This is a repository copy of *The impact of the speech systems approach on intelligibility for children with cerebral palsy: a secondary analysis*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/205689/

Version: Published Version

Article:

Check for updates

The impact of the Speech Systems Approach on intelligibility for children with cerebral palsy: a secondary analysis

Lindsay Pennington, Stuart Cunningham, Shaun Hiu, Ghada Khattab and Vicky Ryan

The impact of the Speech Systems Approach on intelligibility for children with cerebral palsy: a secondary analysis

Lindsay Pennington,[1,2*] Stuart Cunningham,[3]
Shaun Hiu,[1] Ghada Khattab[4] and Vicky Ryan[1]

[1]Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK
[2]Great North Children's Hospital, Newcastle upon Tyne Hospitals NHS Foundation
  Trust, Newcastle upon Tyne, UK
[3]Division of Human Communication Sciences, Health Sciences School, University of
  Sheffield, Sheffield, UK
[4]School of Education, Communication and Language Sciences, Newcastle University,
  Newcastle upon Tyne, UK

[*]Corresponding author

This report should be referenced as follows:

Pennington L, Cunningham S, Hiu S, Khattab G and Ryan V. The impact of the Speech Systems Approach on intelligibility for children with cerebral palsy: a secondary analysis. *Efficacy Mech Eval* 2023;**10**(4). https://doi.org/10.3310/EDAE1779

Efficacy and Mechanism Evaluation

> **Criteria for inclusion in the *Efficacy and Mechanism Evaluation* journal**
> Reports are published in *Efficacy and Mechanism Evaluation* (EME) if (1) they have resulted from work for the EME programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

EME programme

The Efficacy and Mechanism Evaluation (EME) programme funds ambitious studies evaluating interventions that have the potential to make a step-change in the promotion of health, treatment of disease and improvement of rehabilitation or long-term care. Within these studies, EME supports research to improve the understanding of the mechanisms of both diseases and treatments.

The programme supports translational research into a wide range of new or repurposed interventions. These may include diagnostic or prognostic tests and decision-making tools, therapeutics or psychological treatments, medical devices, and public health initiatives delivered in the NHS.

The EME programme supports clinical trials and studies with other robust designs, which test the efficacy of interventions, and which may use clinical or well-validated surrogate outcomes. It only supports studies in man and where there is adequate proof of concept. The programme encourages hypothesis-driven mechanistic studies, integrated within the efficacy study, that explore the mechanisms of action of the intervention or the disease, the cause of differing responses, or improve the understanding of adverse effects. It funds similar mechanistic studies linked to studies funded by any NIHR programme.

The EME programme is funded by the Medical Research Council (MRC) and the National Institute for Health and Care Research (NIHR), with contributions from the Chief Scientist Office (CSO) in Scotland and National Institute for Social Care and Health Research (NISCHR) in Wales and the Health and Social Care Research and Development (HSC R&D), Public Health Agency in Northern Ireland.

This report

Abstract

The impact of the Speech Systems Approach on intelligibility for children with cerebral palsy: a secondary analysis

Lindsay Pennington[1,2*] Stuart Cunningham,[3] Shaun Hiu,[1] Ghada Khattab[4] and Vicky Ryan[1]

[1]Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK
[2]Great North Children's Hospital, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK
[3]Division of Human Communication Sciences, Health Sciences School, University of Sheffield, Sheffield, UK
[4]School of Education, Communication and Language Sciences, Newcastle University, Newcastle upon Tyne, UK

*Corresponding author lindsay.pennington@ncl.ac.uk

Background: The motor speech disorder, dysarthria, is common in cerebral palsy. The Speech Systems Approach therapy programme, which focuses on controlling breath supply and speech rate, has increased children's intelligibility.

Objective: To ascertain if increased intelligibility is due to better differentiation of the articulation of individual consonants in words spoken in isolation and in connected speech.

Design: Secondary analysis.

Setting: University.

Participants: Forty-two children with cerebral palsy and dysarthria aged 5–18 years, Gross Motor Function Classification System I–V.

Intervention: The Speech Systems Approach is a motor learning therapy delivered to individuals by a speech and language therapist in 40-minute sessions, three times per week for 6 weeks. Intervention focuses on production of a strong, clear voice and speaking at a steady rate. Practice changes from single words to increasingly longer utterances in tasks with increasing cognitive load.

Main outcome measures: Unfamiliar listeners' identification of singleton consonants (e.g. nap) and clusters of consonants (e.g. stair, end) at the start and end of words when hearing single words in forced choice tasks and connected speech in free transcription tasks. Acoustic measures of sound intensity and duration.

Data sources: Data collected at 1-week pre- and 1-week post-therapy from three studies: two interrupted time series design, one feasibility randomised controlled trial.

Results: Word initial and word final singleton consonants and consonant clusters were better identified post-therapy. The extent of improvement differed across word initial and word final singleton consonant subtypes. Improvement was greater for single words than connected speech. Change in sound

identification varied across children, particularly in connected speech. Sound intensity and duration increases also were inconsistent.

Limitations: The small sample size did not allow for analysis of cerebral palsy type. Acoustic data were not available for all children, limiting the strength of conclusions that can be drawn. The different but phonetically balanced word lists, used in the original research, created variability in single words spoken across recordings analysed. Low frequencies of plosives, fricatives and affricates necessitated their combination for analysis preventing investigation of the effect of specific consonants. Connected speech was spontaneous, again creating variability within the data analysed. The estimated effects of therapy may therefore be partially explained by differences in the spoken language elicited.

Conclusions: The Speech Systems Approach helped children generate greater breath supply and a steady rate, leading to increased intensity and duration of consonant sounds in single words, thereby aiding their identification by listeners. Transfer of the motor behaviour to connected speech was inconsistent.

Future work: Refining the Speech Systems Approach to focus on connected speech early in the intervention. Personalisation of cues according to perceptual and acoustic speech measures. Creation of a battery of measures that can be repeated across children and multiple recordings.

Study registration: This trial is registered as Research Registry 6117.

Contents

CONTENTS

List of tables

List of figures

Glossary

Worster Drought   A type of cerebral palsy that affects the muscles around the mouth and throat, causing problems with swallowing, feeding, speaking.

Phoneme   The smallest sound units which constitute the building blocks of words in a language.

Vowel   A sound that is made by allowing breath to flow out of the mouth, without closing any part of the mouth or throat.

Consonant   A sound that is made by blocking/constraining air from flowing out of the mouth with the teeth, tongue, lips or palate.

Voicing   Referring to the vibration of the vocal fold during speech. Voiced sounds involve vibration of the vocal folds. Voiceless sounds are made when the vocal folds are abducted.

Place   Position in the vocal tract where articulation occurs.

Manner   Nature of constriction of the vocal tract.

Sonorant   Sonorants are produced when there is free airflow through the vocal tract and the voiced sounds can resonate freely.

Nasal   Consonants that are made when the air is stopped in the oral cavity and flows through the nasal cavity.

Approximant   Consonants that are made by bringing two articulators close together without them touching.

Obstruent   Sounds that are made by constricting the air flow.

Plosive   Consonants that involve complete stoppage then sudden release of airflow.

Fricative   Consonants that involve partial closure of the vocal tract and the creation of turbulence of the airflow.

Affricate   Consonants that are made by stopping the airflow followed by frication.

Labial   Consonants produced involving the lip(s).

Coronal   Consonants made with the anterior part of the tongue.

Dorsal   Consonants produced with the back of the tongue.

Labiodental   Consonants made when the lower lip is in contact with the upper incisors.

Alveolar   Consonants made when the tongue tip is in contact with the alveolar ridge (hard palate behind teeth).

Velar   Consonants made when the back of the tongue is in contact with the soft palate (velum).

Glottal   Consonants articulated with the glottis.

Singleton consonant   Consonants that appear alone at the start or end of a syllable.

Cluster   Groups of consonants appearing at the start or end of a syllable that are not separated by a vowel.

List of abbreviations

CI	confidence interval	MLU	mean length of utterance
CP	cerebral palsy	OR	odds ratio
CS	connected speech	PCA	principal component analysis
CSIM	Children's Speech Intelligibility Measure	PISW	percentage identification of single word
dB	decibel	SE	standard error
GLMM	generalised linear mixed model	SLT	speech and language therapy
GMFCS	gross motor function classification system	SW	single word
IPA	international phonetic alphabet	VOT	voice onset time
LR	likelihood ratio	VPC	variance partition coefficient
LSVT	Lee Silverman Voice Therapy	VPM	voice, place, manner

Plain language summary

Some children with cerebral palsy have speech that sounds weak, slurred and difficult to understand, which seriously impacts their social life and education. We developed a therapy programme to help children control their breathing and how fast they speak. Having more breath should make children's voices stronger. Speaking at a steady rate should give enough time for children to move their jaw, tongue and lips to produce each sound more precisely. Children's speech was easier to understand after the therapy. This study aimed to find out if the therapy worked by helping children to say consonant sounds more clearly.

We used recordings made in previous research to work out which consonants listeners heard correctly. We also looked at waveforms, which showed children's speech as moving pictures, to find out how speech changed.

After therapy, when children spoke in single words, listeners heard almost all types of consonant sounds at the start and end of words more clearly. No particular type of consonant sounds, such as 's' in 'so' or 't' in 'tar', led to better speech clarity. Waveforms showed that some children produced stronger speech sounds, some slowed their speech, and some did both.

Listeners heard some children more clearly after therapy when they spoke in phrases, but found others more difficult to understand. Few consonants were easier to understand after therapy. We saw no clear patterns of change in speech waveforms.

Overall, children produced stronger, more precise speech in single words, but not all transferred this skill to speaking in phrases. Children differed in how they achieved clearer speech.

We used the findings to refine the therapy to focus on phrases early in the programme and to personalise instructions to children's individual speech patterns. We will use waveforms to find where children have most difficulty and to measure improvement.

Scientific summary

Background and introduction

The motor disorders of cerebral palsy (CP) often affect breath control and speech production, causing the speech disorder dysarthria. Dysarthria in CP typically affects all speech systems: respiration, phonation, resonance, prosody and articulation. Respiration is often shallow and lacks co-ordination with phonation, generating weak or inconsistent subglottal pressure. Vocal folds may vibrate slowly and irregularly; air may leak through the folds when they should be adducted, reducing the intraoral pressure and weakening the sound source. The velum may rise slowly or fail to close off the nasal passage during speech. The movements of the articulators (jaw, tongue and lips) may be slow and imprecise. They may also be weak, reducing children's ability to constrict the vocal tract for consonant sounds. The combined effect of these limitations is that children often speak in short phrases, with inappropriate phrasing or rushes of speech if children run out of air. Their voice may sound weak, breathy and sometimes harsh. Speech is often slow with reduced melodic intonation and children may have a restricted range of consonants that they can produce clearly.

Intervention focussing on breath support and speech rate is expected to aid with the co-ordination of three phases of speech production, mainly initiation, phonation and articulation. Greater breath supply and increased air pressure during exhalation should increase subglottal air pressure bringing firmer contact of the vocal folds during phonation to generate a stronger vocal note/sound source. The improved audibility and potential for greater intraoral air pressure arising from this will also help compensate for any weak closures of articulators and reduce 'leakage' of air during speech. A steady speech rate should allow children to move with precision from one articulatory place and manner to another. Thus, as a result of changes in breath supply and rate, phonemes should be acoustically differentiated and listeners better able to perceive the sounds that children are articulating (increased *phonetic intelligibility*). The Speech Systems Approach has been developed to focus on breath control and speech rate and has led to improvements in the intelligibility of children's speech which have been maintained for up to 12 weeks without further intervention. What is not yet known is whether there is a differential effect on types of speech sounds, or whether this effect is moderated by CP type or severity of impairment. Establishing this can then help with further individualisation of therapy and further gains in intelligibility.

Aims

The aim of the study was to ascertain if therapy focusing on breath supply and speech rate is associated with increased differentiation of the articulation of individual phonemes, enabling listeners to better identify individual phonemes in words spoken in isolation and in connected speech (CS).

Methods

Design
The study was a secondary analysis of previously collected data from two phase II studies using an interrupted time series design and one feasibility randomised controlled trial of the Speech Systems Approach.

Participants
Forty-two children and young people aged 5–18 years, who had a diagnosis of CP by a medical practitioner and moderate to severe dysarthria as assessed by their local speech language therapist,

received the Speech Systems Approach in the original studies. Participants were excluded from the studies if they had hearing impairments >50 decibel HL, visual impairments that were not correctable with glasses, or were unable to follow simple verbal instructions. All 42 participants were included in this secondary analysis.

Intervention

Children received individual therapy following the Speech Systems Approach from a registered speech and language therapist three times a week for 6 weeks. Sessions lasted approximately 40 minutes. In two of the original studies the sessions were provided face-to-face, the third study session took place remotely using video conferencing software. In the first session the therapist tried several cues to find the best that elicited a strong, clear voice in an open vowel (ah). Cues included 'strong, big, loud' and a combination of these. For children who had difficulty initiating movement cues were 'nice and easy' or 'smooth'. Once the most appropriate cue had been found for the child, that cue was used to elicit open vowels on command. Practice of the target voice then followed a hierarchy of increasing length of utterance and cognitive load. Children first practiced in single words (SWs), then short phrases and finally longer CS. At each stage they started with repetition, moved to picture naming and description, and then questions and answers and games. Children had to use their target voice 8/10 attempts to move to the next level in the hierarchy. Children were provided with knowledge of results on how their voice sounded and were encouraged to use bio feedback ('How did that feel?'). Feedback was given frequently in the acquisition phase at each level of the hierarchy and then faded to aid retention. Sessions followed a set sequence for practice of the target voice: (1) 10 open vowels; (2) three repetitions of 10 self-selected phrases that children use in daily life; (3) 70–80 words and phrases from the speech task hierarchy; and (4) 10–20 utterances randomly selected from the three preceding tasks.

Procedure

Data were collected at 6- and 1-week pre-therapy and 1, 6- and 12-weeks post-therapy. The original studies suggested that gains in intelligibility were observed immediately after therapy (1 week) and were maintained at 6 and 12 weeks. For this study we analysed data from 1-week pre- and 1-week post-therapy. At each time point, participants' speech was recorded on two separate days and elicited through two tasks. SWs were elicited using the Children's Speech Intelligibility Measure (CSIM) which contains 200 lists of 50 words. The CSIM is a forced choice word recognition task. Listeners heard each word and selected the target word from a list of 12 phonetically similar words. CS was elicited by asking the participants to describe complex pictures and answer questions. The recordings were transcribed live by an expert speech and language therapist and then checked with the child, to create a gold standard transcription of target words. Up to 60 seconds of CS was presented to listeners in phrases separated by pauses of at least three seconds. The CS recognition task was open choice; listeners heard the recordings of CS and wrote down the words they perceived the child to say. Listeners were native speakers of English, aged 18–55 years, and had no hearing difficulties or regular experience of interacting with people with disabilities or speech disorders. In each of the original studies, listeners were randomly allocated three speech recordings, with the limitation that they did not hear the same child more than once. Each recording was heard by three listeners.

Outcomes

Perceptual data

Each word produced by children in the SW and CS tasks was categorised by the single consonants and the clusters of consonants at the start and end of words. Consonants were categorised according to their voicing (voiced = vocal fold vibration; voiceless = vocal folds open), place of articulation (bilabial, labiodental, dental, alveolar, post-alveolar, velar, and glottal), and manner of articulation (plosives, fricatives, affricates, nasals and approximants). The words perceived by listeners were categorised in the same way. A review of the data showed that some manners and places of articulation rarely appeared in the SW lists. We therefore combined manners of articulation into obstruents, which demand constriction/closure of the vocal tract (plosives, fricative and affricates), and sonorants, which allow air

to flow out of the mouth or nose (approximants and nasals) and places of articulation to labial (bilabial, labiodental), coronal (dental, alveolar, post-alveolar) and dorsal (velar). Consonants were then further categorised according to a combination of their voicing, place, and manner characteristics (e.g. voiced labial obstruent). This was done owing to dependencies between voice and manner such that consonants with a sonorant manner are never voiceless. Because of this, we were not able to separate the main effects of voice and manner from each other if they were added as individual explanatory variables in modelling.

The perceptual data were multilevel such that children were level 2 units and each target word-and-listener combination was a level 1 unit. The primary outcome for the perceptual analysis was the identification of words and segments within them (binary outcomes). Words and sounds were identified if there was a match between the target and perceived word/consonant/cluster/voice/place/manner. Secondary outcomes include the percentage identification measures.

We visually inspected radar plots profiling performance on six percentage identification measures comprising the voice, place and manner of word initial and word final singleton consonants at pre- and post-therapy.

Acoustic data
Recordings of 24 out of the 42 children were available pre- and post-therapy and were automatically transcribed and segmented, and manually corrected. Measures of intensity (dB), duration (ms) and speech rate (seconds per syllable) were made at word and segment level. Changes in intensity and/or speech rate were monitored post-therapy and examined as a function of consonant manner, intelligibility, and groups of children based on the perceptual analyses.

Data analysis

Perceptual data
We used generalised linear mixed modelling (GLMM) with a logit link function and random effect of child to examine the effect of the Speech Systems Approach intervention (post-therapy vs. pre-therapy) on identification of word initial and word final singleton consonants and consonant clusters and whether there was evidence that certain types of consonants benefitted more from the therapy than others. We visually inspected radar plots of change in listeners' identification of consonants to group children with similar profiles.

Acoustic data
Acoustic measurements made on SWs and CS, with mean differences and 95% confidence intervals (CIs) calculated between the post- and pre-therapy. Changes in the acoustic variables were then compared to changes in perceptual variables such as single and connected word intelligibility. These comparisons were made for each individual child and groupings derived from the radar plots from the perceptual data analysis.

Results

Perceptual analysis
Using GLMM, we found evidence that the odds of word initial and word final singleton consonants being identified by listeners increased after therapy in both the SW [odds ratio (95% CI) = 1.54 (1.44 to 1.65) and 1.61 (1.51 to 1.73) respectively; all $p < 0.01$] and CS analyses [1.26 (1.15 to 1.39) and 1.27 (1.15 to 1.41) respectively; all $p < 0.01$], even after adjusting for potential confounders. There was also evidence for the effect of therapy on the identification of word initial consonant clusters and word final consonant clusters in the SW analyses [1.84 (1.60 to 2.12) and 1.42 (1.15 to 1.75) respectively; all $p < 0.01$]. Identification of consonant clusters was not examined in the CS data due to the infrequent use of clusters. Additionally, we found evidence of heterogeneity in the effect of therapy on word initial and

word final singleton consonant identification between the subtypes of consonants, categorised according to their voicing, place, and manner, in both SW and CS analyses. Nearly all subtypes of consonants, as either word initial or word final singleton consonants, showed an improvement in the probability of being identified by a listener after therapy in the SW analyses, but only about a handful showed an improvement in the CS analyses.

We identified six groups of children in the SW data and seven groups of children in the CS data. Each group categorised children according to their relative standing at pre-therapy, the direction and magnitude of change in percentage identification measures, and additionally which combination of the six measures displayed a change.

Acoustic analysis

Most of the children whose data were analysed acoustically produced slower speech post-therapy, regardless of gains in intelligibility. There was individual variability in the degree of change in speech rate, with a tendency for greater decreases in the speech rate of children whose initial performance was low. Most children also produced words with higher mean intensity post-therapy regardless of gains in intelligibility. However, the majority of the children with an increase in maximum intensity in initial and final word position had improved intelligibility post-therapy. This was particularly the case for the obstruent category. There was variation in changes to intensity across clusters, word position and manner categories, with more modest increases in the intensity of sonorant sounds pre- and post-therapy compared to obstruents. There were also more variable changes in speech rate, intensity and their relationship with intelligibility in CS compared with SWs.

Conclusions and recommendations

The Speech Systems Approach, which focusses on breath control and speech rate, improved the intelligibility of SWs. Our previous research suggests that gains were maintained for up to 12 weeks following the intensive burst of therapy. This type of intervention is now recommended by National Institute for Health and Care Excellence (NICE) to help address the interaction challenges faced by children with CP. Acoustic analysis of a subset of the data shows change for individuals in intensity and/or duration of individual speech segments. Findings suggest that increased intelligibility was achieved through a stronger vocal signal and allowing children to articulate individual sounds with greater precision. Changes in CS were more modest, suggesting that adaptations to the approach are warranted. The marked individual differences suggest a varying response to therapy between children but also within children across speech segments. Personalisation of the intervention, with cues adapted to the child's performance, should be investigated. Acoustic analysis of speech during the intervention could aid personalisation, showing how speech changes in response to individual cues. Practice should move to CS quickly. Intelligibility, through listener identification of words and constituent sounds, should continue to be assessed in SWs and CS to investigate the impact of utterance length. Two matched lists of high frequency SWs could be used to facilitate comparison across children and time. Using one list at 6-weeks pre-therapy and immediately post-therapy, and a second list immediately pre-therapy and at follow-up would minimise learning effects. Acoustic change should be measured to help understand how change is achieved by individuals.

Study registration

This trial is registered as Research Registry 6117.

Funding

Chapter 1　Introduction and aims

Cerebral palsy (CP) is an umbrella term for disorders of movement and posture arising from damage to the developing brain.[1] It is a life-long condition and the most common cause of motor disorder in childhood, affecting 2/1000 live births.[2] Around one quarter of children with CP have motor disorders that affect control of the respiratory and vocal tracts, causing the speech disorder 'dysarthria'.[3,4] The speech of children with dysarthria can sound weak, slow, slurred, effortful and monotone and is often difficult to understand.[5]

The impact of dysarthria can be severe and long lasting. Children who have communication difficulties, as part of their CP, rate their quality of life lower, especially in the domain of social relationships, than their peers without CP *and* those with CP who speak clearly.[6] They are four times as likely to have mental health problems than their peers.[7] They are also at significant risk of poorer educational outcomes and, as adults, they are more likely to be unemployed and single.[8] Children's communication difficulties have an impact across the family. Parents of children with CP are five times more likely to have clinically significant levels of stress; those with children who also have communication difficulties are ten times more likely.[9]

Children with dysarthria receive National Health Service (NHS) speech and language therapy (SLT) services. SLT aims to improve children's intelligibility – the extent to which their message is perceived correctly by a listener – so they can communicate effectively and participate in family, social, and educational activities like their peers. Increased communicative participation may also have secondary impacts on children's psychological wellbeing, and ultimately their parents' mental health.

The importance of communication and developing effective therapy interventions was highlighted in a 2015 James Lind Alliance research priority setting exercise.[10] Parents of children with neurodisability and professionals working with them identified the top two research priorities to be (1) timing and dosage of therapies and (2) improving communication – selecting the most appropriate communication strategies and helping staff and carers to promote effective communication.

Phase II studies have suggested that SLT interventions focussing on breath support and speech rate may increase the intelligibility of children with dysarthria and positively impact on children's communicative participation and self-confidence.[11] These interventions seek to help children to maximise their intelligibility by increasing control of the volitional movements in speech production, not cure their underlying speech disorder. National Institute for Health and Care Excellence (NICE) acknowledged the impact of communication impairments for children with CP and recommended that those whose speech intelligibility is limited by their motor disorder should be offered this type of therapy, despite the current low levels of evidence.[12] Nevertheless, national surveys of United Kingdom (UK) SLT provision show that although many therapists give advice on the principles underpinning this type of therapy, most do not follow defined intervention programmes.[13] NHS SLT provision to children also varies widely in terms of amount and content.[14,15] It is likely that many children who have CP and dysarthria are not yet receiving direct intervention focussing on breath control and speech rate.

What is the gap this research will address?

We previously observed modest changes in overall voice quality, speech rate, pitch and intensity (perceived by listeners as loudness), which were insufficient to account for changes in intelligibility.[16] This study will examine if the increased intelligibility observed in some children following intervention focussing on breath support and speech rate is associated with the ability to differentiate individual sounds within words.

