating articulation and phonological disorders when the clock is running. Am J Speech Lang Pathol. 2002;11(3):243–9.

7. Hodson BW, Scherz JA, Strattman KH. Evaluating communicative abilities of a highly unintelligible preschooler. Am J Speech Lang Pathol. 2002;11(3):236–42.

8. Morrison JA, Shriberg LD. Articulation testing versus conversational speech sampling. J Speech Hear Res. 1992 4;35(2):259–73. [PubMed: 1573866]

9. Masterson JJ, Bernhardt BH, Hofheinz MK. A comparison of single words and conversational speech in phonological evaluation. Am J Speech Lang Pathol. 2005 8;14(3):229–41. [PubMed: 16229674]

10. Wolk L, Meisler AW. Phonological assessment: a systematic comparison of conversation and picture naming. J Commun Disord. 1998 Jul-Aug;31(4):291–310. [PubMed: 9697041]

11. Stoel-Gammon C, Dunn C. Normal and disordered phonology in children. University Park Press; 1985.

12. Shriberg LD, Kwiatkowski J. Phonological disorders III: a procedure for assessing severity of involvement. J Speech Hear Disord. 1982 8;47(3):256–70. [PubMed: 7186561]

13. Howard S The interplay between articulation and prosody in children with impaired speech: observations from electropalatographic and perceptual analysis. Adv Speech Lang Pathol. 2007;9(1):20–35.

14. Local J, Walker G. Methodological imperatives for investigating the phonetic organization and phonological structures of spontaneous speech. Phonetica. 2005 Apr-Dec;62(2–4):120–30. [PubMed: 16391498]

15. Wells B Junction in developmental speech disorder: A case study. Clin Linguist Phon. 1994;8(1):1–25.

16. Broadbent J Linking and intrusive r in English. UCL Work Pap Linguist. 1991;3:281–302.

17. Brown G Listening to spoken English. 2nd ed London: Longman; 1990.

18. Newton C, Wells B. Between-word junctures in early multi-word speech. J Child Lang. 2002 5;29(2):275–99. [PubMed: 12109372]

19. Smith A Speech motor development: integrating muscles, movements, and linguistic units. J Commun Disord. 2006 Sep-Oct;39(5):331–49. [PubMed: 16934286]

20. Daniloff R, Moll K. Coarticulation of lip rounding. J Speech Hear Res. 1968 12;11(4):707–21. [PubMed: 5719225]

21. Browman CP, Goldstein L. Articulatory gestures as phonological units. Phonology. 1989;6(02):201–51.

22. McCarthy JJ. Linear order in phonological representation. Linguist Inq. 1989;20:71–99.

23. Shattuck-Hufnagel S, Turk AE. A prosody tutorial for investigators of auditory sentence processing. J Psycholinguist Res. 1996 3;25(2):193–247. [PubMed: 8667297]

24. American Speech-Language-Hearing Association (ASHA): Childhood Apraxia of Speech (technical report). 2007.

25. Shriberg LD, Strand EA, Fourakis M, Jakielski KJ, Hall SD, Karlsson HB, et al. A diagnostic marker to discriminate childhood apraxia of speech from speech delay: IV. The pause marker index. J Speech Lang Hear Res. 2017 4;60(4):S1153–69. [PubMed: 28384662]

26. Thoonen G, Maassen B, Gabreëls F, Schreuder R, de Swart B. Towards a standardised assessment procedure for developmental apraxia of speech. Eur J Disord Commun. 1997;32(1):37–60. [PubMed: 9135712]

27. Marquardt TP, Sussman HM, Snow T, Jacks A. The integrity of the syllable in developmental apraxia of speech. J Commun Disord. 2002 Jan-Feb;35(1):31–49. [PubMed: 11949971]

28. Preston JL, Molfese PJ, Gumkowski N, Sorcinelli A, Harwood V, Irwin JR, et al. Neurophysiology of speech differences in childhood apraxia of speech. Dev Neuropsychol. 2014;39(5):385–403. [PubMed: 25090016]

29. Shriberg LD, Aram DM, Kwiatkowski J. Developmental apraxia of speech: II. Toward a diagnostic marker. J Speech Lang Hear Res. 1997 4;40(2):286–312. [PubMed: 9130200]