The intervention is underpinned by the source-filter model of speech production and the impairments of dysarthria in CP, as described below.

The source filter model of speech production

Speech is one of the most complex of human motor behaviours. It depends on rapid, millimetre and millisecond precise movements within and across muscles from the diaphragm to the lips.

The source of vocal sound is created when pressurised exhaled air is forced from the lungs and pushes through adducted vocal folds causing them to vibrate (phonate). Modulation of the tightness of lateral and anterior-posterior adduction balanced against the expired air pressure is responsible for variations in loudness/intensity of voice and the rising and falling intonation of speech. This source vocal note is then filtered, that is, further modified by various partial or complete constrictions along the vocal tract, to create individual vowel and consonant sounds (phonemes). Constrictions may involve all or any so-called place of articulation – pharynx, velum, tongue, or lips.[17,18]

Consonants are categorised by their voicing, the place in the vocal tract where they are made and their manner of articulation:

- Most consonants in English have voiced and voiceless pairs. Voiced consonants are produced when the vocal folds vibrate. Voiceless consonants are produced when the vocal folds are abducted and air flows unimpeded through the larynx.
- Place refers to where the constriction(s) of the vocal tract occurs:

  ◦ Labial sounds involve movement of the lips and include bilabial sounds and labiodentals where the lower lip is in contact with the upper incisors.
  ◦ Coronal consonants are made when the tongue tip or blade approaches or is in contact with the upper incisors (interdental), alveolar ridge (alveolar) behind the top teeth or the anterior hard palate (post-alveolar).
  ◦ Dorsal sounds involve movement of the base of the tongue towards the posterior hard palate (palatal) or the soft palate (velar).
  ◦ In English, there is one glottal consonant *h*, which is made by air being expelled through the open vocal folds and appears before vowels. A second glottal consonant, the glottal stop, is a possible realisation of the *t* sound in English and is made by partial or complete closure at the glottis.

- Manner refers to the nature of the constriction:

  ◦ Obstruents involve an obstruction of the air flow. There is complete stoppage then sudden release of airflow in stops/plosive sounds such as *p*, *b*, *t*, *d*, *k*, *g*; partial closure creating turbulence in fricatives such as *f*, *s*, *sh*; and stoppage closely followed by frication in affricates *ch* and *dg*.
  ◦ Sonorants are produced when there is free airflow through the vocal tract and the voiced sounds can resonate freely. Sonorants include vowels, approximants that involve articulators approaching each other but not touching such as *w*, *l*, *r*, *j*, and nasals *m* and *n* that are made when the oral cavity is closed and air flows through the nasal cavity. Vowel sounds have no constriction but are differentiated by the fine positioning of the tongue to create varying resonances within the oral cavity.

Table 1 shows the voicing, place and manner of each consonant in English and their symbols using the International Phonetic Alphabet (IPA), with examples of their production in single words (SWs). Herein speech sounds will be denoted in the text using their IPA symbol.

Intelligible speech thus depends on careful equilibrium of all the movements and timing of constrictions involved in producing and modifying a vocal note. The speaker must generate and maintain sufficient

TABLE 1 Voice, place and manner of articulation of English consonants

Manner group	Manner	Voicing	Place							
			Labial		Interdental	Coronal		Dorsal		Glottal
			Bilabial	Labio-dental		Alveolar	Post-alveolar	Palatal	Velar	
Obstruent	Plosive	Voiced	b						g	
		Voiceless	p						k	
	Fricative	Voiced		v	ð	z	ʒ			
		Voiceless		f	θ	s	ʃ			h
	Affricate	Voiced					dʒ			
		Voiceless					tʃ			
Sonorant	Nasal	Voiced	m			n			ŋ	
		Voiceless								
	Approximant	Voiced	w				ɹ	j		
		Voiceless								
	Lateral	Voiced				l				
		Voiceless								

air pressure subglottally to drive the vocal cords; in turn they require a fine balance in contraction to resist the subglottal airflow to produce voice and create variations in voice loudness/intensity and pitch/intonation. Clear speech also demands rapid, precise and consistent control of articulatory movements across the vocal tract to signal target sounds. For example, the final phonemes in *kin, kid, kit,* and *kiss* share the same place of articulation, with the tongue tip in contact with or in close proximity to the alveolar ridge. But they vary in the type of constriction (complete vs. narrow: *kid, kit* vs. *kiss*), the presence or absence of phonation (*kin, kid* vs. *kit, kiss*) and whether the velum is lowered or raised (*kin* vs. *kid*). Failure to produce one of these movements with precision will lead to loss of contrast between the target words. For example, if the velum is not raised, *kid* will be perceived as *kin*; incomplete constriction for 't' in *kit* causes listeners to hear *kiss* instead.

In connected speech (CS), one exhalation supports multiple words, sometimes multiple phrases. Exhalation must be controlled to sustain sufficient sound intensity across the phrase and to modulate intensity, pitch and timing of movements. This enables the speaker to indicate word-level stress that signals meaning, for example, *OBject* versus *obJECT* and phrase-level stress to turn a statement into a question and vice versa, for example, when remarking on a four-legged animal: *That's a dog!* versus *That's a dog?*

Once modulated air leaves the mouth of a speaker, the sound waves travel to the listener's ear and are processed through their auditory, perceptual and linguistic system for sound identification and message comprehension. Visual speech analysis allows the study of speech sound waves using spectrography. Spectrographic analysis is a robust method for identifying the characteristics of speech sounds and their articulation. A spectrogram displays a visible representation of the time course of (recorded) speech as a function of frequency and amplitude. It allows the measurement of voicing (as displayed via individual pulses in the waveform and a voice bar on the spectrogram), fundamental frequency (which is perceived as pitch), and intensity (which is perceived as loudness). Categories of sounds share spectrographic characteristics. For instance, in terms of manner categories:

- Stops/plosives are characterised by a closure period that appears as a gap in all energy on the spectrogram. This is followed by a release burst as energy rushes out when the closure is released, and either the resumption of voicing in a following vowel sound or a delay and aspiration if the stop is voiceless. The location of concentration of energy in the burst, along with transitions into and out of a stop, help to identify the place of articulation of the stop.

Pronunciation of phonemes

/b/'<u>b</u>at' /p/'<u>p</u>at' /m/'<u>m</u>an' /w/'<u>w</u>ay' /v/'<u>v</u>iew' /f/'<u>f</u>in' /ð/'<u>th</u>is' /θ/'<u>th</u>ink' /z/'<u>z</u>oo'
/s/'<u>s</u>igh' /n/'<u>n</u>o' /l/'<u>l</u>ie' /ʒ/'mea<u>s</u>ure' /ʃ/'<u>sh</u>eep' /dʒ/'jump' /tʃ/'ha<u>tch</u>' /ɹ/'<u>r</u>an' /j/'<u>y</u>es'
/g/'<u>g</u>o' /k/'<u>k</u>eep' /ŋ/'ha<u>ng</u>' /h/'<u>h</u>appy'

- Fricatives are characterised by high level random energy across a wide range of frequencies; various spectral properties of the noise generated from fricatives (e.g. the location of peak frequencies, their amplitude, and properties of the spectral slope in the transition into or out of neighbouring vowels) aid with the place of articulation of a fricative. Voiced fricatives are typically shorter than voiceless ones due to the difficulty in maintaining the aerodynamic conditions required for vocal fold vibration in the presence of high frequency random energy.
- Nasals combine properties from stops due to closure in the oral cavity with those of vowels due to air escape through the nasal cavity. This is typically manifested as regions of bounced-back energy in the oral cavity, which appear as anti-formants, combined with regions of low-energy resonance frequencies and a wide bandwidth.

- Approximants (w, j, l, r) involve the least constricted type of articulation and have formant-like structure that is similar to that of vowels, albeit with lower amplitude. Formants are bands of resonance that vary in their frequencies depending on the size of the cavities behind and in front of the highest position of the tongue.
- The first two formants (F1 and F2) play a major role in vowel identification in terms of frontness-backness and height, while F3 varies as a function of rounding and the involvement of the back (mostly pharyngeal) cavity.

An examination of the acoustic properties of the speech of children with CP has detected deviations from these expected patterns above, as well as any changes post-therapy, for example, lack of (sufficient) closure for a target plosive (p, t, k, etc.) due to weak articulation, lack of voicing or excessive voicing, low intensity, etc.[19]

It should be noted, however, that there are two caveats when examining acoustic analyses. Firstly, articulatory-to-acoustic correspondence does not always have a one-to-one relationship.[20] Small changes in articulation may lead to large acoustic changes, for example, a slide of the tip of the tongue from alveolar to post-alveolar in a narrow constriction changes *s* to *sh*, two perceptually distinct sounds with major implications for intelligibility in English. By contrast, a change from labiodental to dental frication changes *f* to *th*, two articulatorily distinct sounds but with less impact on English intelligibility given they are often neutralised in child speech and some accents (*I fink* vs. *I think*).[21] Secondly, articulatory differences (and their acoustic consequences) are not always easily perceived by listeners, especially if the differences do not meet the listener's threshold or target-like criterion.[22–24] For example, a child in one of our previous studies signalled the difference between *t* and *d* by nasalising *d*; but both were still perceived by listeners as *d*.

Impaired speech production in dysarthria associated with cerebral palsy
The Surveillance of Cerebral Palsy in Europe classifies children's type of CP by their predominant motor disorder: spastic, dyskinetic and ataxic.[25] As in other neurological disorders, spasticity is associated with upper neurone damage/maldevelopment and causes increased muscle tone, making movements slow and reduced in range. Dyskinesia (comprising dystonia and choreoathetosis in CP typology) is associated with basal ganglia involvement and leads to involuntary changes in muscle tone, with variation in range, speed and precision of movements, and the presence of involuntary movements. Ataxia, arising from cerebellar insults/maldevelopments, causes difficulties in orderly muscular co-ordination; with movements are performed with abnormal force, rhythm and accuracy. As the damage to or maldevelopments of the brain that cause CP are often diffuse, affecting more than one part of the brain, many children have mixed motor disorders[26] and classification of CP is by the predominant motor disorder.[25]

Dysarthria is the label of the speech disorder arising when these motor disorders affect speech production. Historically, five types of dysarthria were defined according to movement patterns and underlying neurophysiology: flaccid dysarthria in bulbar palsy; spastic dysarthria in pseudobulbar palsy; ataxic dysarthria in cerebellar disorders; hypokinetic dysarthria in Parkinson's; and hyperkinetic dysarthria in dystonia and chorea.[27] Each was thought to be associated with a unique profile of perceived impairments, such as harsh or breathy voice and reduced speech rate. However, it is now recognised that there is variation in presence and severity of the characteristics within dysarthria types and that speech intelligibility may be better predicted by aetiology and speech impairment severity than by dysarthria type.[28,29] Furthermore, questions have been raised about applying a classification by adult brain neurophysiology to developmental conditions such as CP.[30] Studies of the perceptual characteristics of speech point towards greater similarities rather than differences between children with spastic and dyskinetic CP, where typically movements of all speech subsystems are affected: respiration, phonation, resonance, and articulation. Children with ataxia, where the predominant difficulty is lack of appropriate intonation and fluency,[31] may show a more defined profile.

Children with dysarthria associated with spastic or dyskinetic CP, respiration is often shallow[32] and lacks co-ordination with phonation, generating weak or inconsistent subglottal pressure. Vocal folds may vibrate slowly and irregularly; air may leak through the folds when they should be adducted, reducing the intraoral pressure and weakening the sound source. The velum may rise slowly or fail to close off the nasal passage during speech. The movements of the articulators – jaw, tongue and lips – may be slow and imprecise. They may also be weak, reducing children's ability to constrict the vocal tract for consonant sounds. Although similarities have been observed in the perceptual characteristics of children with spastic and dyskinetic motor disorders, it is possible that the degree to which systems are affected may differ. Children with spastic type disorders are thought to have greater impairment in voice production and hypernasality; those with dyskinesia have been perceived to have greater impairment in the marking of stress, inappropriate voicing and slow rate.[5,33] However, phenotyping of dysarthria in children with CP is lacking.[30] *Table 2* summarises the effects of impairment of each of the speech systems on the perceptual characteristics of dysarthria in CP and their acoustic-phonetic properties.

The impact of dysarthria may be more noticeable in CS, which demands more complex speech motor control than SWs. For example, voice may become more breathy towards the end of utterances as speakers run out of air and subglottal pressure reduces.[34] Spectrographic analysis also shows that some speakers with CP may consistently produce the sound patterns expected of each phoneme, but with insufficient differentiation from its 'neighbour' leading to misperception. For example, 'ch' may be produced with a burst followed by frication, but the frication may be longer than speakers without dysarthria and perceived as 'sh'.[35] These differences are referred to as 'covert'. This is important since acoustic markers used to differentiate phonemes in articulation (formant ranges, formant slope, articulation rate, fundamental frequency and marking of bursts for stop sounds) are stronger predictors of intelligibility for children with dysarthria and CP than suprasegmental features such as loudness and voice quality.[34–37]

TABLE 2 Speech motor impairments of individual speech system function in dysarthria in CP, their perceptual characteristics, and acoustic correlates

Speech system	Perceptual characteristics	Acoustic-phonetic properties
Respiratory	Frequent inspiration[32,38,39] Inappropriate phrasing[5] Low speech volume[5] Inappropriate volume[5]	Reduced intensity[40]
Phonatory	Voice quality: harsh, breathy, strained-strangled[5,39,41] Voicing of unvoiced consonants[5,35]	Reduced overall F0 variation[42–44] Increased period to period variation (jitter and shimmer)[45] Lack of differentiation in Voice Onset Time of voiced and voiceless syllable initial consonants[35] Prolongation of vowels prior to syllable final voiceless consonants[35]
Resonatory	Hypernasality[5]	
Articulatory	Imprecise articulation[5,39] Fricative versus affricate contrast blurred[35] Imprecise plosives[35] Omission of consonants[5] Slow speech rate[5,39]	Shallow transitions in F2 between vowels and consonants[36] Reduced duration of frication in affricates[35] Low intensity burst[35,42] Longer pauses between phonemes and syllables[42] Longer transitions between phonemes[36,43,46]
Prosody	Monotone – reduced variation in melodic intonation[5,39]	Shallow pitch contours[43] Stress marked by extended duration and increased intensity rather than pitch[43]

Speech and language therapy focussing on breath support and rate: hypothesised mechanisms of action

Application of the source-filter model of speech production to the perceptual characteristics of dysarthria in spastic and dyskinetic CP hypothesises that increased intelligibility could be obtained through greater control of breath supply. According to that model, greater respiratory effort generates greater breath supply and increased air pressure during exhalation. Greater subglottal air pressure together with firmer contact of the vocal folds during phonation generates a stronger vocal note/sound source. The improved audibility and potential for greater intraoral air pressure arising from this will also help compensate for any weak closures of articulators and reduce 'leakage' of air during speech.

A steady speech rate should allow children to move with precision from one articulatory place and manner to another. Thus, as a result of changes in breath supply and rate, phonemes will be acoustically differentiated and listeners will be able to perceive the sounds that children are articulating (increased *phonetic intelligibility*).

Changes to individual phonemes are referred to as 'segmental' changes. Increased breath support and a steady rate (allowing precise movements) should also lead to 'suprasegmental' changes, that is, variables such as stress, loudness and intonation patterns that span across syllables, words and phrases. Improvements should also lead to longer breath groups (syllables spoken on one breath).

These changes should increase speakers' intelligibility. In SWs, listeners identify individual sounds, combine the sounds and then map these memories to representations of sounds of words in their vocabularies in bottom-up processing. In CS, the greater duration of phrases and increased modulation should allow listeners greater access to top down linguistic and world knowledge to process words within a phrase.[47]

Given the children's underlying neurological disorder, intervention does not aim to 'normalise' children's speech. Rather, it seeks to increase precision of the movements under volitional control so that their speech is easier to understand. Due to the complexity of speech production and variation in children's motor impairments, it is possible that the precise nature of changes underlying increases in intelligibility will vary between children.[34] Hence, our aim was also to identify possible individual or subgroup outcomes associated with increased intelligibility that should lead to enhanced personalisation of intervention.

Evidence review

We conducted a systematic review of SLT interventions for children with dysarthria for the Cochrane Collaboration.[11] The review searched the following databases up to July 2015: The Cochrane Central Register of Controlled Trials (CENTRAL; 2015 Issue 7); MEDLINE (Ovid); EMBASE (Ovid); CINAHL (EBSCOhost); ERIC (EBSCOhost); PsychInfo (Ovid); Linguistics and Language Behaviour Abstracts (LLBA) (ProQuest); Science Citation Index (Web of Science); Scopus; Dissertation Abstracts (ProQuest). The review found several phase I and phase II studies of interventions focussing on breath support and speech rate, but no quasi or randomised controlled trials. Since the review, six additional papers[16,37,48–51] reporting phase I and II studies have been published.

Two programmes have been reported most widely – the Speech Systems Approach that we developed in Newcastle, UK, and the Lee Silverman Voice Therapy (LSVT) Loud programme, which has been tested in phase I studies by a collaboration between an independent service provider LSVT Global in Arizona, United States of America (USA) and the University of Alberta, Canada. Both programmes aim to improve speech clarity by generating a sufficiently loud acoustic signal and are based on the source-filter model of speech production described above. The Speech Systems Approach also targets speech rate, to provide sufficient time for the speaker to make the precise movements needed to articulate individual speech sounds.

Lee Silverman Voice Therapy has been tested in three series of single case experimental designs replicated across participants and one group pre-post design, with a total of 22 children. Discussion with the principal investigators revealed that no randomised controlled trial (RCTs) of the intervention are currently planned. The Speech Systems Approach has been studied in two phase II studies using interrupted time series design and one feasibility RCT, involving a total of 53 children.

Both interventions have been associated with improved intelligibility. Boliek and Fox found small effects, with average increases of 7% for 6–10 year olds following LSVT.[48] The Speech Systems Approach, which additionally targets rate, increased intelligibility by an average of 15% for children aged 11–18 years[52] and 11% for children aged 7–11 years,[53] corresponding to a moderate effect size.[50] For children receiving the Speech Systems Approach, the gains have been accompanied by important changes in the frequency and success of social interactions, such as children volunteering answers to questions in class, talking to their friends at break, speaking to people outside their immediate family, and talking on the phone.[51]

Initial exploration of the changes underlying these intelligibility gains have focussed on the source in the source-filter model.

Lee Silverman Voice Therapy has been associated with increases in maximum loudness sustained phonation 5 decibel (dB) SPL and reductions in cycle-to-cycle variation of amplitude and frequency of vocal vibration (jitter and shimmer, relating to loudness and pitch) of approximately 0.5% (3% is indicative of voice disorder) in sustained vowels.[48] Small, significant reductions in jitter and shimmer (−1%) were also observed in SWs following the Speech Systems Approach.[37] However, no change was observed in harmonics to noise ratio (relating to phonatory stability across a vowel rather than cycle-to-cycle variation) for either programme.

In CS, small increases of 7 dB SPL have been observed across phrases[45,48] following LSVT. After the Speech Systems Approach, children produced longer breath groups (mean = 1.1 sec) and produced more syllables per second,[37] but the melodic intonation of their speech showed no change.[16]

Children's voice was rated as stronger after the Speech Systems Approach on the GRBAS (Grade, Roughness, Breathiness, Asthenia, Scale) four-point rating scale of voice disorder,[54] by therapists blinded to the time of recordings (mean reduction in asthenia was 0.3).[55]

The impact of individual acoustic changes following LSVT on children's intelligibility has not been tested to date. In a preliminary study, data from the first 16 children who received Speech Systems Approach found no association between acoustic change in duration, rate and period-to-period variation and intelligibility change.[37] Perceived reduction in asthenia of one point on the four point GRBAS was associated with increase in 11% intelligibility. However, the mean reduction obtained (0.3) had little impact on intelligibility change.[55]

In summary, acoustic and perceptual changes in children's speech that are expected from greater respiratory control have been observed following therapy; these include increased duration of phrases in CS, greater loudness and less cycle-to-cycle variation in SWs. However, the changes are small and do not account for the gains children made in intelligibility. The impact of the intervention on articulation, which is the strongest predictor of intelligibility in children with dysarthria and CP,[36,39] may have greater influence than suprasegmental changes in voice quality, duration and loudness.

Aims

Aim
To examine if patterns of change in segmental articulation are associated with overall gains in intelligibility following intervention that aims to create greater intraoral pressure and increased time for children to coordinate articulatory movements.

For the purposes of this study intelligibility is measured as listeners' identification of words and their constituent sounds spoken in isolation and in CS.

The complexity of speech production and the impairment of dysarthria means that no single change in articulation is expected in an explanatory model. Rather, patterns of change are predicted, and some patterns will have greater impact on intelligibility than others. Patterns may differ across children, potentially associated with type of CP or severity of speech impairment.[28] These anticipated patterns underpin our hypotheses.

Hypotheses

1. Within word articulation of singleton consonants:
   1. word-initial consonants will be identified correctly more frequently following intervention, due to increased intraoral pressure;
   2. word-final consonants will be identified correctly more frequently following intervention, due to intraoral pressure being sustained across the word.

2. Within word articulation of consonant clusters:
   1. word-initial consonant clusters (e.g. st) will be identified correctly more frequently after intervention, as a steady rate and stronger speech signal arising from increased intraoral pressure will allow speakers to differentiate each phoneme in the cluster;
   2. word-final consonant clusters (e.g. nd) will be identified correctly more frequently after intervention, as a steady rate and sustained intra-oral pressure will allow speakers to differentiate phonemes across the word.

3. Manner of articulation. Obstruent consonants (plosives, affricates and fricatives), which require the tightest constriction and intraoral pressure, will have greater increases in identification than sonorants (nasals and approximants) that have relatively free passage of sound through the oral or nasal cavity.
4. Voicing. We predict modest change in listeners' perception of voicing as greater subglottal air pressure should enable greater vocal fold abduction/adduction. Maintaining a steady rate of articulation may also enable children to combine the intricate subglottal movements needed to start and stop vocal fold vibration for the production of voiced and voiceless consonants. However, vocal fold vibration may be compromised by children's motor disorder and outside their volitional control.
5. Place of articulation. We predict no change in listeners' perception of the place of articulation. The Speech Systems Approach does not teach children how to articulate individual or groups of speech sounds. Movement of the lips, tongue tip and blade and base of the tongue that are vital for the consonant production in English are under the control of different cranial nerves, which may be impaired to varying degrees. For example, control of the base of the tongue by the Vagus nerve may be less impaired than control of the lips by the Facial nerve.

We predict that similar findings will be seen across SW utterances and CS.

Chapter 2 Methods

Design

Secondary analysis of previously collected data from three interrupted time series studies of the Speech Systems Approach.[50,52,53]

Methods

Participants

Forty-two children and young people who had received Speech Systems Approach intervention in three previously reported studies[50,52,53] provided data for this secondary analysis. To be eligible for the research the children had to have a diagnosis of CP made by a medical practitioner and moderate to severe dysarthria, as assessed by their local SLT. Children were excluded from the studies if they had hearing impairments >50 dB HL, visual impairments that were not correctable with glasses, or if they were unable to follow simple verbal instructions. *Table 3* shows the characteristics of the participants, including their motor function and mean length of utterances (MLUs) in spontaneous speech. Motor function was classified using the gross motor function classification system (GMFCS),[56] a five-level categorical scale that describes children's gross motor performance and mobility. Level 1 on the GMFCS indicates that children walk without limitations, in Level 2 children walk with handheld mobility devices and Level 5 shows severe limitations in head and trunk control and the need for a wheelchair. The mean number of words children produced in spontaneous speech (MLU) was calculated from productions in data collection sessions prior to the start of the Speech Systems Approach intervention.

Perceptual data

In the original studies providing data for this secondary analysis, children's speech was recorded on two separate days at six weeks and one week prior to therapy commencing and on two separate days at 1-, 6- and 12-weeks following the completion of the Speech Systems Approach. For the present study we used recordings taken at one week before and one week after therapy. Children were recorded saying SWs and producing CS.

Single words

Single words were elicited using the Children's Speech Intelligibility Measure (CSIM),[57] which contains 200 lists of 50 words. Lists are balanced in phonetic complexity and word length. Forty words in each list are monosyllabic; 10 words are polysyllabic. In the original studies lists were randomly allocated to children with the proviso that they were never allocated the same list twice. The recordings comprised 100 SWs spoken pre-therapy and 100 words produced post-therapy per child, which were saved in WAV (Waveform Audio File) format.

The CSIM is a forced choice word recognition task. Listeners hear each word and select the target word from 12 phonetically similar words (e.g. item 1 in the list is one of the following: born, corn, door, floor, for, form, horn, sore, storm, swarm, torn, warm. Listeners see the full list and select the word they think they have heard). The foils allow us to examine which consonants are perceived correctly in each word position, as singletons and in clusters, and whether phonemes are routinely misperceived as other phonemes for example, /b/is misperceived as/d/.

Connected speech data

Each child was recorded, describing complex pictures and answering questions using CS. The recordings were transcribed live by an expert speech and language therapist and then checked with the child, to create a gold standard transcription. Children's phrases contained a mean of 5.7 words (range 1–11)

TABLE 3 Child characteristics

Variable	Total (N = 42 children)	Pennington 2010 (N = 16 children)	Pennington 2013 (N = 15 children)	Pennington 2019 (N = 11 children)
Age, years				
Mean (SD)	11.0 (3.5)	14.6 (2.2)	8.9 (2.1)	8.8 (2.1)
Median (IQR)	11.0 (8.0–13.0)	14.5 (13.0–16.5)	9.0 (7.0–11.0)	9.0 (7.0–10.0)
Range	5.0–18.0	11.0–18.0	5.0–11.0	6.0–13.0
Female, N (%)	21 (50.0)	9 (56.3)	6 (40.0)	6 (54.6)
CP type, N (%)				
Spastic	23 (54.8)	9 (56.3)	8 (53.3)	6 (54.6)
Dyskinetic	11 (26.2)	2 (12.5)	4 (26.7)	5 (45.4)
Spastic+Dyskinetic	4 (9.5)	4 (25.0)	0	0
Ataxic	1 (2.4)	0	1 (6.7)	0
WD	3 (7.1)	1 (6.2)	2 (13.3)	0
Study source, N (%)				
Pennington 2010	16 (38.1)	–	–	–
Pennington 2013	15 (35.7)	–	–	–
Pennington 2019	11 (26.2)	–	–	–
GMFCS, N (%)				
1	6 (14.3)	1 (6.2)	0	5 (45.4)
2	12 (28.6)	4 (25.0)	8 (53.3)	0
3	6 (14.3)	2 (12.5)	2 (13.3)	2 (18.2)
4	14 (33.3)	6 (37.5)	5 (33.3)	3 (27.3)
5	4 (9.5)	3 (18.8)	0	1 (9.1)
MLU				
Mean (SD)	5.6 (2.0)	6.2 (1.2)	5.2 (2.7)	5.2 (1.9)
Median (IQR)	5.7 (4.2–7.0)	6.4 (5.8–7.0)	4.9 (2.4–7.1)	5.2 (4.1–6.9)
Range	1.9–10.4	3.4–8.4	2.0–10.4	1.9–7.9

produced over one or multiple breath groups. Up to 60 seconds of CS was presented to listeners in phrases separated by pauses of at least 3 seconds. The CS recognition task was open choice. Listeners heard the recordings of CS and wrote down the words they perceived the child to say.

The characteristics of the unique target words appearing in the SW and CS datasets are presented in *Table 4*. There were 601 unique target words in the SW data and 1156 unique target words in the CS data. Across both datasets, there was a total of 1625 unique target words. Although only 600 words exist in the CSIM, there was one target word mis-recorded as 'sall' for one participant at pre-therapy, due to an error in data entry. Attempts at correcting this error were unsuccessful due to multiple competing possibilities.

For word frequency, complexity, and density, we computed tertiles for use in the analyses (see *Percentage identification strategy*).

TABLE 4 Unique target word characteristics

Characteristics	Total (1625 words)	SW (601 words)	CS (1156 words)
Number of syllables			
Monosyllabic	943	464 (77.20)	602 (52.08)
Polysyllabic	682	136 (22.63)	554 (47.92)
Missing	1	1 (0.17)	0
Frequency			
Median (IQR)	22.5 (4–116.50)	10 (2–56)	39 (7–200)
Range	0–39,358	0–25,789	0–39,358
Available n	1552	600	1084
Tertile 1	7	–	–
Tertile 2	71	–	–
Articulatory complexity			
Median (IQR)	–	6 (4–7)	–
Range	–	0–17	--
Available n	–	600	–
Tertile 1	–	5	–
Tertile 2	–	7	–
Neighbourhood density			
Median (IQR)	–	–	192 (101–389)
Range	–	–	1–6796
Available n	–	–	823
Tertile 1	–	–	131
Tertile 2	–	–	311
Word class			
Content	–	–	975 (84–34)
Function	–	–	181 (15.66)
Initial singleton consonant n (%)	1192/1623 (73.44)	454/600 (75.67)	848/1155 (73.42)
Initial cluster n (%)	280/1623 (17.25)	117/600 (19.50)	179/1155 (15.50)
Final singleton consonant n (%)	962/1623 (59.27)	419/600 (69.83)	636/1155 (55.06)
Final cluster n (%)	284/1623 (17.50)	56/600 (9.33)	238/1155 (20.61)

Transcription of acoustic data

Recordings of the SWs and phrases in CS that were played to listeners were first phonetically transcribed automatically using acoustic models which were adapted with speech of adults with CP using the method described by Sehgal.[58] These models were then used to produce a forced-alignment of the sequence of phonemes that corresponded to the word (or words) in the recording. This forced-alignment was converted to a Praat TextGrid with two tiers, one indicating the word boundaries and another indicating the phoneme boundaries. Human transcribers were then asked to correct the initial machine-produced transcription and to align the phoneme and word boundaries to the acoustic

waveform. This process was completed by eight transcribers who were recruited from the School of Education, Communication and Language Sciences at Newcastle University. All had experience of completing phonetic transcription, and some had prior experience of listening and transcribing the speech of people with dysarthria. Transcribers were also asked to annotate a third tier of the Praat TextGrid to mark the omission of expected sounds, or the inclusion of unexpected sounds.

Transcribers were allocated equal amounts of pre-therapy and post-therapy recordings. Over 10% of each recording session (of SWs or CS) was transcribed by two transcribers. These portions of the data, that were transcribed in parallel, were then used to assess the agreement between all of the transcribers. Bland-Altman plots showed high agreement between pairs of transcribers, with no systematic disagreement within pairs.

Procedure

Adults who had no regular experience of interacting with people with disabilities or speech disorders participated as listeners in the original studies. All listeners were native speakers of English and were aged 18–55 years. All confirmed that they had no hearing difficulties (e.g. did not need to turn up the volume of the radio or television louder than other family members). Data on actual age and gender were not recorded in the studies and hence not reported here. In each study, listeners were randomly allocated three recordings, with the constraint that they did not hear a child more than once, and each recording was heard by three listeners. In the first two studies,[52,53] a researcher played the recordings to listeners individually at the same volume. Each word in the CSIM and phrase in the CS was played only once. In the third study,[50] listeners accessed the recordings via a secure web platform. Listeners were instructed to wear headphones and not to turn up the volume (although this could not be controlled). Again, each word and phrase could be played only once.

Data processing

Single words
A database was created to record perceptual characteristics of the 100 target words spoken by each child pre-therapy and the 100 words spoken post-therapy, and the words each listener perceived the child to say. In the database each row contained data on one word spoken by a child and the word one listener perceived, comprising: child ID (1–42), recording (1 = first pre-therapy recording; 2 = second pre-therapy recording; 3 = first post-therapy recording; 4 = second post-therapy recording); item number in the CSIM list (1–50); the target word; listener (1–3); the word perceived by the listener.

The perceptual characteristics of the target and perceived words were also recorded:

- the number of syllables in the word (1 or 2);
- For words starting with a singleton consonant, initial consonant:
  - voicing (voiced or voiceless);
  - place of articulation: labial (bilabial or labiodental); coronal (alveolar or post-alveolar) or dorsal (velar);
  - manner: obstruent (plosive, fricative, affricate) or sonorant (approximant, nasal);
  - manner: plosive, fricative, affricate, approximant, or nasal.

- for words starting with a consonant cluster:
  - number of consonants in the cluster (2 or 3);
  - for two consonant clusters, the voice, place and manner (obstruent or sonorant) of each consonant in the cluster.

- for words ending with a singleton consonant, final consonant:
  ◦ voicing (voiced or voiceless);
  ◦ place of articulation: labial (bilabial, labiodental); coronal (alveolar, post alveolar) or dorsal (velar);
  ◦ manner: obstruent (plosive, fricative, affricate) or sonorant (approximant, nasal);
  ◦ manner: plosive, fricative, affricate, approximant, or nasal.

- for words ending with a consonant cluster:
  ◦ number of consonants in the cluster (2 or 3);
  ◦ for two consonant clusters, the voice, place and manner (obstruent or sonorant) of each consonant in the cluster.

Three consonant clusters appear in two items of the 50-item CSIM (12/600 words in the full corpus), meaning that children would usually produce only one word with a three consonant cluster. The constituent phonemes of three consonant clusters were therefore not coded due to their low frequency in the dataset.

Phonetic complexity of the whole word was calculated using the formula by Kent,[17] which assigns four levels of complexity to consonants using the principles of motor speech difficulty. Level 1 consonants:/p, m, n, w, h/; level 2 consonants:/b, d, k, g, f, j/; level 3 consonants:/t, ŋ, r, l/; and level 4 consonants:/s, z, v, ʃ, θ, ð, tʃ, dʒ/. Level 1 to 4 represents increasingly refined articulatory adjustments and interarticulator co-ordination. The complexity of each word in the CSIM was calculated by summing the score of each constituent consonant (e.g. skill/*skɪl/* = 4+2 + 3 = 9).

As word frequency has been found to influence word recognition,[59] the frequency with which each word appears in spoken and written English per million words was also recorded using The British National Corpus.[60]

Connected speech

A second database was constructed for the words spoken in CS. The database showed each word spoken by each child and the words perceived by each listener.

In addition to the data recorded for SWs, the CS database included information on lexical class may influence both word production and word recognition. Function words (articles, prepositions, conjunctions, and pronouns e.g. 'a', 'the', 'of', 'my') are usually monosyllabic, and therefore may be easier for speakers with dysarthria to produce following the source filter model. They appear frequently in spoken language; therefore, speakers may have a motor programme they can call upon for consistent production. These function words are also closed class, facilitating their recognition. However, they may be produced with less intensity,[61] making them more difficult for listeners to perceive. Content words (nouns, verbs, adjectives, and adverbs) in contrast, are open class words, meaning they contain a larger number of words (for example, 'sunflower', 'cup', 'pretty') and are more difficult to process using top-down knowledge.[62] Word class was split into content and function words in the CS database. The position of a word in a phrase was also dichotomised into phrase initial, where breath support should be strongest (first or second word in a phrase) or phrase final (third word and beyond), where breath support may be tailing off, based on the mean number of words per utterance of 5.6 (see *Table 3*). In free choice transcription tasks listeners initially decode words sound by sound whilst simultaneously attempting to match the sounds they hear to words in their lexicon. If words share sounds with many other words, further decoding will be required to identify the target word. Words that share similarities with many other words have high neighbourhood densities; those with few similar sounding 'neighbours' have low neighbourhood densities. We measured the neighbourhood density of words in the CS from CELEX2 linguistic corpora.[63]

Measures

Independent measures

Child factors: age in years; sex (male or female); type of motor disorder (spastic, dyskinetic, spastic and dyskinetic, ataxic or Worster Drought).

Word factors: number of syllables; phonetic complexity; word frequency; word class.

Segmental factors: VPM of singleton consonants and of pairs of consonants in two consonant clusters in word initial or word final position.

Dependent measures

The primary outcome for the perceptual analysis was the identification of words and segments within them (binary outcomes). Words and sounds were identified if there was a match between the target and perceived word/consonant/cluster/voice/place/manner. Secondary outcomes include percentage identification measures and are described below (see *Definition of percentage identification*).

Data analysis for perceptual data

The SW and CS data were analysed using the same strategies, so we describe an overall data analysis plan for the perceptual data. We adopted two broad methods in our analyses:

1. using generalised linear mixed modelling (GLMM) to examine the effect of therapy (post-therapy vs. pre-therapy) of the Speech Systems Approach intervention; and
2. using child-level summaries of speech performance based on measures of percentage identification to identify subgroups of children.

Generalised linear mixed modelling strategy

The primary research objective is to examine if patterns of change in segmental articulation are associated with overall gains in percentage identification of words spoken in isolation and in CS following intervention that should create a greater intraoral pressure and provide time for children to coordinate articulatory movements. Following intervention, listeners will recognise more frequently:

- singleton consonants at the start of words;
- singleton consonants at the end of words;
- consonant clusters at the start of words;
- consonant clusters at the end of words;
- obstruent consonants;
- voicing of consonants.

We predicted no change in place of articulation following intervention.

We tested five hypotheses using the perceptual data.

The general modelling strategy to test all hypotheses was as follows:

Let the binary outcome be denoted as Y_{ij} – whether a listener correctly identified the initial (or final) singleton consonant (or consonant cluster) in the jth word spoken by the ith child. Under a latent variable formulation,[64] the binary outcome Y_{ij} may be thought of as a dichotomous re-expression of an underlying unobserved continuous latent variable Y^{*}_{ij} – representing some '*propensity*' for the child to articulate the

initial (or final) singleton consonant (or consonant cluster) and have it correctly identified by a listener such that upon exceeding some threshold in Y^*_{ij} that $Y_{ij} = 1$ is realised:[65]

$$Y_{ij} = \begin{cases} 1, & \text{if } Y^*_{ij} > 0 \\ 0, & \text{if } Y^*_{ij} \leq 0 \end{cases} \tag{1}$$

We used a GLMM with a logit link function, where child was treated as a random effect. We began with the null model to assess the proportion of variation in this latent variable of 'propensity' attributable to between-child variance [variance partition coefficient (VPC)].

$$logit\,(\pi_{ij}) = \beta_0 + b_i \tag{2}$$

$$b_i \;\sim\; N(0, \sigma^2_b) \tag{3}$$

$$VPC = \frac{\sigma^2_b}{\sigma^2_b + \sigma^2_{e*}} \tag{4}$$

Where i denotes child [i = 1, 2,... 42] and target word is denoted as j [j = 1, 2,... n_i]. π_{ij} denotes the probability of the outcome for j^{th} target word vocalised by the ith child. Parameter b_i represents the between-child (level 2) residual and is assumed to be normally distributed with mean 0 and variance σ^2_b. For a two-level logistic regression model with a random intercept, the proportion of variance attributable to between-child variation, under the latent variable formulation, is computed assuming $\sigma^2_{e*} \approx 3.29$.

After which, explanatory variables were added including:

- time point (pre- vs. post-therapy);
- child age, sex, CP type, and study source (i.e. Pennington *et al*. 2010, 2013, and 2019);
- word complexity (SW data only);
- word neighbourhood density (CS data only);
- word position (CS data only);
- word class (CS data only);
- word syllables (mono- vs. polysyllabic);
- singleton consonant (or consonant cluster) voice, place, manner (VPM) combination.

These variables were incrementally adjusted for in a series of hierarchical GLMM models beginning with a model with the random effect of child only (null model), followed by a model with time point only (model 1) and cumulatively adding child characteristics and study source (model 2), word-level characteristics (model 3), and consonant (or cluster)-level characteristics (model 4). For each model 1–4, we reported the estimated odds ratios (OR) and 95% confidence intervals (CIs), *p*-values, and estimate of the between-child variance and the standard error (SE).

For modelling purposes, the ataxic and Worster Drought CP types were combined into an 'others' category due to their low frequencies (1 and 3 children respectively). The Spastic+Dyskinetic CP group were combined with the Dyskinetic group to form an overall 'Dyskinetic' group. We did not include a random effect of listener as all listeners were naive to dysarthria, having no prior experience of conversing with people with speech disorders, and inter-rater agreement between listeners was high in the original research [0.83 (95% CI 0.78 to 0.87);[52] 0.88 (95% CI 0.8 to 0.91[53])].

We had intended to examine individual manners of articulation (e.g. fricative, approximant) and individual places of articulation (e.g. bilabial, alveolar). Exploration of the data revealed low frequencies of some places and manners of articulation in the SW data, as consonants are not equally distributed in word initial and word final position across the 200 CISM lists. We therefore grouped consonants by

their superordinate place (labial, coronal, dorsal) and manner (obstruent and sonorant). An initial data analysis of consonant voice, place, and manner variables revealed dependencies between voice and manner such that consonants with a sonorant manner are never voiceless. Because of this, we were not able to separate the main effects of voice and manner from each other if they were added as individual explanatory variables. Thus, we created a single variable containing the combinations of VPM in all GLMMs. For example, the VPM combination of the initial consonant/t/would be classified as 'voiceless-coronal-obstruent'. The VPM combination for consonant clusters follows the structure: Voicing of the first consonant, voicing of the second consonant, place of first consonant, place of second consonant, manner of first consonant, and manner of second consonant in the cluster. For example, the VPM combination of the initial consonant cluster/tr/ would be classified as 'voiceless-voiced coronal-coronal obstruent-sonorant'.

Although word frequency was available, this was not included as a covariate in the models due to its extremely right-skewed distribution; attempts at transforming this variable to meet the linearity with the logit of the outcome assumption were unsuccessful.

Hypothesis 1

To test hypotheses 1a and 1b, we analysed subsets of the data including only words with an initial singleton consonant and final singleton consonant respectively. For hypothesis 1a, the outcome was a binary variable indicating whether the initial singleton consonant of a word spoken by a child was correctly identified by a listener (initial singleton consonant correct). For hypothesis 1b, the outcome was a binary variable indicating whether the final singleton consonant of a word spoken by a child was correctly identified by a listener (final singleton consonant correct).

A likelihood ratio (LR) test was performed in models 1–4 to test the null hypothesis that the model with the time point variable does not significantly improve model fit. A statistically significant LR test provides evidence for the effect of the therapy on the identification of initial/final singleton consonants.

Hypothesis 2

To test hypotheses 2a and 2b, we analysed subsets of the data including only words with an initial consonant cluster (of two consonants in length) and final consonant cluster (of two consonants in length), respectively. Consonant clusters of three consonant length or greater were relatively infrequent and were not analysed. For hypothesis 2a, the outcome was a binary variable indicating whether the initial consonant cluster of a word spoken by a child was correctly identified by a listener (initial cluster correct). For hypothesis 2b, the outcome was a binary variable indicating whether the final consonant cluster of a word spoken by a child was correctly identified by a listener (final cluster correct).

A LR test was performed in models 1–4. A statistically significant LR test provides evidence for the effect of the therapy on the identification of initial/final consonant clusters (of two consonant length).

Hypotheses 3, 4 and 5

To test hypotheses 3–5, we extended the models used to test hypothesis 1 and added a time point x VPM combination interaction term to test the null hypothesis that the effect of therapy was homogeneous across the consonant subtypes. We tested both hypotheses using the initial singleton consonant correct and final singleton consonant correct outcomes.

A LR test was performed in models 1–4 to test the null hypothesis that the model with the interaction term does not significantly improve model fit. A statistically significant LR test provides evidence for heterogeneity of the effect of therapy on singleton consonant subtypes. Predicted probability plots were reported to visualise the interactions. The random effect of child was fixed at zero during the computation of the predicted probability.

Percentage identification strategy

Definition of percentage identification

The perceptual data were nested data whereby target words were nested within children, and for each target word there was at least one listener who listened to that target word (under the study design there are three listeners, but this was not always the case for some participants). Therefore, a single *observation* in the perceptual data was a combination of child-time point-target word-listener. Measures of percentage identification are child-level summaries and are defined as 'the percentage of *observations* whose target word contained some characteristic that was correctly identified by a listener divided by the total number of *observations* whose target word contained that characteristic'. The term characteristic in this context could mean the entire target word itself or a component of the word such as the initial consonant.

To elaborate further, an example using the percentage identification of SWs (PISW) is given. The *PISW* at a particular time point is calculated as the percentage of target words correctly identified by all listeners. Let i denote child (i = 1, 2 …, 42), j denote the time point (pre-therapy: j = 0; post-therapy: j = 1), k_{ij} denote the number of listeners who listened to child i at time point j, W_{ij} be the total number of target words the ith child verbalised at time point j, and X_{nij} denote the number of target words correctly identified by the nth listener (n = 1, …, k) for the ith child at time point j. The *PISW* for the ith child at time point j is computed as:

$$PISW_{ij} = \frac{1}{k_{ij}} \left(\frac{\sum_{n=1}^{k_{ij}} X_{nij}}{W_{ij}} \right) * 100\% \tag{5}$$

Using this definition, we generalised this approach to compute percentage identification measures of subtypes of words (e.g. PISWs which are monosyllabic; *PISW_mono*) and characteristics of words (e.g. percentage identification initial consonant of words with initial singleton consonants; *PIIC-Single*). So, the computation of *PIIC-Single*$_{ij}$ follows that of *PISW*$_{ij}$ except X_{nij} is now the number of words beginning with initial consonant singletons whose initial consonant was correctly identified by the nth listener for the ith child at time point j, and W_{ij} is the number of target words beginning with initial consonant singletons which the ith child verbalised at time point j. Additionally, we also computed PI measures where the criterion for 'correctness' X_{nij} may be the manner, place, or voicing of the consonant/cluster. For example, if a child verbalised/*ramp*/but a listener thought they had said/*damp*/, although the initial consonant was incorrectly identified the initial consonants of the target and identified words had the same voicing and place, and thus the identified initial consonant would be counted as having correct voice and correct place. On the other hand, it would be counted as manner incorrect. The tertiles used to categorise words based on their frequency, density, and complexity are presented in *Table 4*.

A caveat about the percentage identification measures should be taken into account. For each child, it is assumed that *if* the number of words presented to them tended to infinity, the performance over an infinite number of trials would converge on a 'true' value of percent identification, based on ideas from classical test theory.[66] Percentage identification is derived by averaging over a fixed number of *observations* (the denominator). When the value of the denominator is 'small', the estimated percent identification may not be a reliable representation of a child's ability. Furthermore, trivially small changes in the number of words (or consonants/clusters) correctly identified will correspond to large changes in percentage identification.

Principal component analysis

A subset of percentage identification measures (pre-therapy values) was brought forward to a principal component analysis (PCA) to reduce the dimensionality of the data and obtain a smaller number of orthogonal variables (principal components) by taking the weighted linear combination of the original data. These principal components were then brought forward into a cluster analysis to identify subgroups of children. Owing to the caveat of percentage identification described above, we

set a guideline for the subset of percentage identification measures that they should be derived over a minimum of 20 observations in every child. This cut-off was judged to be a compromise between coverage (of the available percent identification measures) and reliability.

K-means cluster analysis
We used k-means clustering to identify subgroups of children based on their performances on the principal component scores derived from the PCA. To determine the optimal number of clusters, we assessed a screen plot of the within-cluster sums of squares against the number of cluster solutions.[67]

Radar plots
Radar plots (also known as spider charts) allowed for the visualisation of several percentage identification measures simultaneously for each child at both pre- and post-therapy and offered a qualitative description of a child's performance profile on those percentage identification measures. The number of axes in each plot was equal to the number of percentage identification measures under consideration and they spanned from a common central point (percentage identification = 0%) to the outward-most point (percentage identification = 100%). Pre- and post-intervention values were visualised using different colours within the same plot for individual children. The radar plots were then used to group children who shared similar profiles.