30. Shriberg LD, Green JR, Campbell TF, McSweeny JL, Scheer AR. A diagnostic marker for childhood apraxia of speech: the coefficient of variation ratio. Clin Linguist Phon. 2003 Oct-Nov;17(7):575–95. [PubMed: 14608800]

31. Murray E, McCabe P, Heard R, Ballard KJ. Differential diagnosis of children with suspected childhood apraxia of speech. J Speech Lang Hear Res. 2015 2;58(1):43–60. [PubMed: 25480674]

32. Iuzzini-Seigel J, Hogan TP, Green JR. Speech inconsistency in children with childhood apraxia of speech, language impairment, and speech delay: depends on the stimuli. J Speech Lang Hear Res. 2017 5;60(5):1194–210. [PubMed: 28395359]

33. Shriberg LD, Strand EA, Fourakis M, Jakielski KJ, Hall SD, Karlsson HB, et al. A diagnostic marker to discriminate childhood apraxia of speech from speech delay: I. Development and description of the pause marker. J Speech Lang Hear Res. 2017 4;60(4):S1096–117.

34. Shriberg LD, Strand EA, Fourakis M, Jakielski KJ, Hall SD, Karlsson HB, et al. A diagnostic marker to discriminate childhood apraxia of speech from speech delay: II. Validity studies of the pause marker. J Speech Lang Hear Res. 2017 4;60(4):S1118–34. [PubMed: 28384803]

35. Shriberg LD, Strand EA, Fourakis M, Jakielski KJ, Hall SD, Karlsson HB, et al. A diagnostic marker to discriminate childhood apraxia of speech from speech delay: III. Theoretical coherence of the pause marker with speech processing deficits in childhood apraxia of speech. J Speech Lang Hear Res. 2017 4;60(4):S1135–52. [PubMed: 28384751]

36. Barrett C, McCabe P, Masso S, Preston J. Connected Speech Transcription Protocol (CoST-P). Sydney, Australia, 2018.

37. McCabe P, Preston J, Evans P. Ultrasoud or ReST? Which Treatment is better for older children with CAS? A pilot RCT – preliminary data. Poster session presented at: Speech Pathology Australia Annual National Conference; 2016 May 15–18; Perth, Australia.

38. Shriberg LD, Potter NL, Strand EA. Prevalence and phenotype of childhood apraxia of speech in youth with galactosemia. J Speech Lang Hear Res. 2011 4;54(2):487–519. [PubMed: 20966389]

39. Brown L, Sherbenou R, Johnsen S. Test of nonverbal intelligence. 3rd ed Austin (TX): PRO-ED; 1997.

40. Semel E, Wiig EH, Secord WA. Clinical Evaluation of Language Fundamentals fourth edition, Australian Standardised Edition (CELF-4 Australian). Marrickville, Australia, Harcourt Assessment, 2006.

41. Goldman R, Fristoe FM. 2: Test of Articulation. Circle Pines (MN): American Guidance Service. Inc; 2000.

42. Patel R, Connaghan K. Park Play: a picture description task for assessing childhood motor speech disorders. Int J Speech Lang Pathol. 2014 8;16(4):337–43. [PubMed: 24665912]

43. Rose Y, MacWhinney B, Byrne R, Hedlund G, Maddocks K, O'Brien P, et al. Introducing Phon: A software solution for the study of phonological acquisition; in Bamman D, Magnitskaia T, Zaller C(eds): 30th Annual Boston University Conference on Language Development Somerville, MA, Cascadilla Press, 2006, pp 489–500.

44. Munson B, Brinkman KN. The influence of multiple presentations on judgments of children's phonetic accuracy. Am J Speech Lang Pathol. 2004 11;13(4):341–54. [PubMed: 15719900]

45. Byun T, Rose Y: Analyzing clinical phonological data using phon. Semin Speech Lang 2016;37;17:085–105. 10.1055/s-0036-1580741.

46. Baker E Children's independent and relational phonological anlaysis (CHIRPA). Sydney, Australia: General American English; 2016.

47. Shriberg LD, Austin D, Lewis BA, McSweeny JL, Wilson DL. The percentage of consonants correct (PCC) metric: extensions and reliability data. J Speech Lang Hear Res. 1997 8;40(4):708–22. [PubMed: 9263938]

48. Ingram D The measurement of whole-word productions. J Child Lang. 2002 11;29(4):713–33. [PubMed: 12471970]

49. Masso S Word-level analysis of polysyllables. Sydney, Australia: Author; 2016.

50. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977 3;33(1):159–74. [PubMed: 843571]

51. Mehler J, Jusczyk P, Lambertz G, Halsted N, Bertoncini J, Amiel-Tison C. A precursor of language acquisition in young infants. Cognition. 1988 7;29(2):143–78. [PubMed: 3168420]

52. Redford MA, Oh GE. Children's abstraction and generalization of English lexical stress patterns. J Child Lang. 2016 3;43(2):338–65. [PubMed: 26027880]

53. Stoel-Gammon C. Lexical and phonological development at 30 months: A crosslinguistic analysis of word shapes; International Clinical Phonetics and Linguistics Association Conference; 2016 Jun 15–18; Halifax, Canada.

54. Lenneberg EH. Biological foundations of language. New York: Wiley; 1967 10.1080/21548331.1967.11707799.

55. Stoel-Gammon C, Cooper JA. Patterns of early lexical and phonological development. J Child Lang. 1984 6;11(2):247–71. [PubMed: 6746776]

56. Velleman SL. Lexical and phonological development in children with childhood apraxia of speech —a commentary on Stoel-Gammon's 'Relationships between lexical and phonological development in young children'. J Child Lang. 2011 1;38(1):82–6. [PubMed: 20950498]

57. Murray E, McCabe P, Ballard KJ: A systematic review of treatment outcomes for children with childhood apraxia of speech. Am J Speech Lang Pathol / ASHA 2014;23:486 10.1044/2014_AJSLP-13-0035.

58. Vuolo J, Goffman L. An exploratory study of the influence of load and practice on segmental and articulatory variability in children with speech sound disorders. Clin Linguist Phon. 2017;31(5):331–50. [PubMed: 27960554]

59. Lewis BA, Freebairn LA, Hansen AJ, Iyengar SK, Taylor HG. School-age follow-up of children with childhood apraxia of speech. Lang Speech Hear Serv Sch. 2004 4;35(2):122–40. [PubMed: 15191325]

60. Marquardt TP, Jacks A, Davis BL. Token-to-token variability in developmental apraxia of speech: three longitudinal case studies. Clin Linguist Phon. 2004 3;18(2):127–44. [PubMed: 15086134]

61. Hodson BW, Paden EP. Phonological processes which characterize unintelligible and intelligible speech in early childhood. J Speech Hear Disord. 1981;46(4):369–73.

62. McCabe P, Rosenthal JB, McLeod S. Features of developmental dyspraxia in the general speech-impaired population? Clin Linguist Phon. 1998;12(2):105–26. [PubMed: 21434785]

63. Gerken L The metrical basis for children's subjectless sentences. J Mem Lang. 1991;30(4):431–51.

64. Velleman SL, Shriberg LD. Metrical analysis of the speech of children with suspected developmental apraxia of speech. J Speech Lang Hear Res. 1999 12;42(6):1444–60. [PubMed: 10599626]

65. Chen NF, Slifka JL, Stevens KN. Vowel nasalization in American English: acoustic variability due to phonetic context. Speech Commun. 2007:905–18.