We grouped the figures, by eye, on the size and shape of the plots, corresponding to the correct identification of voice, place and manner of consonants at the start and end of words. This allowed us to investigate similarities in the children's ability to mark each characteristic and their response to therapy. For example, the plots would show if listeners could identify voice and place correctly at the start of words but were unable to identify manner of articulation correctly. These groups were taken forward in the acoustic analysis, to investigate how children marked the articulatory contrasts within voicing, place and manner, and how their productions changed post-therapy.

Mean change in percentage identification
We reported the mean (SD) of percentage identification measures relating to the voice, place and manner of singleton consonants at the start and end of words at pre- and post-therapy along with the mean change (95% CI).

Acoustic measurements
We analysed word initial and final consonants from monosyllabic words. Production of phonemes is affected by the sounds that immediately precede and follow them (coarticulation), so separate characteristics were measured for word initial and final consonants.[17,68,69] Our analysis was two-fold. Firstly, we examined the intensity and duration of sounds, following the hypothesis that sounds would become stronger and longer with the emphasis on breath control and rate in the Speech Systems Approach. Secondly, we carried out an acoustic profiling which looks for the presence of the main characteristics of voice, place and manner of consonant sounds pre- and post-therapy. The acoustic measures for both SWs and words spoken in CS were:

- Intensity (in dB) of individual words and the intensity rms amplitude of the constituent phonemes of the words.
- Duration of words (in ms), the individual phonemes and any pauses in speaking that occurred within an utterance.
- Measures specific to manner:
  - Plosives: Plosives are signalled by a closure, followed by a release burst, and in the case of voiceless stops, aspiration. For initial stops we measured burst intensity, and f0 (mean fundamental frequency, in Hertz) and formant frequencies in the following vowel. For final stops we measured the duration of the preceding vowel along with formants at offset, closure duration and burst intensity.[35]

- ◦ Fricatives: Fricatives are recognised by the presence and duration of frication. Fricatives are differentiated from affricates by their duration and their rate of onset of the frication (rise time). Fricatives have longer rise time than affricates. We measured the total duration of frication and the rise time, as the point of frication to its maximum.[35] We also measured spectral moments to explore place of articulation.
  - ◦ Affricates: Affricates involve a stop-like burst followed by frication followed. Using the contrast information for fricatives, we measured burst intensity and duration, the duration of frication, rise time, and spectral moments.[35]
  - ◦ Nasals: Nasals have low amplitude formant-like structure and antiformants due to the extra nasal cavity. We measured duration, formants and a range of amplitude and harmonic measures ($A1$-$P1$, $A1$-$P0$; $B1$; $H1$-$A1$) in the nasal and preceding/following vowel in order to examine voice quality effects of nasal production and potential excess breathiness or creak.[70]

- Voicing: Voiced and voiceless stops are distinguished by voice onset time (VOT) in ms, f0 and the duration of the preceding/following vowel. Voicing in fricatives is distinguished by their overall duration, energy and their influence on f0, formants and duration of following vowels. Cues for voicing were measured in the consonants themselves as well as their following vowels.
- Place: While place of articulation is less reliably determined in acoustic analysis, we used known measures in stop bursts and the onset/offset of surrounding vowels (e.g. location of maximum intensity; formant transition in the vowels) in order to examine changes in place of plosives.[71]
- Consonant clusters: Acoustic profiling looked for evidence of occurrence of two separate consonants. In the absence of separate clusters we looked for cues of partial acquisition of clusters in terms of duration, voicing patterns, and intensity of the consonant that is produced and the surrounding vowels.
- For utterances we measured speech rate using syllables per second.

Scripts were written in Praat[72] to automate the measurements of the acoustic waveforms of the isolated words and CS. Separate scripts were written to make whole-utterance (CS) and whole-word (SW) measurements of duration, intensity and fundamental frequency. These used the word boundary tier of the Praat TextGrids produced by the transcribers.

For segmental-level measurements individual scripts were written for different phonetic classes (plosives, fricatives, nasals, vowels, etc.). These scripts used the phoneme boundaries from the transcriptions, produced by the transcribers, to isolate individual portions of the waveform and make appropriate measurements using the built-in functions of Praat to estimate intensity, fundamental frequency, rise-time, and formant frequencies.

Together the scripts were used to produce databases of measurements for the SWs and CS for all the children. These databases could then be used to produce descriptive statistics for measurements for individual children, recordings, and classes of speech sound.

Chapter 3 Results: perceptual analysis of single word data

Aim

To examine if patterns of change, in segmental articulation, are associated with overall gains in recognition of words spoken in isolation following intervention that aims to create greater intraoral pressure and increased time for children to coordinate articulatory movements.

Hypotheses

We predicted that following therapy for SWs we would observe:

1. more frequent correct identification of:
    1. word-initial singleton consonants;
    2. word-final singleton consonants.

2. more frequent correct identification of:
    1. word-initial consonant clusters (e.g. st);
    2. word-final consonant clusters (e.g. nd).

3. greater increases in identification of obstruents than sonorants;
4. more frequent correct identification of voicing of consonants;
5. no change in listeners' identification of the place of articulation.

Generalised linear mixed modelling strategy

Descriptive statistics
Analyses were conducted on the SW dataset containing 23,700 observations across 42 children – with a range of 300–600 observations per child.

We inspected the frequencies of each voice, place, and manner combination in singleton consonants and consonant clusters within the SW data. The results indicated absent combinations – either because these combinations do not exist, (e.g. sonorants are not voiceless in English) or are perhaps simply unrepresented in the given SW data. All linguistically possible combinations of voice, place and manner for initial and final singleton consonants were represented in the study data (see *Table 5*). For initial consonant clusters (length two), 11/14 possible combinations of voice, place, and manner had non-zero frequencies (see *Table 8*). For final consonant clusters, only 8/22 possible combinations had non-zero frequencies (see *Table 8*).

Hypothesis 1a: effect of therapy on identification of initial singleton consonants
There were 17,856 observations which included words with an initial singleton consonant. To ensure comparability of models, all models were based on *n* = 17,323 observations across 42 children with available data on the outcome, demographic variables, target word syllables, target word complexity, and the combined voice, place, and manner variable (with a range of 222–477 observations per child). The frequencies of the various voice, place and manner combinations for initial singleton consonants were inspected to ensure there were sufficient observations for modelling purposes (see *Table 5*).

TABLE 5 Frequency of combined VPM categories for initial and final singleton consonants in the GLMM

| VPM combination | Frequency | |
	Initial singleton consonants (*n* = 17,323 observations) (%)	Final singleton consonants (*n* = 16,623 observations) (%)
Voiced coronal obstruent (d, z, ð, dʒ)	1781 (10.28)	2694 (16.21)
Voiced coronal sonorant (n, j, r, l)	2955 (17.06)	3683 (22.16)
Voiced dorsal obstruent (g)	506 (2.92)	54 (0.32)
Voiced dorsal sonorant (ŋ)	NA	291 (1.75)
Voiced labial obstruent (b, v)	2131 (12.30)	706 (4.25)
Voiced labial sonorant (m, w)	2082 (12.02)	966 (5.81)
Voiceless coronal obstruent (t, s, θ, ʃ, tʃ)	3941 (22.75)	5495 (33.06)
Voiceless dorsal obstruent (k)	1406 (8.12)	1039 (6.25)
Voiceless labial obstruent (p, f)	2521 (14.55)	1695 (10.20)

A null model, with only the random effect of child, was fitted first and the estimate of the between-child (level 2) residual variance was 0.67, indicating that between-child variability explained approximately 16.92% of the total variability in the 'propensity' to correctly articulate an initial singleton consonant. This percentage varied from 14.99% to 17.34% over the four hierarchical models as fixed effects of therapy, demographics, and word-level and consonant level features were sequentially added. Generally, the therapy effect was statistically significant in all four models based on the LR tests (see *Table 6*). The estimated therapy effect was robust and had minimal changes between the models, such that the odds of the initial consonant in words, with an initial singleton consonant being accurately perceived by a listener, was 1.5 times greater post-therapy than at pre-therapy after adjusting for covariates (95% CI 1.44 to 1.65).

Hypothesis 1b: effect of therapy on identification of final singleton consonants

There were 16,624 observations which included words with a final singleton consonant. Likewise for the analysis of final singleton consonants, to ensure comparability of models, all models were based on *n* = 16,623 observations across 42 children with available data on the outcome, demographic variables, target word syllables, target word complexity, and the combined voice, place, and manner variable (with a range of 240–456 observations per child).

A null model, with only the random effect of child, was fitted first and the estimate of the between-child (level 2) residual variance was 0.56, indicating that between-child variability explained approximately

TABLE 6 Effect of therapy on identification of initial singleton consonants (*n* = 17,323 observations over *N* = 42 children)

	Post-therapy (vs. pre-therapy) OR (95% CI)	*P*-value	Between-child variance estimate (SE)
Model 1	1.51 (1.41 to 1.61)	< 0.01	0.69 (0.15)
Model 2	1.51 (1.41 to 1.61)	< 0.01	0.58 (0.13)
Model 3	1.51 (1.41 to 1.62)	< 0.01	0.59 (0.13)
Model 4	1.54 (1.44 to 1.65)	< 0.01	0.65 (0.14)

Note
P-value derived from LR test.
Model 1: includes random effect of child.
Model 2: Model 1 with fixed effects of age, sex, study source, and CP type.
Model 3: Model 2 with fixed effects of target word complexity and target word syllables.
Model 4: Model 3 with fixed effects of initial consonant voice, place, and manner combination.
See *Appendix 1*, *Table 18* for full result.

TABLE 7 Effect of therapy on identification of final singleton consonants (*n* = 16,623 observations over *N* = 42 children)

	Post-therapy (vs. pre-therapy) OR (95% CI)	*P*-value	Between-child variance estimate (SE)
Model 1	1.59 (1.49 to 1.71)	< 0.01	0.56 (0.13)
Model 2	1.59 (1.49 to 1.71)	< 0.01	0.48 (0.11)
Model 3	1.60 (1.49 to 1.71)	< 0.01	0.49 (0.11)
Model 4	1.61 (1.51 to 1.73)	< 0.01	0.50 (0.11)

Note
P-value derived from LR test.
Model 1: includes random effect of child.
Model 2: Model 1 with fixed effects of age, sex, study source, and CP type.
Model 3: Model 2 with fixed effects of target word complexity and target word syllables.
Model 4: Model 3 with fixed effects of final consonant voice, place, and manner combination.
See *Appendix 1*, *Table 19* for full results.

14.55% of the total variability in the 'propensity' to articulate a final singleton consonant and have it accurately perceived by a listener. This percentage varied from 12.72–14.55% over the four hierarchical models as fixed effects of therapy, demographics, and word-level and consonant level features were sequentially added (see *Table 7*). Generally, the therapy effect was statistically significant in all four models based on the LR tests. The estimated therapy effect was robust and had minimal changes between the models, such that the odds of the final consonant in words with a final singleton consonant being accurately perceived by a listener was 1.61 times greater post-therapy than at pre-therapy after adjusting for covariates, 95% CI (1.51 to 1.73).

Hypothesis 2a: effect of therapy on identification of initial consonant clusters (two consonants in length)

There were 4262 observations which included words with an initial consonant cluster (two consonants in length). We approached the modelling of the correct identification of initial consonant clusters much in the same way as initial singleton consonants. All models were based on *n* = 4261 observations across *N* = 42 children with available data on the outcome, demographic variables, target word syllables, target word complexity, and the combined consonant cluster VPM variable (with a range of 54–132 observations per child). The frequencies of the various VPM combinations for initial consonant clusters were inspected to ensure there were sufficient observations during modelling (see *Table 8*).

A null model, with only the random effect of child, was fitted first and the estimate of the between-child (level 2) residual variance was 0.87, indicating that between-child variability explained approximately 20.91% of the total variability in the 'propensity' to articulate an initial consonant cluster (of two consonants in length) and have it accurately perceived by a listener. This percentage varied from 17.34–21.67% over the four hierarchical models as fixed effects of therapy, demographics, and word-level and consonant cluster-level features were sequentially added (see *Table 9*). Generally, the therapy effect was statistically significant in all four models based on the LR tests. The estimated therapy effect was robust and had minimal changes between the models, such that the odds of the initial consonant cluster in words with an initial consonant cluster being accurately perceived by a listener was 1.84 times greater post-therapy than at pre-therapy after adjusting for covariates, 95% CI (1.60 to 2.12).

Hypothesis 2b: effect of therapy on identification of final consonant clusters (two consonants in length)

There were 1999 observations which included words with a final consonant cluster (two consonants in length). We approached the modelling of the correct identification of final consonant clusters much in the same way as initial consonant clusters. All models were based on *n* = 1949 observations across *N* = 42 children with available data on the outcome, demographic variables, target word syllables, target word complexity, and the combined consonant cluster VPM variable (with a range of 27–75

TABLE 8 Frequency of combined VPM categories for initial and final consonant clusters (two consonants) in the GLMM

| Acronym | VPM combination | | | Frequency | |
	Voicing	Place	Manner	Initial clusters (*n* = 4261 observations)	Final clusters (*n* = 1949 observations)
VVCCOS	Voiced-Voiced	Coronal-Coronal	Obstruent-Sonorant	203 (4.76%)	414 (21.24%)
VVDCOS	Voiced-Voiced	Dorsal-Coronal	Obstruent-Sonorant	848 (19.90%)	0
VVLCOS	Voiced-Voiced	Labial-Coronal	Obstruent-Sonorant	234 (5.49%)	0
VlVlCCOO	Voiceless-Voiceless	Coronal-Coronal	Obstruent-Obstruent	375 (8.80%)	820 (42.07%)
VlVlCLOO	Voiceless-Voiceless	Coronal-Labial	Obstruent-Obstruent	225 (5.28%)	42 (2.15%)
VlVlDCOO	Voiceless-Voiceless	Dorsal-Coronal	Obstruent-Obstruent	0	207 (10.62%)
VlVlLCOO	Voiceless-Voiceless	Labial-Coronal	Obstruent-Obstruent	0	143 (7.34%)
VlVlCDOO	Voiceless-Voiceless	Coronal-Dorsal	Obstruent-Obstruent	191 (4.48%)	84 (4.31%)
VlVCCOS	Voiceless-Voiced	Coronal-Coronal	Obstruent-Sonorant	801 (18.80%)	0
VlVCLOS	Voiceless-Voiced	Coronal-Labial	Obstruent-Sonorant	490 (11.50%)	0
VlVDCOS	Voiceless-Voiced	Dorsal-Coronal	Obstruent-Sonorant	384 (9.01%)	0
VlVDLOS	Voiceless-Voiced	Dorsal-Labial	Obstruent-Sonorant	118 (2.77%)	0
VlVLCOS	Voiceless-Voiced	Labial-Coronal	Obstruent-Sonorant	392 (9.20%)	0
VVlCCSO	Voiced-Voiceless	Coronal-Coronal	Sonorant-Obstruent	0	146 (7.49%)
VVlLLSO	Voiced-Voiceless	Labial-Labial	Sonorant-Obstruent	0	93 (4.77%)

Note
V, voiced; Vl, voiceless; C, coronal; L, labial; D, dorsal; O, obstruent; S, sonorant.
For consonant clusters, the VPM combinations are reported as acronyms for ease of reporting whose structure is as follows: voice of first consonant, voice of second consonant, place of first consonant, place of second consonant, manner of first consonant, and manner of second consonant in the cluster. For example, the initial consonant cluster/st/has/s/ as the first consonant and/t/ as the second consonant.

TABLE 9 Effect of therapy on identification of initial consonant clusters (n = 4261 observations over N = 42 children)

	Post-therapy (vs. pre-therapy) OR (95% CI)	P-value	Between-child variance estimate (SE)
Model 1	1.77 (1.55 to 2.03)	< 0.01	0.91 (0.21)
Model 2	1.77 (1.55 to 2.03)	< 0.01	0.69 (0.17)
Model 3	1.79 (1.56 to 2.05)	< 0.01	0.70 (0.17)
Model 4	1.84 (1.60 to 2.12)	< 0.01	0.78 (0.19)

Note
P-value derived from LR test.
Model 1: includes random effect of child.
Model 2: Model 1 with fixed effects of age, sex, study source, and CP type.
Model 3: Model 2 with fixed effects of target word complexity and target word syllables.
Model 4: Model 3 with fixed effects of initial consonant cluster voice, place, and manner combination.
See *Appendix 1*, *Table 20* for full results.

observations per child). The frequencies of the various VPM combinations for final consonant clusters were inspected to ensure there were sufficient observations during modelling (see *Table 8*).

A null model, with only the random effect of child, was fitted first and the estimate of the between-child (level 2) residual variance was 0.99, indicating that between-child variability explained approximately 23.13% of the total variability in the 'propensity' to articulate a final consonant cluster (of two consonants in length) and have it accurately perceived by a listener. This percentage varied from 18.77–23.49% over the four hierarchical models as fixed effects of therapy, demographics, and word-level and consonant cluster-level features were sequentially added (see *Table 10*). Generally, the therapy effect was statistically significant in all four models based on the LR tests. The estimated therapy effect was robust and had minimal changes between the models, such that the odds of the final consonant cluster in words with a final consonant cluster being accurately perceived by a listener was 1.42 times greater post-therapy than at pre-therapy after adjusting for covariates, 95% CI (1.25 to 1.85).

Hypotheses 3–5: moderating effect of consonant voice, place, and manner in the identification of initial and final singleton consonants

For the identification of initial singleton consonants, there was a statistically significant therapy by VPM-combination interaction when the interaction term was included in Model 4 of *Table 6*, indicating heterogeneity of therapy effect across the consonant subtypes (LR test for interaction: $X^2(7) = 25.84$, p < 0.01). The interaction plot indicates that initial singleton consonants which were

TABLE 10 Effect of therapy on identification of final consonant clusters (n = 1949 observations over N = 42 children)

	Post-therapy (vs. pre-therapy) OR (95% CI)	P-value	Between-child variance estimate (SE)
Model 1	1.38 (1.13 to 1.70)	< 0.01	1.01 (0.26)
Model 2	1.38 (1.13 to 1.70)	< 0.01	0.76 (0.20)
Model 3	1.37 (1.11 to 1.68)	< 0.01	0.84 (0.22)
Model 4	1.42 (1.15 to 1.75)	< 0.01	0.85 (0.22)

Note
P-value derived from LR test.
Model 1: includes random effect of child.
Model 2: Model 1 with fixed effects of age, sex, study source, and CP type.
Model 3: Model 2 with fixed effects of target word complexity and target word syllables.
Model 4: Model 3 with fixed effects of final consonant cluster voice, place, and manner combination.
See *Appendix 1*, *Table 21* for full results.

voiced-labial-obstruent (b, v) had, descriptively, the highest probability of being correctly perceived by listeners at pre- and post-therapy, whilst initial singleton consonants which were voiceless-coronal-obstruent (t, s, θ, ʃ, ʧ) or voiceless-labial-obstruent (p, f) had, descriptively, the lowest probability of being correctly perceived by listeners at pre- and post-therapy (see *Figure 1*). Whilst all eight initial singleton consonant subtypes appeared on average to benefit from therapy, the effect of therapy was relatively larger in the voiced-labial-sonorant (m, w), voiceless-coronal-obstruent (t, s, θ, ʃ, ʧ), and voiceless-labial-obstruent subtypes (p, f) (see *Table 11*).

For the identification of final singleton consonants, there was also a statistically significant therapy by VPM-combination interaction when the interaction term was included in Model 4 of *Table 7* (LR test for interaction: $X^2(8) = 26.54$, $p < 0.01$). The interaction plot indicates that final singleton consonants which were voiced-coronal-obstruent (d, z, ð, ʤ) or voiceless-dorsal-obstruent (k) had, descriptively, the highest probability of being correctly perceived by listeners at pre- and post-therapy, whilst those which were voiced-dorsal-sonorant (ŋ) had, descriptively, the lowest probability at both time points (see *Figure 2*). All subtypes except for the voiced-dorsal-obstruent (g) subtype appeared to benefit from the therapy with the latter experiencing no change, on average, in the probability of being correctly perceived. The effect of therapy was relatively larger in the voiced-labial-obstruent (b, v) and voiced-labial-sonorant (m, w) subtypes (see *Table 11*).

Percentage identification strategy

Child-level percentage identification descriptive statistics, N = 42 children
Of the 76 percentage identification measures computed, 30 were derived over a minimum of 20 observations per child at pre- and post-therapy (see *Appendix 2*, *Table 24*).
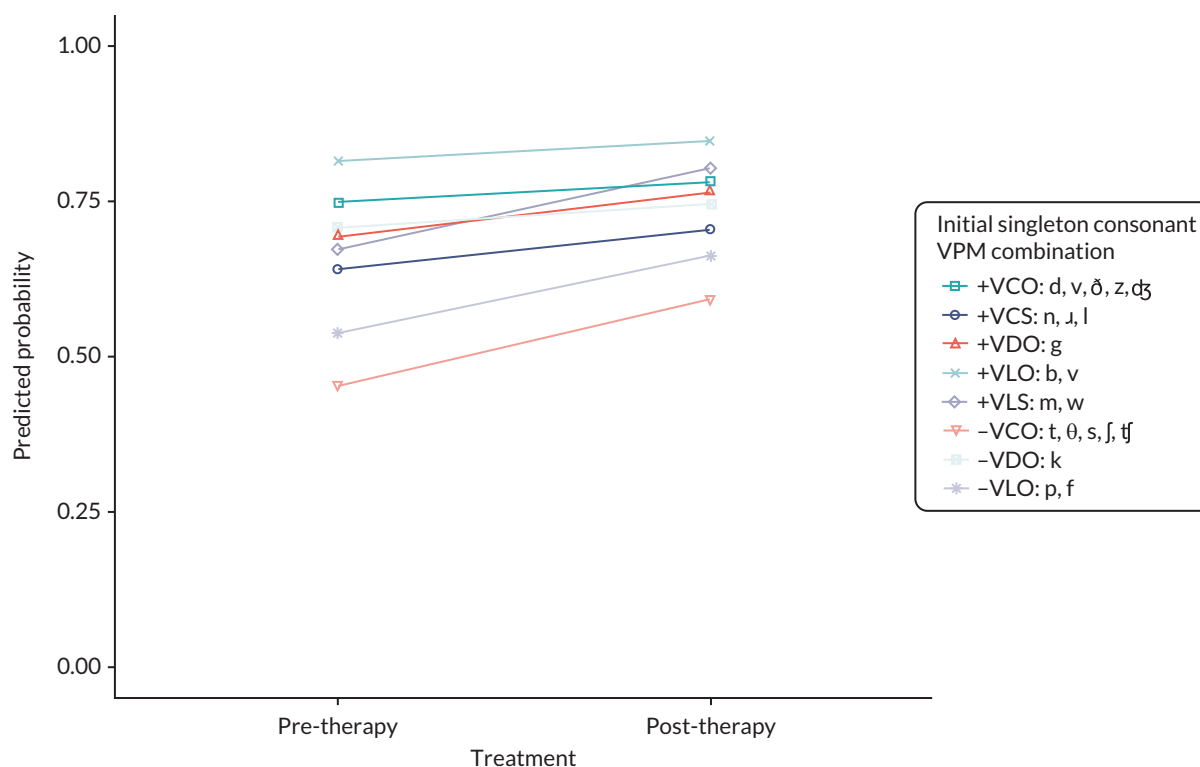
FIGURE 1 Interaction plot for therapy and initial singleton consonant VPM combination. +V, voiced; -V, voiceless; C, coronal; L, labial; D, dorsal; O, obstruent; S, sonorant. Phonemes represented by the different VPM combinations are displayed in the legends.