66. Newton C Between-word processes in children with speech difficulties: insights from a usage-based approach to phonology. Clin Linguist Phon. 2012 8;26(8):712–27 [PubMed: 22774930]
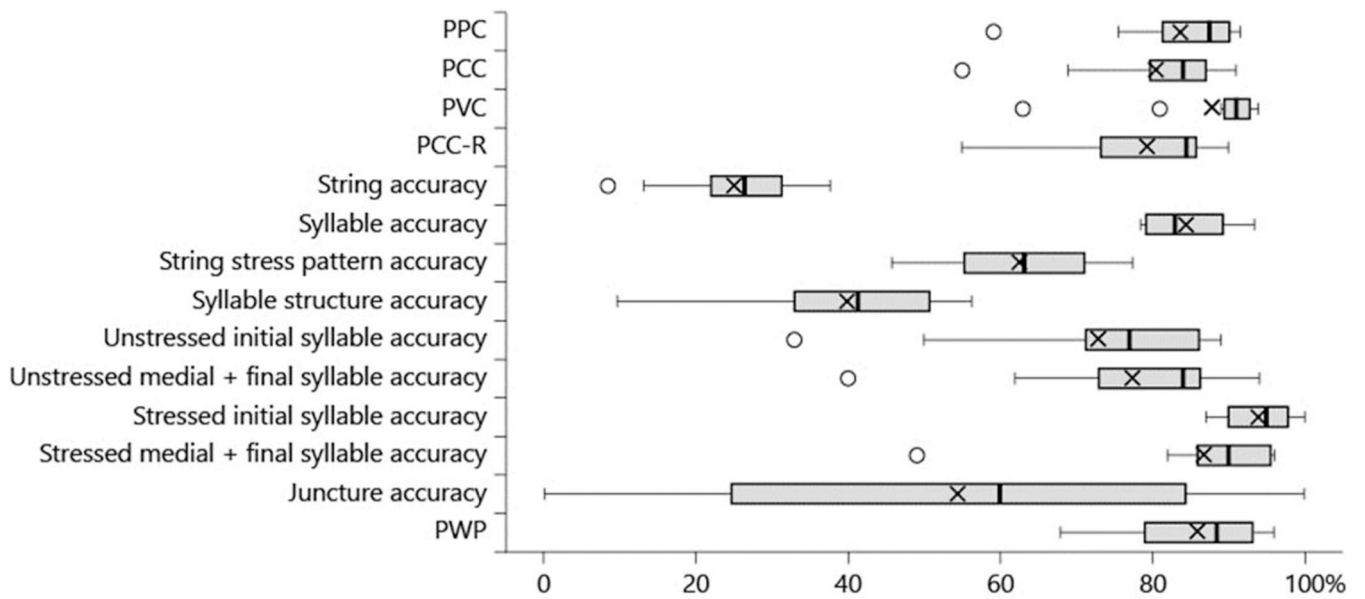
Fig. 1.
Relational analyses captured following transcription using the CoST-P (*n* = 10). PPC, percentage of phonemes correct; PCC, percentage of consonant correct; PVC, percentage of vowels correct; PCC-R, percentage of consonant correct-revised; PWP, proportion of whole-word proximity. Please also note that × reflects the group mean, and ○ represents outliers.

Table 1.

Participant demographic information

	Participant											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Age	9;11	9;1	6;0	8;3	10;7	11;0	9;10	12;9	9;0	6;4	6;6	6;5
Sex	M	M	F	M	M	M	M	F	F	M	M	M
GFTA-2	79	103	57	70	41	62	81	73	40	62	–	–
PPVT-4	114	98	105	103	113	97	93	96	85	104	96	85

Age is given as years;months. GFTA-2, Goldman-Fristoe Test of Articulation – second edition (standard score); PPVT-4, Peabody Picture Vocabulary Test – fourth edition (standard score).

Table 2.

CoST-P iterations

#	Changes	Testing procedure
1	This initial version included five broad recommendations and 20 sub-recommendations structured in list format	Further research regarding unique connected speech characteristics
2	Additional recommendations added; this version included six broad recommendations, 30 sub-recommendations, and 15 additional recommendations; of particular note was the consideration of linguistically/contextually appropriate speech alterations (e.g., reduced forms such as "dunno") and the addition of examples to explain recommendations	Areas for further change determined by the research team (through consensus)
3	Table formatting and contingencies for sub-recommendations	Suitability of each structure determined by consensus
4	Decision to transcribe words as connected where they are naturally joined in connected speech; spaces to be used to indicate appropriate inter-word pause, and pause markers to be used to indicate inappropriate segmentation	Transcription and analysis assessed using test file
	Decision to transcribe dysfluencies in all tiers (to eliminate impact on relational phoneme analysis) and to note the number of dysfluencies in each utterance (e.g., "$ $")	Transcription and analysis assessed using test file
	Decision to transcribe stereotyped speech or sing-song-style speech and include a note below indicating the speech type	Transcription and analysis assessed using test file
5	Decision to use audio (rather than video) recordings to eliminate effects of auditory illusions (e.g., McGurk-like effects) which may occur due to inadvertent audio/video misalignment or visual dominance in information processing [43]	Audio was isolated from files; transcription and analysis assessed using test file
	Decision for unintelligible phonemes or utterances to be transcribed with use of consistent codes; for example, if the manner of consonants is audible, but the consonant place is unidentifiable, transcriber should use generic marker	Transcription and analysis assessed using test file
6	Decision to use diacritics to transcribe allophonic variation only where it is linguistically inappropriate (e.g., /æ/ in /mæn/ does not require a hyper-nasal diacritic as it is appropriate for the vowel preceding a nasal consonant to be nasalised)	Decision made by author consensus; transcription and analysis assessed using test file
7	Decision to transcribe primary and secondary stress markers in both the children's target productions and actual productions	Transcription and analysis assessed using test file
8	Recommendations to ensure an error is accounted for only once in computer-based analysis	Assessed with use of a test file; analyses run in Phon and compared to manual analysis
9	CoST-P finalised and implemented for all test transcriptions; protocol fidelity conducted on this final version of the CoST-P	

Table 3.

Independent speech analyses captured following transcription using the CoST-P ($n = 12$)

Feature	Range	Mean	SD
Segmental features			
Total No. phonemes	353–2,170	1,028	491
Phoneme inventory	25–41	37	4
Consonant inventory	12–26	22	5
Vowel inventory	13–19	15	1
Syllables, words, and strings			
No. strings	83–185	108	27
No. syllables	199–598	382	127
No. syllables/string	1.2–5.9	3.7	1.2
No. words	235–476	344	82
No. polysyllabic words (3+ syllables)	7–23	14	6
Suprasegmental features			
No. primary stressed syllables	141–389	258	77
No. secondary stressed syllables	7–59	33	15
No. unstressed syllables	42–158	28	39
Instances of intra-word segregation	0–28	4	8
Instances of inter-word segregation	5–67	29	21
Other			
Instances of juncture	0–15	7	5
No. dysfluencies	1–23	8	7
No. voice diacritics	0–24	8	8

Table 4.

Phonological process analyses (captured following transcription using the CoST-P) ($n = 10$)

Phonological process	Error occurrence			Percent of total opportunities		
	range	mean	SD	range	mean	SD
Segmental processes						
Stopping	2–49	17	15	2–30	10	9
Deaffrication	0–5	1.5	2	0–100	32	39
Voicing	1–16	6	5	<1–9	3	3
Gliding	1–19	9	6	1–28	14	9
Structural processes						
Deletion (onsets)	4–41	17	11	1–10	5	3
Deletion (nucleus)	0–28	6	8	0–9	2	2
Deletion (coda)	9–81	32	22	3–58	17	15
Deletion (total)	17–129	55	34	1–18	6	5
Cluster reduction (onset)	0–7	3	3	0–39	14	13
Cluster reduction (coda)	3–16	8	4	16–44	28	9
Cluster reduction (total)	3–20	11	5	7–39	23	9

Table 5.

Word-by-word transcription versus CoST-P transcription example

	Word-by-word transcription
Orthography	Mum I want to colour in two elephants with a hat on. I want to colour all them orange.
IPA Target	/mʌm aɪ wɒnt tə kʌlə ɪn tu ɛləfənts wɪð ə hæt ɒn/ /aɪ wɒnt tə kʌlə al ðɛm ɒ.ɹəndʒ/
IPA Actual	/mʌm aɪ̃ wɒnt tə kʌ.lə ɪn tu ɛləfənts wɪd ə hætʰ ɒn/ /aɪ wɒnt tə kʌlə al dɛm ɒ̃.ɹənd/
Notes	2× instances of inappropriate hyper-nasality present (/aɪ̃/ and /ɒ̃.ɹənd/)
	1× instance of inappropriate intra-word segregation present (/kʌ.lə/)
	1× instance of inappropriate aspiration present (/hætʰ/)

	CoST-P transcription
Orthography	Mum I want to colour in two elephants with a hat on. I want to colour all them orange.
IPA Target	/mʌmaɪwɒnəkʌlə.ɹɪntuwɛləfəntswɪðəhætɒn/ /aɪwɒnəkʌlə.ɹaldɛmɒ.ɹəndʒ/
IPA Actual	/mʌmaɪwɒnə^kʌ.lə.ɹɪntuwɛləfəntswɪdəhætɒn/ aɪŵɒnəkʌlə.ɹaldɛmɒ.ɹənd/
Notes	2× instances of inappropriate inter-word pauses (/wɒnə^kʌ.lə/ and /aɪŵɒnə/)
	3× instances of juncture (2× /ɹ/, 1× /w/) (/kʌ.lə.ɹɪn/, /tuwɛləfənts/, and /kʌlə.ɹal/)
	1× instance of intra-word segregation (/kʌ.lə/)