TABLE 11 Predicted probability (95% CI) of therapy and initial and final singleton consonant VPM combination subtype strata

	Initial singleton consonant		Final singleton consonant	
	Pre-therapy	**Post-therapy**	**Pre-therapy**	**Post-therapy**
Voiced coronal obstruent (d, z, ð, dʒ)	0.75 (0.69 to 0.80)	0.78 (0.73 to 0.83)	0.52 (0.46 to 0.58)	0.60 (0.54 to 0.66)
Voiced coronal sonorant (n, j, r, l)	0.64 (0.58 to 0.70)	0.71 (0.65 to 0.76)	0.71 (0.67 to 0.76)	0.78 (0.74 to 0.82)
Voiced dorsal obstruent (g)	0.70 (0.62 to 0.77)	0.77 (0.70 to 0.83)	0.54 (0.34 to 0.74)	0.54 (0.34 to 0.74)
Voiced dorsal sonorant (ŋ)	NA	NA	0.40 (0.30 to 0.50)	0.49 (0.39 to 0.59)
Voiced labial obstruent (b, v)	0.82 (0.77 to 0.86)	0.85 (0.81 to 0.89)	0.51 (0.43 to 0.58)	0.70 (0.64 to 0.77)
Voiced labial sonorant (m, w)	0.67 (0.61 to 0.73)	0.81 (0.76 to 0.85)	0.53 (0.46 to 0.60)	0.75 (0.69 to 0.81)
Voiceless coronal obstruent (t, s, θ, ʃ, tʃ)	0.46 (0.39 to 0.52)	0.59 (0.53 to 0.66)	0.60 (0.54 to 0.65)	0.70 (0.65 to 0.75)
Voiceless dorsal obstruent (k)	0.71 (0.65 to 0.77)	0.75 (0.69 to 0.80)	0.69 (0.63 to 0.75)	0.77 (0.72 to 0.83)
Voiceless labial obstruent (p, f)	0.54 (0.47 to 0.60)	0.66 (0.60 to 0.72)	0.56 (0.49 to 0.62)	0.69 (0.64 to 0.75)
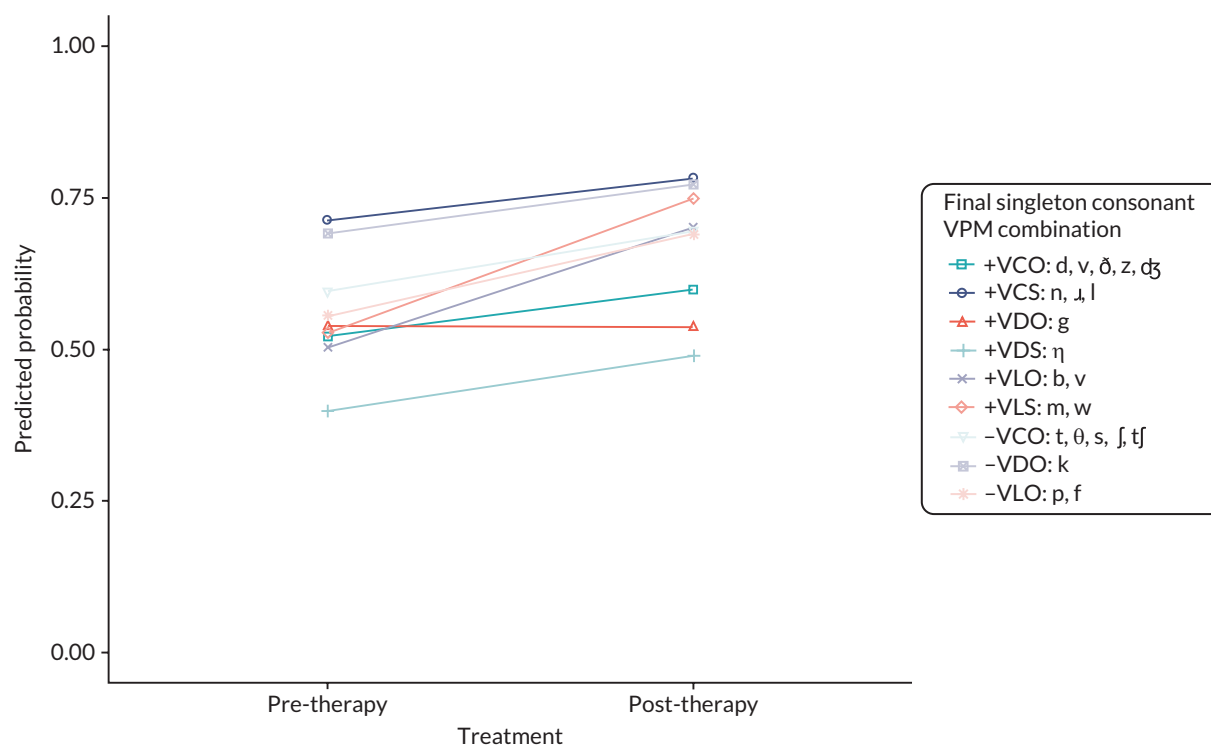
FIGURE 2 Interaction plot for therapy and final singleton consonant VPM combination. +V, voiced; -V, voiceless; C, coronal; L, labial; D, dorsal; O, obstruent; S, sonorant. Phonemes represented by the different VPM combinations are displayed in the legends.

The mean (SD) PISW was 43.02% (20.29%) at pre-therapy and 53.67% (21.71%) at post-therapy. The mean difference on PISW from pre- to post-therapy was 10.6% (SD 10.63%); the standard deviation of the differences indicates the large variation between children in the difference from pre-to post-PISW. Based on this data the 95% CI for the mean difference in PISW from pre- to post-intervention is 7.32–13.94% (see *Figure 3*).

Principal component analysis

From the 30 candidate percentage identification measures, nine measures at pre-therapy were brought forward into a PCA. The percentage identification measures retained were related to the articulatory complexity of the word (complexity and syllables) and segmental articulatory characteristics specified in the hypotheses – the voice, place and manner of singleton consonants and the production of clusters.

- SWs of complexity > tertile 2 (upper third);
- polysyllabic words;
- initial consonant voicing in words with initial singleton consonants;
- initial consonant place in words with initial singleton consonants;
- initial consonant manner in words with initial singleton consonants that have obstruent manner;
- final consonant voicing in words with final singleton consonants;
- final consonant place in words with final singleton consonants;
- final consonant manner in words with final singleton consonants that have obstruent manner;
- initial consonant cluster (two consonant length) in words with initial clusters.

The PI measures were strongly inter-correlated with the smallest bivariate correlation of $r = 0.58$ (see *Appendix 2*, *Figure 18*). Assumptions of bivariate linearity were met and there were no obvious outliers. The results suggested that a single component was sufficient to explain 88% of the total variance. The Kaiser-Meyer-Olkin measure of sampling adequacy was 0.90. The component loadings of each variable suggest that each is given approximately the same weight in the computation of a component score.
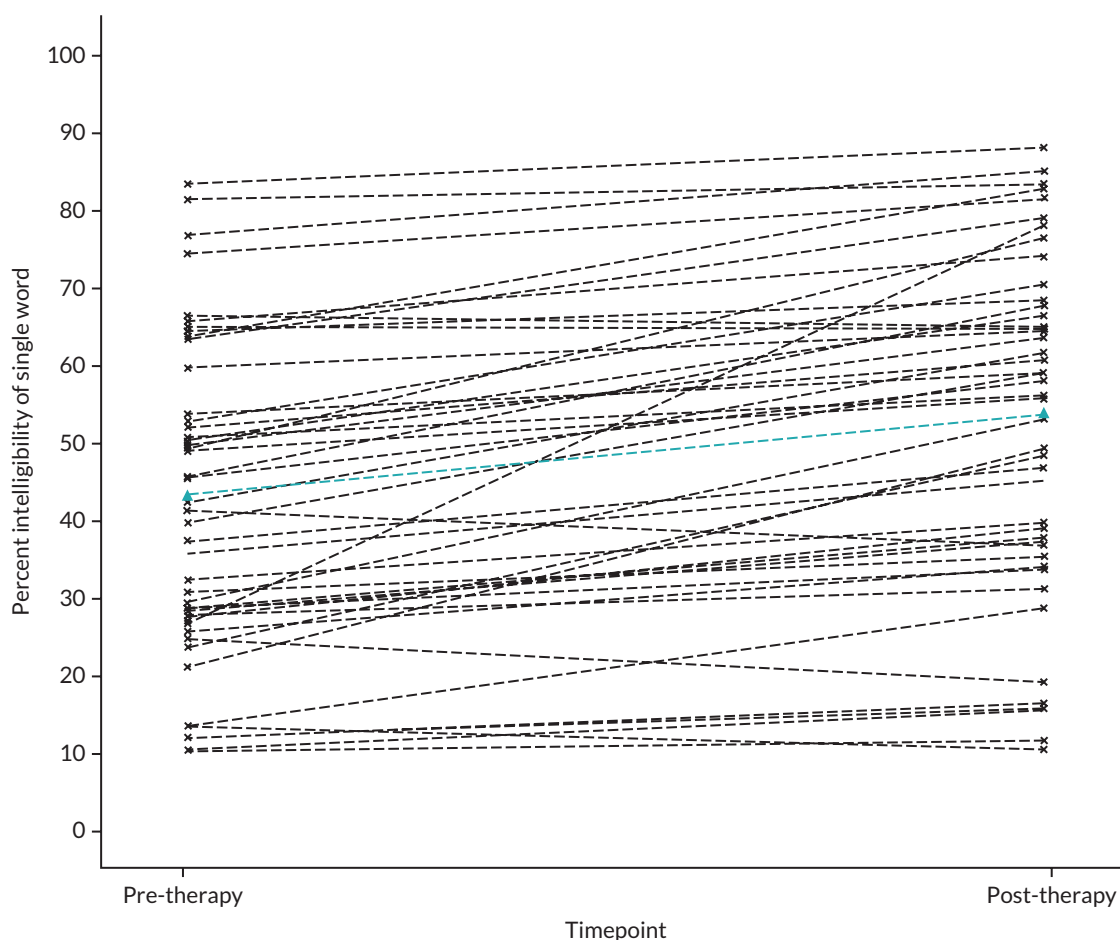
FIGURE 3 Spaghetti plot of individual change in PISWs from pre- to post-therapy, *N* = 42. Each black cross and dashed line describes the change in a participant's PISW over time. Light blue triangles indicate the mean PISW at pre- and post-therapy.

Descriptive statistics of these PI measures at pre-therapy and their loadings on the first component are presented in *Appendix 2*.

Cluster analysis

We computed a component score for each child at pre-therapy and post-therapy. To compute the component score at a given time point, the nine variables were first standardised using their pre-therapy means and standard deviations. Next we took the linear combination of these variables, weighted by their loadings. We used K-means clustering to identify subgroups of children based on their pre-therapy and post-therapy component scores. However, with only 42 children it was not possible to determine reliable, distinct clusters and hence the K-means cluster analysis was not pursued.

Radar plots

To further understand the patterns of change, pre- and post-therapy information on six identification measures were included in the radar plots for all 42 children, accompanied by additional information on their ID and CP type. One radar plot was generated for each child showing the percentage identification (from the vertical axis at the 12 o'clock position and reading anti-clockwise) of:

- initial consonant voicing in words with initial singleton consonants;
- initial consonant place in words with initial singleton consonants;
- initial consonant manner in words with initial singleton consonants;
- final consonant voicing in words with final singleton consonants;
- final consonant place in words with final singleton consonants;
- final consonant manner in words with final singleton consonants.

Pre- and post-intervention values were visualised using different colours within the same plot for individual children. Based on the resulting profiles, the plots were grouped by eye by two researchers who did not see the other's grouping. Six groups were identified.

Following the categorisation by VPM, we added the children's percentage word identification to the descriptions of the groups to help interpret overall intelligibility and possible reasons for incorrect word identification. Children with spastic, dyskinetic and spastic+dyskinetic CP type were found in each group. Children with Worster Drought and ataxia were also spread across the groups. Figure 4 shows illustrative examples of the plots from each group.

Group 1 comprised four participants (12, 13, 28, 42), whose overall SW identification was low pre-therapy (15–25%). Listeners identified the manner and place of articulation correctly in < 50% of both word initial and word final consonants pre-and post-therapy. Voicing was more frequently identified (50–75%) in word initial and final consonants before intervention. There was little change in the identification of any individual characteristic in word initial or final consonants post-therapy and on average almost no change in overall word identification (mean change = 0.75%; range −6 to 4%).

Group 2 comprised seven children (6, 8, 16, 20, 34, 40, 47). Identification of SWs for this group ranged between 15% and 40% pre-therapy. Listeners more frequently identified voicing in word initial and word final consonants (approximately 75%) than place or manner (60–70%). This suggests that possibly two out of the three characteristics were usually correctly, and one of the two features perceived correctly was the voicing contrast. Post-therapy we observed increases in the identification of place of articulation in either word initial or word final consonants, and a slight overall mean increase in word identification (mean change = 5%, range −4 to 15%).

Group 3 comprised 11 children: (21, 23, 24, 27, 31, 32, 36, 38, 46, 48, 49). Listeners identified 50–83% of their SWs pre-therapy. Their radar plots showed 50–75% correct identification of place and manner of word initial and word final consonants at both pre-and post-therapy. Like Group 1, voicing was slightly better identified (75–97%). Little or no change in any characteristic was identified post-therapy. Some children made gains in overall intelligibility (mean = 4.3%, range −1 to 9%), suggesting some children achieved control of voicing, place and manner within some words after intervention.

Group 4 comprised seven children (1, 5, 10, 14, 18, 25, 53) who had low SW identification pre-therapy, ranging from 10% to 29%. The identification of their place and manner of word initial and word final consonants was approximately 50% pre-therapy, with voicing slightly higher (up to 75%), possibly suggesting that similar to Group 2 a maximum of two characteristics (usually including voicing) were identified correctly for each consonant. Clear increases across most VPM variables were observed post-therapy, which enabled increases in overall word identification (mean change = 16% range 4–28%).

Group 5 comprised four children (2, 7, 15, 19). Their SW identification was low pre-therapy (26–45%) but identification of individual voice, place and manner for word initial and word final consonants was high (approximately 75%), suggesting that possibly two out of the three characteristics were usually correctly identified but that the two features were not consistent (e.g. voice and place correct in word initial consonant for one word; manner and place correct in a second word). Identification VPM of word initial and word final consonants increased to nearly 100% after intervention, with correspondingly large increases in word identification (mean change = 25%, range 19–31%).

Group 6 comprised nine children (3, 4, 9, 11, 17, 22, 26, 29, 30). Similar to Group 3, SW identification for children in Group 6 ranged 44–73% pre-therapy and identification of voice, place and manner in word initial and final positions pre-therapy ranged 50–75%. Group 6, however, had increases of approximately 5–12% in identification of most articulatory characteristics post-therapy, but substantial increases in word identification (mean change = 14.6%, range 8–20%). With smaller increases in voice, place and manner identification, other changes must also be contributing to changes in full word
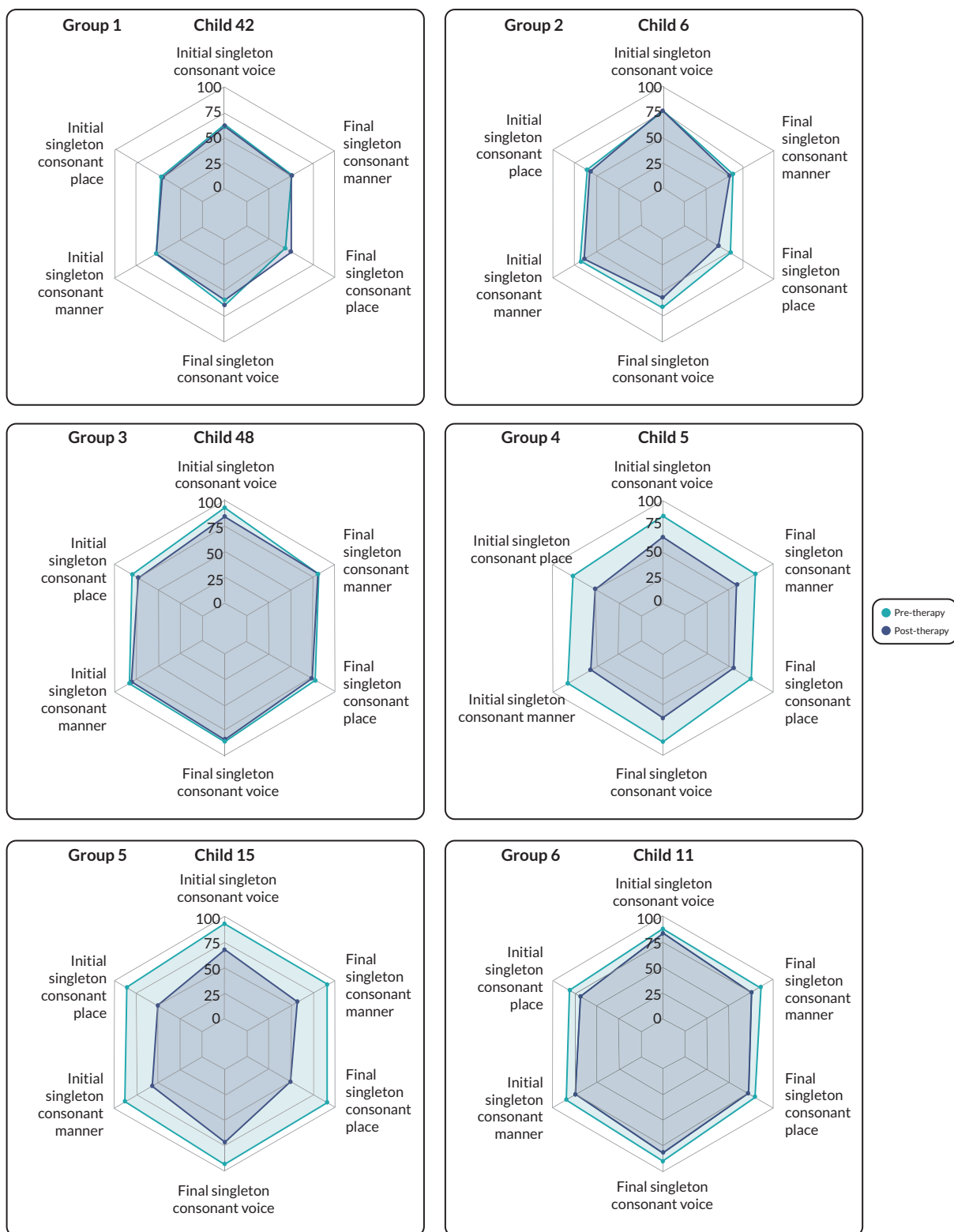
FIGURE 4 Radar plots using SW data.

identification. Inspection of raw scores of other variables showed that the identification of word initial and final clusters increased for this group.

Change in percentage identification measures from pre- to post-therapy

On average, participants showed improvements in articulation from pre- to post-therapy. In particular, on average they displayed positive changes in their ability to articulate whole words, regardless of the number of syllables; initial and final singleton consonants; and initial and final consonant clusters (see *Table 12*).

TABLE 12 Mean change in pre-specified percentage identification measures

Percent identification (%) of	Pre-therapy, mean (SD)	Post-therapy, mean (SD)	Mean change (95% CI), post minus pre
SWs	43.04 (20.29)	53.67 (21.71)	10.63 (7.32 to 13.94)
Monosyllabic words	43.37 (20.16)	53.25 (46.40)	9.88 (6.32 to 13.45)
Polysyllabic words	42.00 (22.05)	55.07 (22.09)	13.06 (9.17 to 16.96)
Initial singleton consonants	61.45 (17.79)	69.19 (18.19)	7.73 (4.52 to 10.95)
Initial singleton consonant voicing	80.85 (11.90)	85.42 (11.37)	4.56 (2.10 to 7.03)
Initial singleton consonant place	68.09 (15.54)	74.60 (15.67)	6.51 (3.30 to 9.72)
Initial singleton consonant manner	73.06 (14.28)	78.95 (14.23)	5.88 (3.02 to 8.75)
Obstruent initial singleton consonants	75.18 (14.28)	80.68 (13.16)	5.50 (2.85 to 8.15)
Final singleton consonants	59.08 (16.90)	68.63 (16.84)	9.54 (6.51 to 12.57)
Final singleton consonant voicing	78.46 (11.44)	84.21 (10.59)	5.75 (3.43 to 8.07)
Final singleton consonant place	67.03 (14.79)	75.89 (13.30)	8.87 (6.02 to 11.72)
Final singleton consonant manner	72.40 (13.04)	79.71 (12.71)	7.31 (4.58 to 10.04)
Obstruent final singleton consonants	70.55 (14.42)	78.63 (14.15)	8.08 (4.81 to 11.35)
Initial cluster	49.68 (22.21)	61.39 (23.04)	11.71 (6.47 to 16.95)
Final cluster[†]	51.43 (23.39)	58.18 (24.11)	6.75 (1.08 to 12.42)

Note
† PIFCL_ClusterL2 was not derived from a minimum number of 20 observations at pre- and post-therapy for all children and should be interpreted with caution.

Chapter 4　Results: perceptual analysis of connected speech data

Aim

To examine if patterns of change in segmental articulation are associated with overall gains in recognition of words spoken in CS following intervention that aims to create greater intraoral pressure and increased time for children to coordinate articulatory movements.

Hypotheses

We predicted that for words produced in CS following therapy we would observe:

1. more frequent correct identification of:
    1. word-initial singleton consonants;
    2. word-final singleton consonants.

2. more frequent correct identification of:
    1. word-initial consonant clusters (e.g. st);
    2. word-final consonant clusters (e.g. nd).

3. greater increases in identification of obstruents than sonorants;
4. more frequent correct identification of voicing of consonants;
5. no change in listeners' identification of the place of articulation.

Generalised linear mixed modelling strategy

Descriptive statistics

Analyses were conducted on the CS dataset containing 17,576 observations across 42 children – with a range of 103–1116 observations per child. For initial consonant clusters (length two), 13/14 possible combinations of voice, place, and manner had non-zero frequencies. For final consonant clusters, 15/22 possible combinations had non-zero frequencies.

Hypothesis 1a: effect of therapy on identification of initial singleton consonants of words in connected speech

There were 11,209 observations which included an initial singleton consonant. To ensure comparability of models, all models were based on n = 8457 observations across 42 children with available data on the outcome, demographic variables, target word syllables, target word density, and the combined VPM variable (range 40–564 observations per child). *Table 13* shows the frequencies of the VPM combinations.

A null model, with only the random effect of child, was first fitted and the estimate of the between-child (level 2) residual variance was 1.46, indicating that between-child variability explained approximately 30.74% of the total variability in the 'propensity' to correctly articulate an initial singleton consonant. This percentage varied from 28.01% to 30.64% over the four hierarchical models as fixed effects of therapy, demographics, and word-level and consonant level features were sequentially added. Generally, the therapy effect was statistically significant in all four models based on the LR tests (see *Table 14*).

TABLE 13 Frequency of combined VPM categories for initial and final singleton consonants in the GLMM

VPM combination	Frequency	
	Initial singleton consonant (n = 8457 observations) (%)	Final singleton consonant (n = 7205 observations) (%)
Voiced coronal obstruent (d, z, ð, dʒ)	2239 (26.48)	1106 (15.35)
Voiced coronal sonorant (n, j, r, l)	1235 (14.60)	2015 (27.97)
Voiced dorsal obstruent (g)	414 (4.90)	171 (2.37)
Voiced dorsal sonorant (ŋ)	NA	530 (7.36)
Voiced labial obstruent (b, v)	834 (9.86)	327 (4.54)
Voiced labial sonorant (m, w)	1550 (18.33)	340 (4.72)
Voiceless coronal obstruent (t, s, θ, ʃ, tʃ)	1152 (13.62)	1955 (27.13)
Voiceless dorsal obstruent (k)	433 (5.12)	445 (6.18)
Voiceless labial obstruent (p, f)	600 (7.09)	316 (4.39)

TABLE 14 Effect of therapy on identification of initial singleton consonants in CS data (n = 8457 observations over N = 42 children)

	Post-therapy (vs. pre-therapy) OR (95% CI)	P-value	Between-child variance estimate (SE)
Model 1	1.26 (1.15 to 1.39)	< 0.01	1.45 (0.34)
Model 2	1.26 (1.15 to 1.39)	< 0.01	1.28 (0.30)
Model 3	1.27 (1.15 to 1.39)	< 0.01	1.30 (0.31)
Model 4	1.26 (1.15 to 1.39)	< 0.01	1.31 (0.31)

Note
P-value derived from LR test.
Model 1: includes random effect of child.
Model 2: Model 1 with fixed effects of age, sex, study source, and CP type.
Model 3: Model 2 with fixed effects of word position in phrase, natural log of target word density, and target word syllables.
Model 4: Model 3 with fixed effects of initial consonant voice, place, and manner combination.
See *Appendix 1*, *Table 22* for full results.

The estimated therapy effect was robust and had minimal changes between the models, such that the odds of the initial consonant in words with an initial singleton consonant being accurately perceived by a listener was 1.26 times greater post-therapy than at pre-therapy after adjusting for covariates, 95% CI (1.15 to 1.39).

Hypothesis 1b: effect of therapy on identification of final singleton consonants
There were 8712 observations which included a final singleton consonant. To ensure comparability of models, all models were based on n = 7205 observations across 42 children with available data on the outcome, demographic variables, target word syllables, target word density, and the combined voice, place, and manner variable (with a range of 40–465 observations per child). The frequencies of the various VPM combinations for final singleton consonants were inspected to ensure there were sufficient observations during modelling (see *Table 13*).

A null model, with only the random effect of child, was first fitted and the estimate of the between-child (level 2) residual variance was 1.30, indicating that between-child variability explained approximately 28.32% of the total variability in the 'propensity' to correctly articulate a final singleton consonant. This percentage varied from 25.90–28.32% over the four hierarchical models as fixed effects of therapy, demographics, and word-level and consonant level features were sequentially added. Generally, the therapy effect was statistically significant in all four models based on the LR tests (see *Table 15*). The estimated therapy effect was robust and had minimal changes between the models, such that the odds of the final consonant in words with a final singleton consonant being accurately perceived by a listener was 1.27 times greater post-therapy than at pre-therapy after adjusting for covariates, 95% CI (1.15 to 1.41).

Hypothesis 2a: effect of therapy on identification of initial consonant clusters (two consonants in length)

We had intended to approach the modelling of the correct perception of initial consonant clusters in much the same way as initial singleton consonants. However, the low number of words with initial consonant clusters in the CS data (179/1155), with $n = 1072$ observations over $N = 42$ children; range of 3–99 observations per child, and the distribution of the observations across the combined VPM categories, meant that it was not feasible to test this hypothesis in a multivariable GLMM in the same way (see *Appendix 3*, *Tables 25* and *26*).

Hypothesis 2b: effect of therapy on identification of final consonant clusters (two consonants in length)

Similarly for final consonant clusters in the CS data. The low number of words with final consonant clusters in the data (238/1155), with $n = 1900$ observations over $N = 42$ children; range of 5–120 observations per child, and the distribution of the observations across the combined VPM categories meant that it was also not feasible to test this hypothesis in a multivariable GLMM in the same way (see *Appendix 3*).

Hypotheses 3–5: moderating effect of consonant voice, place, and manner on identification of initial and final singleton consonants

There was evidence of a statistically significant therapy by VPM-combination interaction in the analysis of initial singleton consonants when the interaction term was included in Model 4 of *Table 14* (LR test for interaction: $X^2(7) = 20.31$, $P < 0.01$). This suggests that the effect of the therapy was heterogeneous across the initial singleton consonant subtypes. Descriptively, initial singleton consonants whose VPM were voiced-coronal-obstruent (d, z, ð, dʒ), voiced-coronal-sonorant (n, j, r, l), voiced-labial-sonorant (m, w), and voiceless-coronal-obstruent (t, s, θ, ʃ, tʃ) showed improvements in the predicted probability of correct identification of the initial consonant (range of pre-post change: 0.05–0.12). The predicted

TABLE 15 Effect of therapy on identification of final singleton consonants in CS data ($n = 7205$ observations over $N = 42$ children)

Model 1	1.29 (1.17 to 1.43)	< 0.001	1.30 (0.31)
Model 2	1.29 (1.16 to 1.43)	< 0.001	1.16 (0.28)
Model 3	1.29 (1.16 to 1.43)	< 0.001	1.15 (0.28)
Model 4	1.27 (1.15 to 1.41)	< 0.001	1.19 (0.29)

Note
P-value derived from LR test.
Model 1: includes random effect of child.
Model 2: Model 1 with fixed effects of age, sex, study source, and CP type.
Model 3: Model 2 with fixed effects of word position in phrase, natural log of target word density, and target word syllables.
Model 4: Model 3 with fixed effects of final consonant voice, place, and manner combination.
See *Appendix 1*, *Table 23* for full results.

probability of listeners' correct identification of initial singleton consonants which were voiceless-dorsal-obstruent (k) were approximately the same pre- and post-therapy. Initial singleton consonants which were voiced-dorsal-obstruent (g), voiced-labial-obstruent (b, v), and voiceless-labial-obstruent (p, f) displayed relatively minor reductions in the probability of their correct identification after therapy; the pre-post therapy changes in the predicted probability of these three subtypes ranged from −0.03 to −0.02 (see *Table 16*). These changes are visualised in *Figure 5*.

There was also evidence of a statistically significant interaction in the analysis of final singleton consonants when the interaction term was included in Model 4 of *Table 15* (LR test for interaction: $X^2(8) = 21.29$, *p* < 0.01). Descriptively, most of the final singleton consonant subtypes appeared to benefit from therapy with the largest improvements in the predicted probability of correct identification observed in final singleton consonants whose VPM were voiced-coronal-obstruent (d, z, ð, ʤ), voiced-labial-obstruent (b, v), voiceless-coronal-obstruent (t, s, θ, ʃ, ʧ), and voiceless-dorsal-obstruent (k) (range of pre-post change: 0.02 to 0.13). Similar to the analysis of initial singleton consonants, some final singleton consonant subtypes displayed a reduction in the predicted probability of correct identification after therapy, that is, voiced-dorsal-sonorant (ŋ), voiced-labial-sonorant (m, w), and voiceless-labial-obstruent (p, f). The pre-post therapy changes in the predicted probability for these three subtypes ranged from −0.04 to −0.02 (see *Table 16*). These changes are visualised in *Figure 6*.

Percentage identification strategy.

Percentage identification descriptive statistics
Of the 76 PI measures computed, 13 were derived over a minimum of 20 observations per child at pre- and post-therapy (see *Appendix 4*, *Table 27*). It should be noted that none of the consonant cluster-related PI measures met the criterion of 20 observations per child, signalling the infrequent use of words with consonant clusters in CS by children with dysarthria.

The mean (SD) PISW was 32.15% (23.03%) at pre-therapy and 37.32% (25.26%) at post-therapy. The mean difference from pre- to post-therapy was 5.18% (SD 17.05%); the standard deviation of the differences indicates the large variation between children in the difference from pre- to post-PISW. Based on these data the 95% CI for the mean difference in PISW from pre- to post intervention is −13.56% to 10.49% (see *Figure 7*).

Principal component analysis
From the 13 candidate percentage identification measures, eight measures at pre-therapy were brought forward into a PCA. The PCA comprised the percentage identification of:

1. phrase initial words;
2. phrase final words;
3. initial singleton consonant voicing in words with initial consonant singletons;
4. initial singleton consonant place in words with initial consonant singletons;
5. initial singleton consonant manner in words with initial singleton consonants that have obstruent manner;
6. final singleton consonant voicing in words with final singleton consonants;
7. final singleton consonant place in words with final singleton consonants;
8. final singleton consonant manner in words with final singleton consonants that have obstruent manner.

Although the percentage identification measures of phrase final words, initial singleton consonants of obstruent manner, and final singleton consonants of obstruent manner were not derived from ≥20 observations, these were included. It was still important to gauge the potential influence of therapy on longer stretches of speech which require a higher level of breath co-ordination than for words in

TABLE 16 Predicted probability (95% CI) of therapy and initial and final singleton consonant VPM combination subtype strata

	Initial singleton consonants		Final singleton consonants	
	Pre-therapy	**Post-therapy**	**Pre-therapy**	**Post-therapy**
Voiced coronal obstruent (d, z, ð, dʒ)	0.38 (0.27 to 0.46)	0.45 (0.36 to 0.55)	0.27 (0.19 to 0.35)	0.41 (0.31 to 0.50)
Voiced coronal sonorant (n, j, r, l)	0.341 (0.25 to 0.43)	0.46 (0.36 to 0.56)	0.43 (0.34 to 0.52)	0.46 (0.37 to 0.55)
Voiced dorsal obstruent (g)	0.49 (0.37 to 0.60)	0.48 (0.35 to 0.58)	0.49 (0.34 to 0.63)	0.51 (0.37 to 0.64)
Voiced dorsal sonorant (ŋ)	NA	NA	0.37 (0.27 to 0.47)	0.33 (0.23 to 0.43)
Voiced labial obstruent (b, v)	0.43 (0.33 to 0.54)	0.41 (0.31 to 0.51)	0.35 (0.23 to 0.46)	0.43 (0.32 to 0.55)
Voiced labial sonorant (m, w)	0.45 (0.36 to 0.55)	0.53 (0.43 to 0.62)	0.30 (0.20 to 0.39)	0.28 (0.18 to 0.38)
Voiceless coronal obstruent (t, s, θ, ʃ, tʃ)	0.37 (0.27 to 0.46)	0.42 (0.32 to 0.52)	0.35 (0.27 to 0.44)	0.43 (0.34 to 0.52)
Voiceless dorsal obstruent (k)	0.49 (0.37 to 0.61)	0.50 (0.39 to 0.61)	0.38 (0.27 to 0.48)	0.487 (0.37 to 0.58)
Voiceless labial obstruent (p, f)	0.467 (0.36 to 0.58)	0.44 (0.33 to 0.54)	0.35 (0.24 to 0.46)	0.312 (0.21 to 0.42)
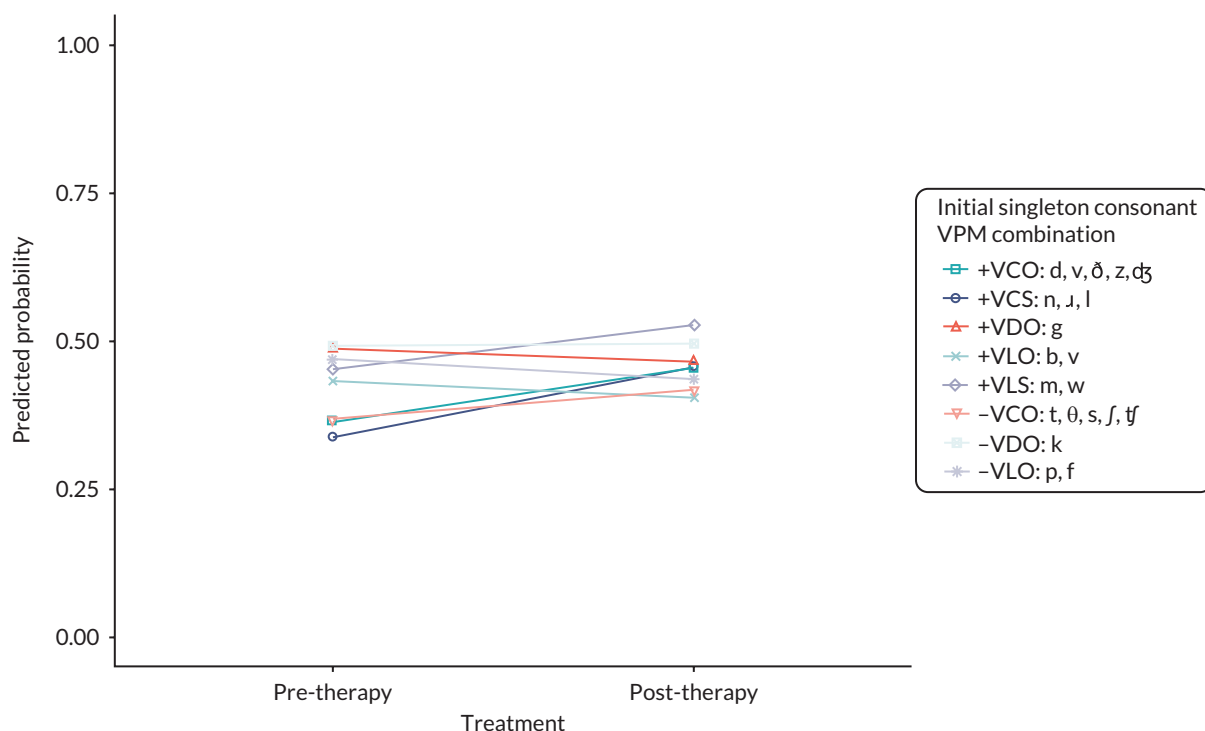
FIGURE 5 Interaction plot for therapy and initial singleton consonant VPM combination. +V, voiced; -V, voiceless; C, coronal; L, labial; D, dorsal; O, obstruent; S, sonorant. Phonemes represented by the different VPM combinations are displayed in the legends.
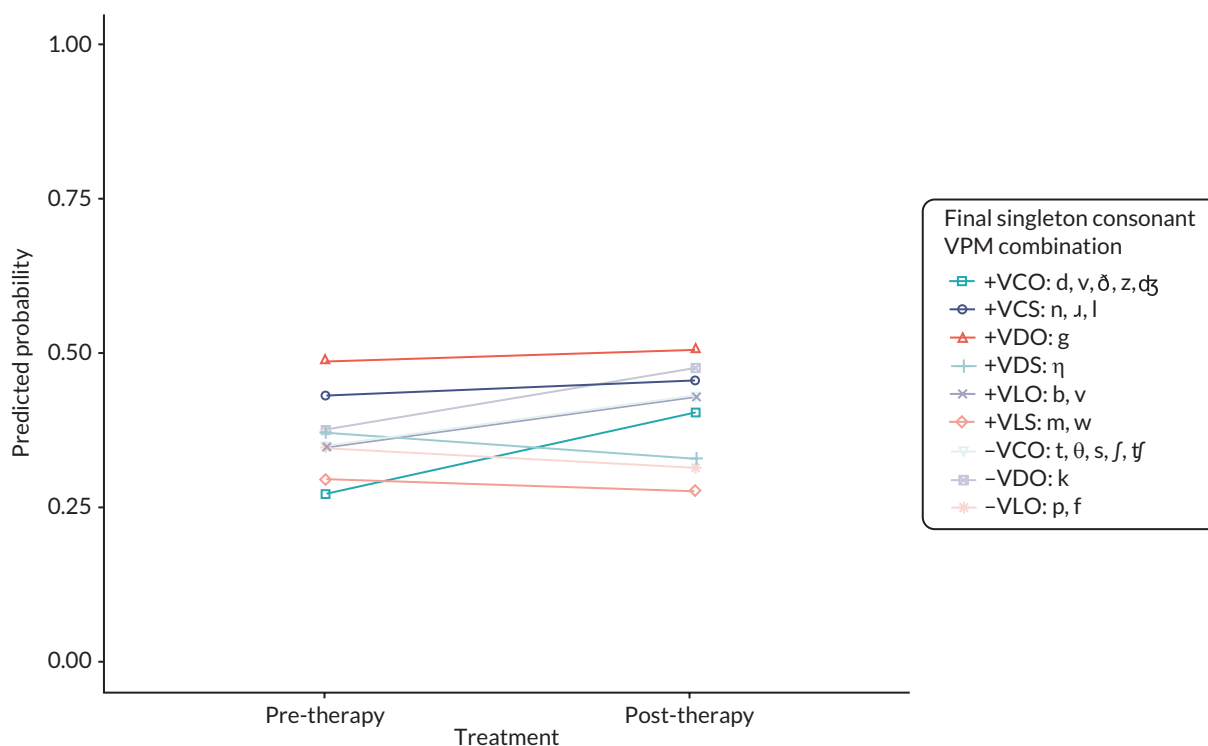
FIGURE 6 Interaction plot for therapy and final singleton consonant VPM combination. +V, voiced; -V, voiceless; C, coronal; L, labial; D, dorsal; O, obstruent; S, sonorant. Phonemes represented by the different VPM combinations are displayed in the legends.
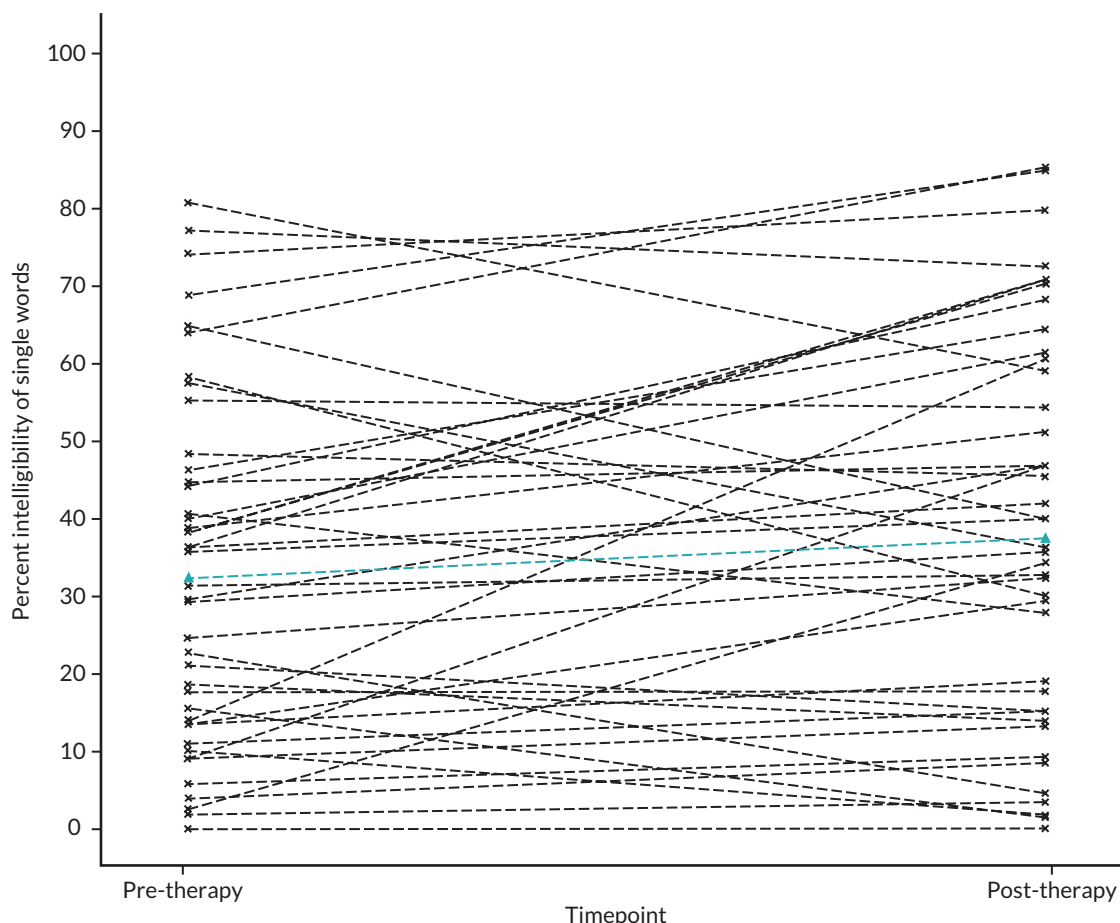
FIGURE 7 Spaghetti plot of individual change in percent identification of CS from pre- to post-therapy, N = 42. Each black cross and dashed line describes the change in a participant's PISW over time. Light blue triangles indicate the mean PISW at pre- and post-therapy.

isolation and to allow for comparability between the PCA results in the SW analysis. Complexity was not available in the CS data and one participant did not produce a polysyllabic word at post-therapy and therefore these were not used in the present PCA. The infrequent prevalence of words with clusters in the CS data also did not allow for inclusion of the percentage identification of initial consonant cluster.

The eight percentage identification measures were all strongly positively inter-correlated, supporting the use of PCA (see *Appendix 4*, *Figure 19*). The results suggested that one component was sufficient to explain 96% of the variance. The Kaiser-Meyer-Olkin measure of sampling adequacy was 0.91. The loadings indicates that all eight variables were positively related to the first principal component one. Descriptive statistics of these PI measures at pre-therapy and their loadings on the first component are provided in *Appendix 4*.

Radar plots

Similar to the analysis of the SW data, the same 6% identification measures (VPM of word initial and word final singleton consonants) were included to generate radar plots for each of the 42 children. Seven groups were formed from visual inspection of the individual plots. Most children had similar identification scores across VPM in word initial and final consonants pre-therapy, as demonstrated in the regular hexagons in the radar plots below. Groups differed in their initial identification scores and change in scores after therapy, and the identification of the whole word. Change in word identification and VPM identification showed similarities, due to the lack of consonant clusters in the CS data. Children with spastic, dyskinetic and spastic+dyskinetic type CP were found in each group. The four children with Worster Drought and ataxia were also spread across the groups. Plots for one member of each group are shown in *Figure 8*, as illustrative examples.
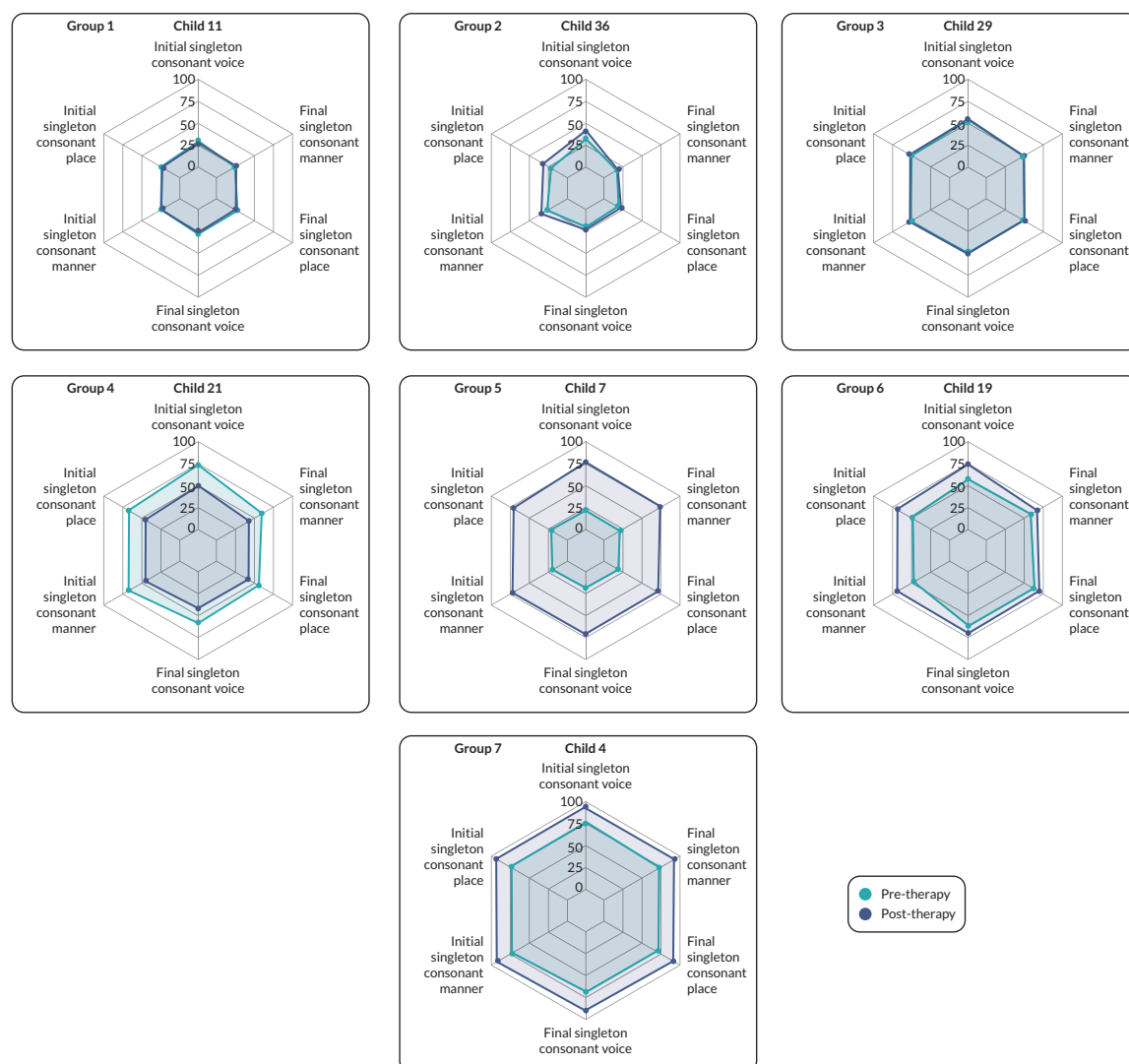
FIGURE 8 Example radar plots for CS groups.

Group 1 comprised four children (1, 11, 12, 34). These children had low identification of words in CS pre-therapy (0–18%) and low identification of the VPM of initial and final singleton consonants in words in CS (< 25% on all variables). They showed no visible change post-therapy in the radar lots and no change in word identification (–4 to 4%).

Group 2 also started with low levels of word identification (0–18%) pre-therapy and no/minimal change after intervention (0 to 4%). Percentage identification of initial and final consonant VPM pre-therapy (< 25%) but had small improvements post-therapy, usually across voice, place and manner characteristics but on *either* initial or final consonant rather than *both* word initial and final. This group comprised six children (6, 8, 28, 36, 40, 42).

Group 3 comprised five children (3, 9, 29, 46, 49). These children had higher word identification than Group 1 and 2 (24–77%), but again showed minimal change post-therapy (–5 to 8%). Their pre-therapy identification scores on word initial and word final consonant VPM ranged 50–75%, and no changes were observed in any feature post-therapy.

Group 4 comprised 10 children (13, 18, 20, 21, 24, 25, 30, 31, 38, 48). Their pre-therapy word identification scores ranged from 21 to 81% and decreased post-therapy (–22 to –3%). Similarly, their pre-therapy identification of VPM of word initial and final consonants ranged from 25 to 75%, and all

children had lower identification of VPM following intervention. Reductions for three children were noted for initial consonants only (13, 18, 20), with slight increases in final consonant voicing, place and manner observed. For other children, reductions in identification were observed across both word initial and word final consonants.

Group 5 comprised four children (7, 10, 14, 47), who started with low levels of identification of all VPM features pre-therapy (< 25%) but made clear improvements in the identification of all features post-therapy. The whole word identification started with low scores pre-therapy (2–14%) and increased by 15–46%.

Group 6 comprised ten children (2, 5, 15, 16, 19, 23, 26, 27, 32, 53). These children also made improvements post-therapy but started with higher pre-therapy word identification scores (29–46%) and VPM in word initial and word final consonants (25–50%) than Group 5. Seven of the children in Group 6 had slightly smaller change in final consonant VPM identification than word initial.

Group 7 comprised three children (4, 17, 22), who had high levels of identification of words in CS (64–74%) and of initial and final consonant voice, place and manner (approximately 75%) and whose identification scores increased post-therapy across whole words and all voice, place and manner features in both word initial and word final consonants.

Change in percentage identification measures from pre- to post-therapy for connected speech

On average, there was evidence showing improvements in some measures of articulation, defined using percent identification outcomes, from pre- to post-therapy based on the CS data (see *Table 17*). Generally, the magnitudes of the changes in percent identification were much smaller in the analysis of the CS data compared to the SW data.

TABLE 17 Mean change in pre-specified percentage identification measures

Percent identification measure	n	Pre-therapy, mean (SD)	Post-therapy, mean (SD)	Mean difference (95% CI), post minus pre
Individual words	42	32.15 (23.02)	37.33 (25.26)	5.18 (–13.56 to 10.49)
Monosyllabic	42	32.64 (22.79)	37.95 (25.08)	5.31 (0.02 to 10.59)
Polysyllabic[a]	41	29.99 (25.43)	36.21 (28.17)	6.22 (–0.89 to 13.34)
Initial consonant	42	39.09 (23.64)	44.89 (26.50)	5.80 (0.06 to 11.54)
Initial consonant voicing	42	43.43 (22.66)	49.32 (25.09)	5.89 (0.57 to 11.20)
Initial consonant place	42	39.99 (23.45)	46.15 (25.91)	6.16 (0.46 to 11.85)
Initial consonant manner	42	41.44 (22.85)	47.66 (25.35)	6.22 (0.56 to 11.88)
Initial consonant obstruent manner[a]	42	41.90 (23.64)	46.86 (25.60)	4.96 (–0.81 to 10.73)
Final consonant	42	34.58 (23.51)	41.12 (24.23)	6.54 (1.88 to 11.21)
Final consonant voicing	42	39.79 (22.81)	46.19 (22.95)	6.40 (1.81 to 10.99)
Final consonant place	42	38.74 (22.44)	45.43 (23.09)	6.69 (2.07 to 11.32)
Final consonant manner	42	38.00 (22.87)	44.37 (23.25)	6.37 (1.75 to 10.98)
Final consonant obstruent manner[a]	42	36.46 (23.23)	44.50 (24.58)	8.04 (2.60 to 13.48)
Initial consonant cluster (two consonant length)[a]	38	27.13 (30.64)	29.38 (28.32)	2.25 (–5.29 to 9.79)
Final consonant cluster (two consonant length)[a]	42	30.38 (23.86)	32.90 (23.96)	2.52 (–4.48 to 9.52)

a The PI measure was not derived from a minimum number of 20 observations at either pre- or post-therapy for all children and should be interpreted with caution.

Chapter 5 Results: acoustic analysis of single word and connected speech data

Participants

Audio recordings from 24 children were included in the acoustic analysis of SWs, and 16 of these children were included in the analysis of CS. These represent a subset of the total number of children who participated in the earlier studies, and were the number of children whose recordings were able to be retained for further research. Of the 24 children whose SW speech was measured and analysed there was 1 child with a diagnosis of CP type ataxic, 4 with dyskinetic, 16 with spastic, 2 with Worster Drought, and 1 with dyskinetic+spastic CP. From the 16 children whose CS was measured, there were 4 children with a diagnosis of CP type dyskinetic, 11 with spastic, and 1 with mixed type CP.

Single words

The recordings of SW productions made pre- and post-therapy from 24 children were measured acoustically; in this section we report measurement of speech rate (the duration of words/number of syllables) and the median intensity of words. Overall, 13 of the 21 children whose SW identification increased post-therapy produced SWs with a slower speech rate (see *Figure 9*, panel A). Two of the three children whose post-therapy SW identification was lower produced slower speech.

The change in rate of speech is shown in *Figure 9* panel B for the groupings of children based on the change in perceptual variables (see *Chapter 3*, section 'Radar plots' for a description of these groups). All but one of the groups (Group 5) contain children who produce SWs at a slower rate as well as children whose speech rate does not change or increase. Children in groups 1 and 2 have the largest differences between their pre- and post-therapy speaking rates. For the majority of the children in groups 3–6 their
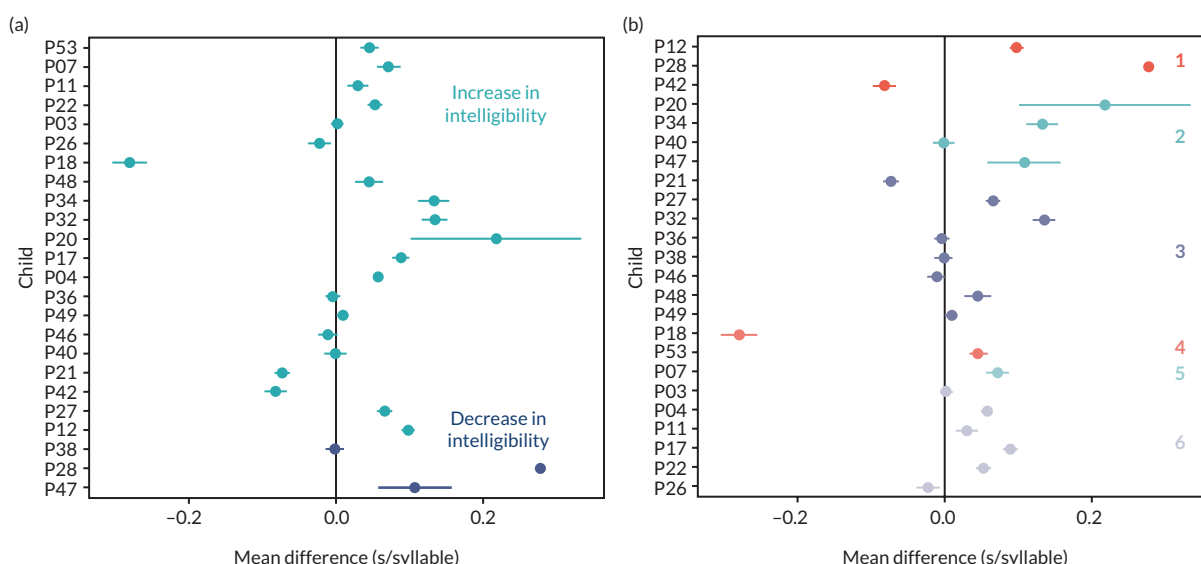
FIGURE 9 Mean difference in speech rate between post-therapy and pre-therapy for SWs. The mean difference and 95% CIs for the mean difference is shown for each child. Panel A: children ordered from lowest (child 47) to highest change (child 53) in SW intelligibility post-therapy. Children whose SW intelligibility increased post-therapy are shown in light blue, and children whose SW intelligibility decreased are shown in dark blue. Panel B: children grouped according to the six SW perceptual groupings, each labelled by number (see *Chapter 3* section 'Radar plots' for a description of these groups).

speech rate changes a little, apart from child 18 who produces speech at a much faster rate post-therapy [*Mean diff* = −0.280 s/syllable, 95% CI (−0.305 to −0.255 s/syllable)].

In the recordings of SWs taken post-therapy the majority of children were found to be producing words with a higher overall intensity. Fifteen of the 24 children produced more intense utterances post-therapy, as shown in *Figure 10* panel A (CIs of the mean difference were both positive). The child who had the largest increase in their intensity post-therapy [child 38, *Mean diff* = 16.40 dB, 95% CI (13.6 to 19.2 dB)] was in the group of children whose post-therapy SW intelligibility decreased. Three children who produced less intense speech post-therapy had increased SW intelligibility post-therapy (children 17, 20, and 48, CIs of the mean difference were both negative). The changes in intensity, according to the perceptual groupings (see *Figure 10* panel B), are not systematic. All groupings (except Group 5) contain children whose speech intensity increased or decreased post-therapy.

Change in intensity in obstruent and sonorant consonants in single words

The changes in intensity between pre- and post-therapy recordings were measured for each phoneme. In this section we report on the differences in obstruents and sonorant consonants according to their position within the word being spoken. The differences in the maximum intensity of obstruents according to word position is shown in *Figure 11*. Twelve of the 24 children produced obstruents in word initial position with a higher maximum intensity post-therapy. Of these 12, eleven belong to the group of children who had higher SW intelligibility post-therapy (see *Figure 11*, panel A). Nine children produced obstruents in word final position with a higher maximum intensity post-therapy, and again only one of these children belonged to the group with lower post-therapy SW intelligibility. The largest changes in maximum intensity of obstruents in either position was produced by children in groups 1 and 3 (see *Figure 11*, panels C and D). This pattern is consistent with the children who had the largest differences in overall intensity pre- and post-therapy (see *Figure 10* panel B).

The differences in the maximum intensity of sonorants according to word position is shown in *Figure 12*. Overall, there is less change in the intensity of sonorant sounds pre- and post-therapy compared to obstruents. Only seven of the 24 children produced sonorants in word initial position with a higher maximum intensity post-therapy. Of these seven, six belong to the group of children who had higher SW
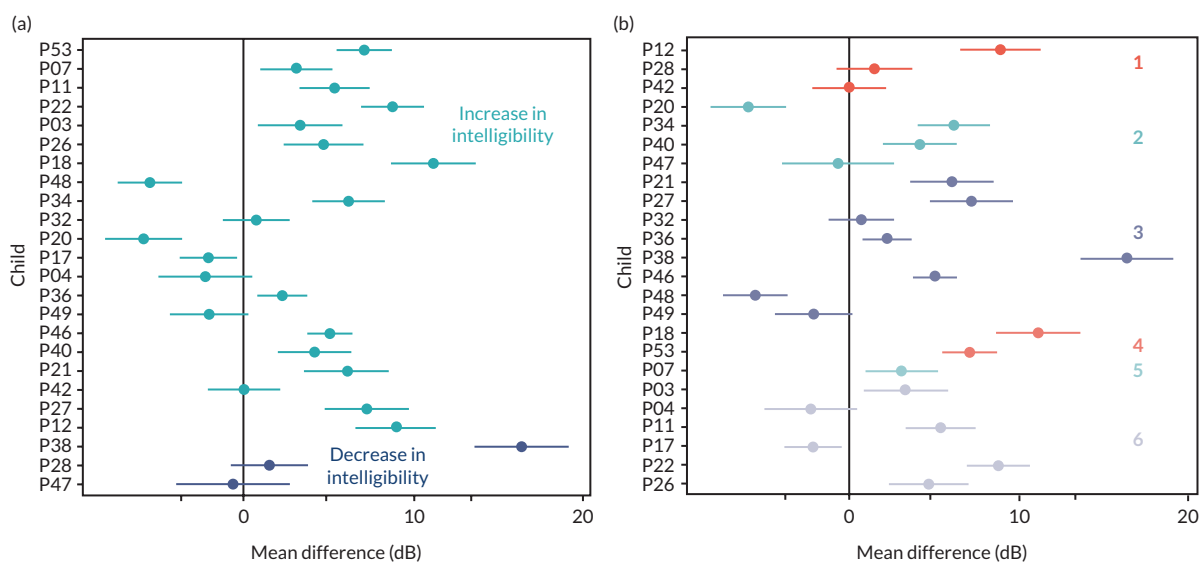
FIGURE 10 Mean difference in median intensity between post-therapy and pre-therapy for SWs. The mean difference and 95% CIs for the mean difference is shown for each child. Panel A: children ordered from lowest (child 47) to highest (child 53) change in SW intelligibility post-therapy. Children whose SW intelligibility increased post-therapy are shown in light blue, and children whose SW intelligibility decreased are shown in dark blue. Panel B: children grouped according to the six SW perceptual variable groupings, each labelled by number.
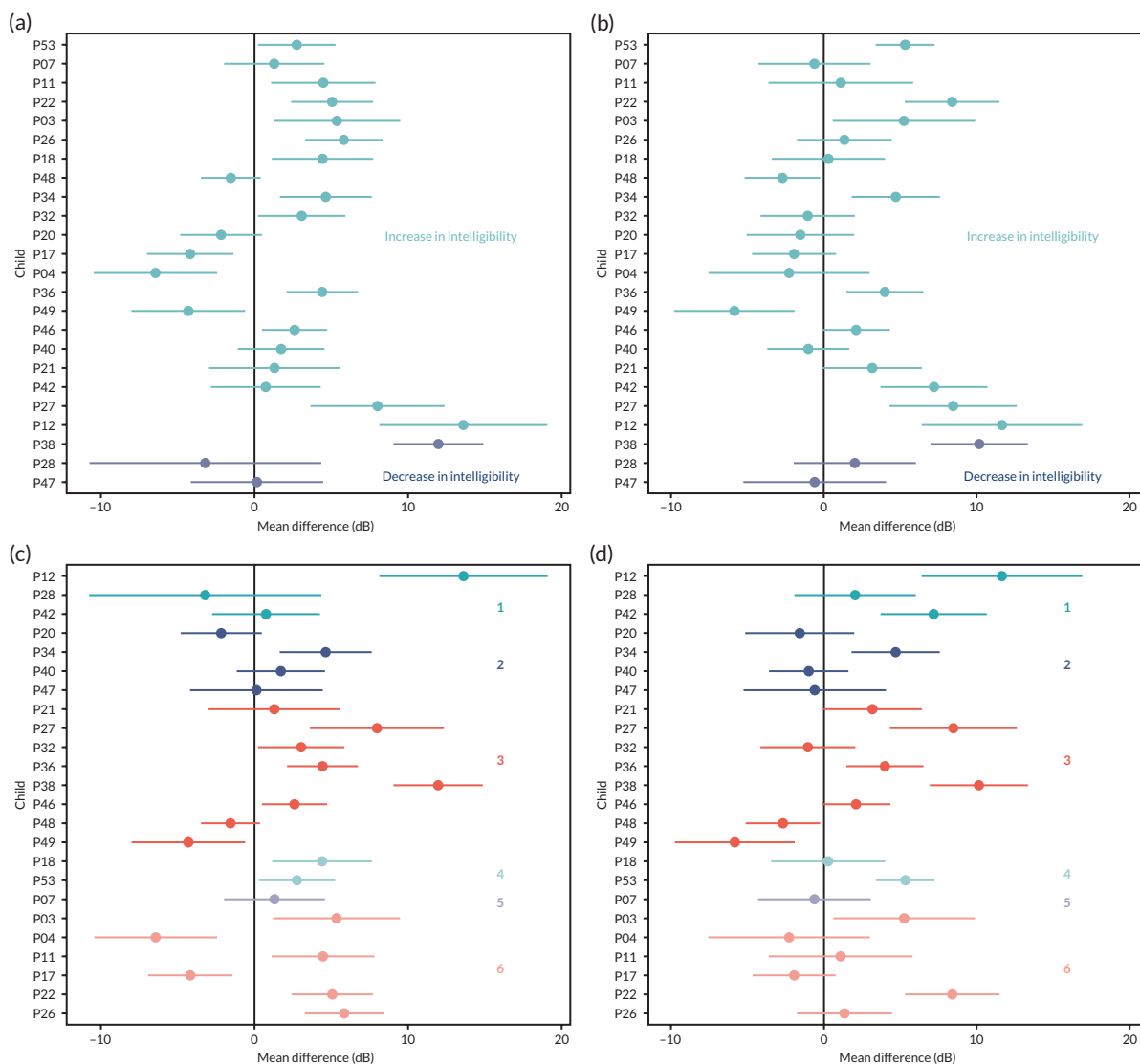
FIGURE 11 Mean difference in the maximum intensity of obstruent sounds in word initial and word final positions. The mean difference and 95% CIs for the mean difference is shown for each child. Panels A and B: the difference in word initial and word final obstruent intensity with children ordered from lowest (child 47) to highest (child 53) change in SW intelligibility post-therapy. Children whose SW intelligibility increased post-therapy are shown in light blue, and children whose SW intelligibility decreased are shown in dark blue. Panels C and D: difference in word initial and word final obstruent intensity with children grouped according to the six SW perceptual groupings.

intelligibility post-therapy (see *Figure 12*, panel A). Eleven of children produced sonorants in word final position with a higher maximum intensity post-therapy, including two children belonging to the group with lower post-therapy SW intelligibility. However, these are estimated from fewer observations than final position obstruent sounds as word final sonorants occurred less frequently in the SWs spoken by the children.

The largest changes in maximum intensity of sonorants, in either position, was produced by children in groups 3 and 4 (see *Figure 12*, panels C and D). This pattern is not consistent with the children who had the largest differences in overall intensity pre- and post-therapy (see *Figure 10*, panel B) but may reflect the typically higher intensity of sonorant sounds. Most children tend to produce sonorants with a higher maximum intensity post-therapy, but this trend is maintained in word final position sonorants (19 children in both cases).
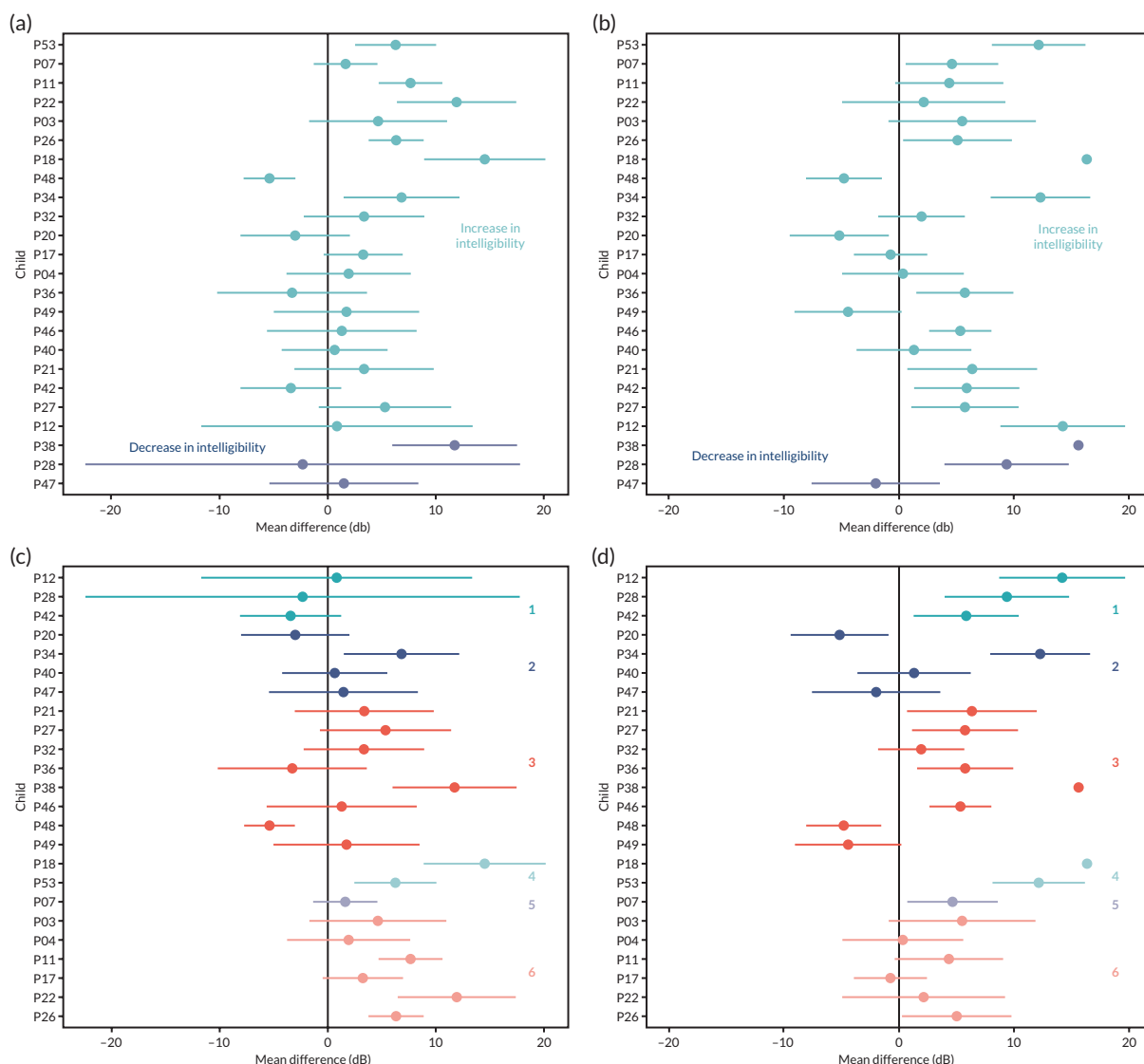
FIGURE 12 Mean difference in the maximum intensity of sonorant sounds in word initial and word final positions. The mean difference and 95% CIs for the mean difference is shown for each child. Panels A and B: difference in word initial and word final sonorant intensity with children ordered from lowest to highest change in SW intelligibility post-therapy. Children whose SW intelligibility increased post-therapy are shown in light blue, and children whose SW intelligibility decreased are shown in dark blue. Panels C and D: difference in word initial and word final sonorant intensity with children grouped according to the six SW perceptual grouping.

Connected speech

The recordings of CS made pre- and post-therapy from 16 children were acoustically measured. Overall, four of the 10 children whose CS intelligibility increased post-therapy produced speech with a slower rate (see *Figure 13*, panel A). One child whose CS intelligibility was unchanged post-therapy produced speech with a slower rate. Overall, five children produced CS at a faster rate post-therapy, and for three of these children their intelligibility decreased.

There is a consistent lowering of speech rate in the three children in group 1 (see *Figure 13*, panel B, see *Chapter 4*, *Radar plots* for a description of these groups). Most of the children who produce speech at
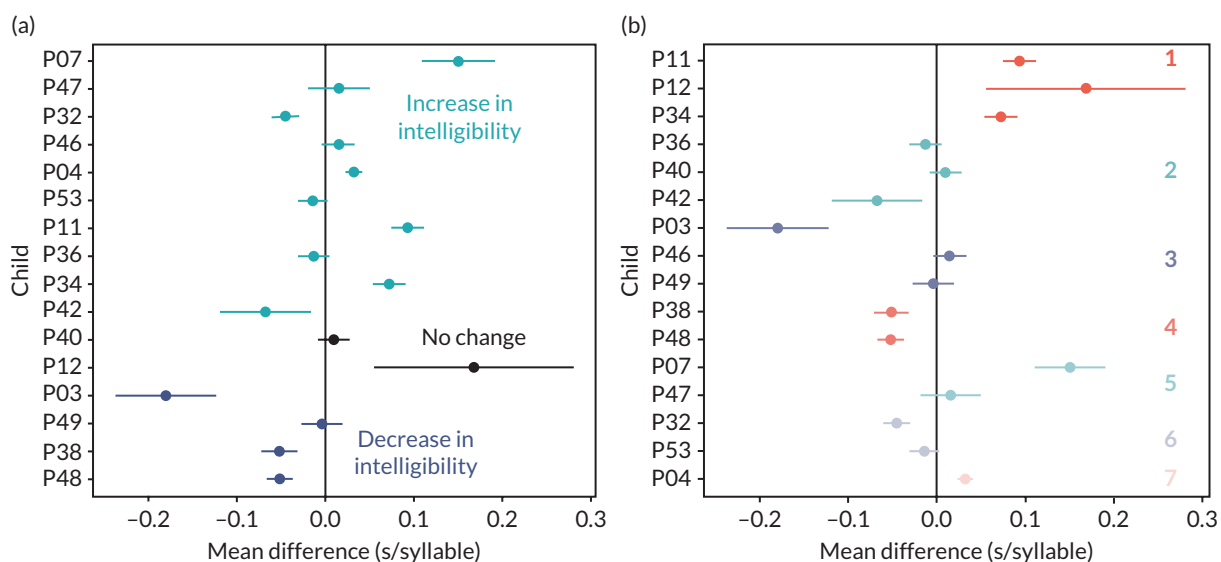
FIGURE 13 Mean difference in speech rate between post-therapy and pre-therapy for CS. The mean difference and 95% CIs for the mean difference is shown for each child. Panel A: children ordered from lowest to highest change in CS intelligibility post-therapy. Children whose CS intelligibility increased post-therapy are shown in light blue, children for whose CS intelligibility decreased are shown in dark blue, and children for whom there was no change in intelligibility are shown in black. Panel B: children grouped according to the seven CS perceptual variable groupings (see *Chapter 4*, *Radar plots* for a description of these groups), labelled by number.

a slightly slower or faster rate post-therapy are in groups 4 and 6. Groups 2, 3 and 5 contain individual children whose mean difference in speech rate is much greater than the rest of the group.

In CS, post-therapy, the majority of children were producing words with an intensity similar to that measured pre-therapy. Although 10 of the children had a mean intensity that was higher post-therapy, this difference was only significant for three children, who respectively had increased, unchanged and decreased CS intelligibility. Groups 1, 4 and 6 contain the largest range of intensity differences with groups 3 and 5 comprising children who had the smallest change in intensity post-therapy (see *Figure 14*, panel B).
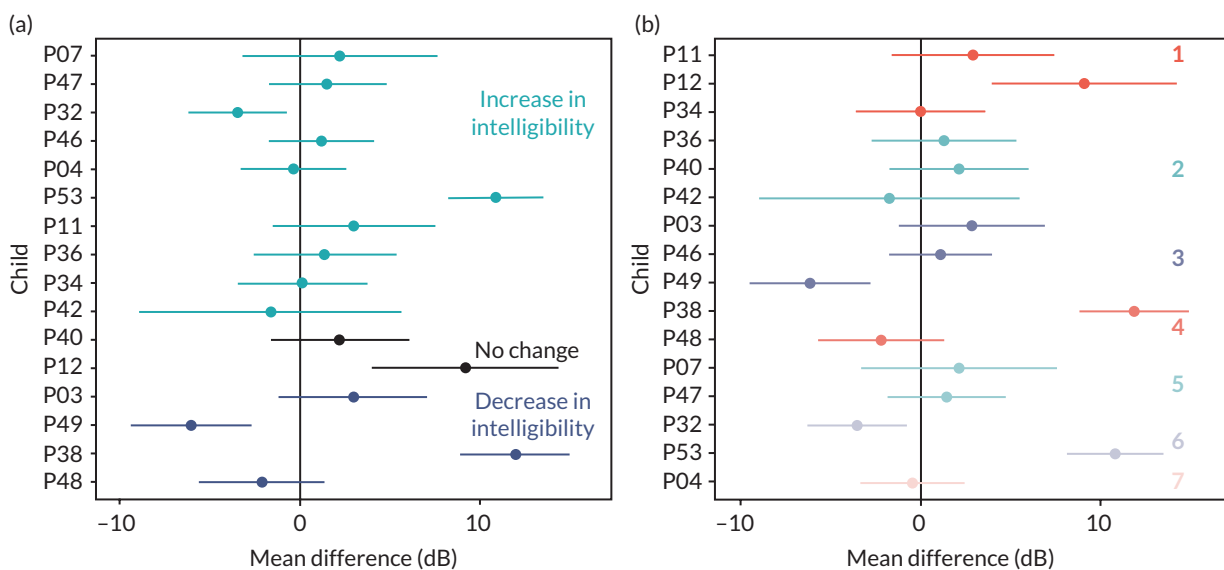
FIGURE 14 Mean difference in median intensity between post-therapy and pre-therapy for words in CS. The mean difference and 95% CIs for the mean difference is shown for each child. Panel A: children, ordered from lowest to highest change in CS intelligibility post-therapy. Children whose CS intelligibility increased post-therapy are shown in light blue, children whose CS intelligibility decreased are shown in dark blue, and the children for whom there was no change are shown in black. Panel B: children grouped according to the seven CS perceptual variable groupings, labelled by number.

Change in intensity in obstruent sounds in connected speech

The changes in intensity between pre- and post-therapy recordings were measured for each phoneme according to their acoustic characteristics in CS. In this section we report on the differences in obstruents according to their position within the word being spoken. Within the CS recordings many children did not produce words with either initial or final sonorants to enable those sounds to be compared across the different groups. The differences in the maximum intensity of obstruents according to word position is shown in *Figure 15*. Most of the children produced word initial and word final obstruents with similar intensity in pre- and post-therapy recordings (panels A and B). Unlike for SWs where children tended to produce obstruents with greater intensity post-therapy in both word initial and word final positions. For both word initial and word final obstruents children in group 1 tend to have the largest amount of change, however, for some children the number of observed obstruents in each position is small.
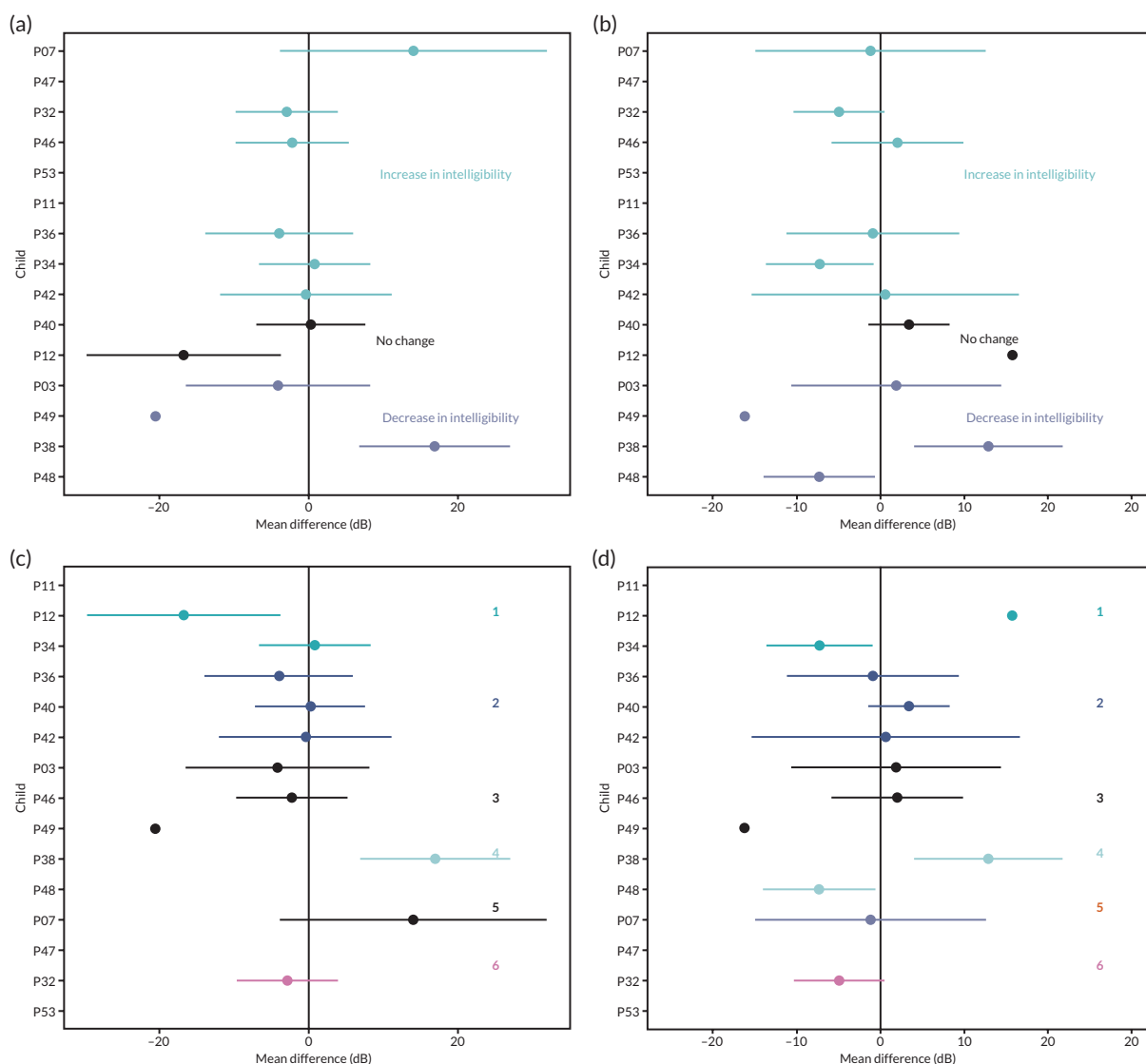
FIGURE 15 Mean difference in the maximum intensity of obstruent sounds in word initial and word final positions in CS. The mean difference and 95% CIs for the mean difference is shown for each child. Panels A and B: the difference in word initial and word final obstruent intensity with children ordered from lowest to highest change in CS intelligibility post-therapy. Children whose CS intelligibility increased post-therapy are shown in light blue, and children whose CS intelligibility decreased are shown in dark blue. Panels C and D: difference in word initial and word final obstruent intensity with children grouped according to the seven CS perceptual groupings.

Comparison of change in single words and connected speech post-therapy

The change in the speech rate and speech intensity between the SW and CS recordings is show in *Figures 16* and *17*. These scatter plots illustrate the amount of change in both variables according to the change in intelligibility, perceptual group or cluster to which the child belongs. For SWs, where the majority of children produce more intelligible speech post-therapy, most children also produce slower and more intense speech (13 out of 24). There is no obvious relationship between the absolute changes in these variables and the change in SW identification, nor do they correspond clearly with either the groups. There is, however, a greater spread of points in the SW speech measurements due to some of the children making larger changes in both variables.

The children with largest changes in SW intelligibility were those that produced slower speech with a higher intensity. Children who only changed the rate or intensity of their speech had smaller changes in SW identification. It is also clear that among these children, some who had the largest changes in
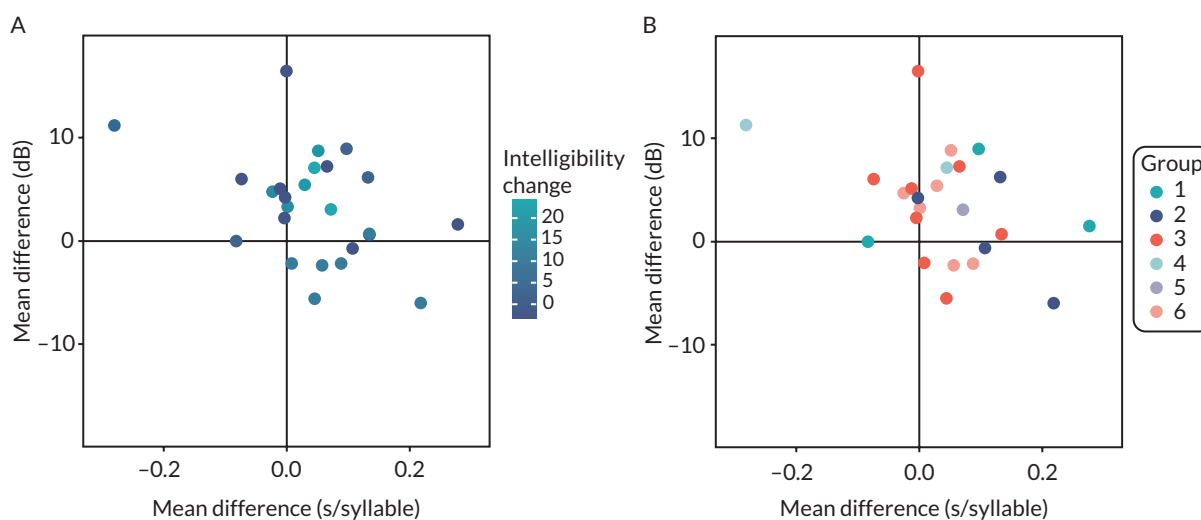
FIGURE 16 Scatter plots of the mean difference in speech rate and mean difference in intensity for the speech of 24 children when speaking SWs. Panel A colours each point according to the change in SW intelligibility post-therapy, where a lighter colour indicates a greater increase in intelligibility. Panel B colour the points according to the SW perceptual grouping the child belongs.
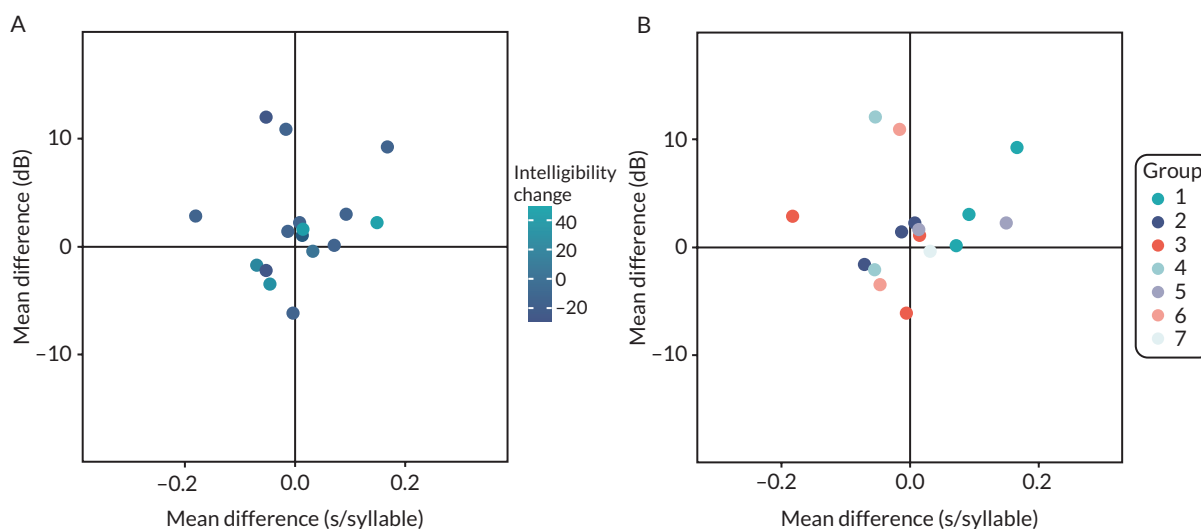
FIGURE 17 Scatter plots of the mean difference in speech rate and mean difference in intensity for the speech of 16 children producing CS. Panel A colours each point according to the change in CS intelligibility post-therapy, where a lighter colour indicates a greater increase in intelligibility. Panel B colour the points according to the connected perceptual grouping the child belongs.

these acoustic measures did not have corresponding changes in their SW identification. For instance, children in groups 5 and 6 had the largest changes in SW identification, but in terms of speech rate and intensity had only small changes in their post-therapy speech. Whereas the children in groups 1, 2 and 3 exhibited a larger amount of change in rate and intensity but had a smaller overall change in their SW identification.

There is a lack of a clear relationship between change in intensity or speech rate and changes in intelligibility of CS (see *Figure 17*). Here the majority of children shown in *Figure 17* do produce more intelligible speech but there are smaller changes in intensity and rate than for SW speech. A minority of the children have a slower speech rate and increased intensity (7 out of 16 children). However, in the context of SWs and CS the majority of children produced more intelligible speech post-therapy, but not all those whose CS is more intelligible change on both intensity and speech rate.

As with the results for SWs, children who had some of the largest changes in their speech post-therapy did not have a corresponding increase in intelligibility (see children in groups 1, 2 and 3 in *Figure 17* panel B). The children whose CS intelligibility was lower post-therapy did, however, have the common characteristic of having an increased rate of speech post therapy. The majority of the children whose CS intelligibility was higher post therapy (groups 5, 6 and 7) produced slower and more intense CS.

Chapter 6　Patient and public involvement in the study

The Speech Systems Approach was developed and evaluated in consultation with young people with dysarthria, parents and speech and language therapists. The parent carers of young people who took part in our previous research,[50,52,53] and a network of speech and language therapists, were consulted about the secondary analysis of data for this study. They supported the aim of the research and were strongly supportive of further analysis of the data which could lead to personalisation of the therapy programme.

Aim

During the study we consulted with young people, parent-carers and speech and language therapists. The aims of the Patient and Public Involvement in the study were to help interpret the results and to advise on the potential optimisation of the therapy protocol.

Methods

Participants
We recruited:

Five young people with neurodisability (two female; three male) aged 11–19 years. One young person had CP and dysarthria; all had friends and peers outside the group who had dysarthria.

Five parents (four female; one male) of young people with dysarthria aged 9–19 years.

Five specialist speech and language therapists working with young people with special educational needs in England (all female) were recruited from a local authority youth voice group, parent carer networks, and clinical networks respectively.

Procedure
We had planned to consult with patient and public involvement (PPI) groups three times during the project: once at the beginning to introduce them to the study and its aims; once to discuss the analysis plans and preliminary results; and once at the end to discuss the results and plans for refining the therapy. The lay member of the Steering Group, who is a parent of a child with dysarthria and teacher of pupils with special educational needs, advised that three workshops would be unnecessary for the project and that one consultation on the aims, results and therapy development would be a more efficient use of young people and families' time for a secondary analysis, especially as the statistical plan was complex. The speech and language therapist member of the Steering Group agreed, and we changed plans to consult with each group of stakeholders once, at the end of the study.

The young people's group took place face-to-face and was supported by three adults who knew the young people. The interviews with parents were individual and took place online via Teams; one parent was not able to attend due to last minute family commitments. Two groups were planned with speech and language therapists also online via Teams. One therapist was not able to attend due to illness; consequently, one individual and one group interview were conducted. Lindsay Pennington conducted all interviews, which lasted 40–75 minutes. She described the therapy and proposed mechanism of action, using images to support the explanation. She also described the results of the study, explaining that improvements were more consistent in SWs, where most sounds were easier to understand at

the start and end of words as singletons and in clusters; change in CS was inconsistent: some children improved, some were more difficult to understand.

All participants were remunerated at rates recommended by the NIHR.

Results

All groups recognised the description of speech production of young people with dysarthria. They agreed that breath control is often a challenge and that young people's voice often sounds weak and that clarity of individual sounds reduces in phrases, especially in longer utterances. They also agreed that individual consonants could be difficult for children and young people with dysarthria to produce, but which consonants are most difficult varies from child to child. The advisory groups agreed with the research team's interpretation of the results of the study and the suggestions for areas to focus on in future trials. They thought that children and young people would be able to understand instructions to use a big/strong voice to end of words. For example, '*Make it strong to the end*'.

Young people and parents were surprised to see the extent of the variation in children's response to the therapy in CS and suggested that individualisation of the therapy should focus on the use of the target voice in phrases and longer utterances early in the therapy programme. They also suggested that it would be helpful to have more frequent measurement of speech change within the programme as the type and extent of perceptual and acoustic change could inform individualisation. For example, if breath support can sustain only a small number of syllables taking a breath after each phrase, for example, [breath] 'the man' [breath] 'is running' [breath]; or if the child has greater breath supply, focussing on maintaining a steady rate in any length of utterance. The variation in response was not unexpected for the speech and language therapists, who reported observing similar difficulties in transferring behaviours from SWs to CS in their own clients.

PPI recommendations for refining the intervention

- use visual representations of words/words in phrases to show targets, for example, the start and end of words;
- use young people's interests in therapy exercises. For example: picture naming words could all relate to horse riding or dance; CS could involve descriptions of a favourite activity; and generalisation activities could include using the new voice in these activities with familiar people outside the family;
- use video clips to elicit connect speech in therapy and for outcome measurement for example, funny YouTube clips, TikToks of animals;
- use visual feedback to show the loudness of speech achieved against the target loudness;
- use emojis to give feedback;
- provide feedback on results at the end of each week, for example, '*Your voice has been strong. I have heard the end of words clearly when you have been naming pictures/describing pictures/when we have been chatting about your horse riding this week*';
- try individual strategies to help children relax and focus whilst acquiring and practicing their target voice. Young people talked about playing music to help them concentrate;
- create communities for young people trying the therapy so that they can practice together and provide peer support. Provide adult support to moderate groups and to ensure safety online;
- focus on CS to ensure transfer of skills. Move to CS as quickly as possible in the therapy;
- encourage young people to video their practice on a smart phone. Videos could be used as a diary and a record of change. Videos could be shared with the therapist.

Discussion

PPI informed the interpretation of the results of this secondary analysis and the recommendations for refining the therapy in future trials. The single sessions comprising aims, results and refinement of the therapy programme worked well. The topics were covered within an hour and participants described similarities in results in their own, their friends', their child's, or their clients' speech, suggesting generalisation of the findings. The PPI had clear benefits for the refinement of the therapy programme and its evaluation in terms of content and methods. All suggestions were taken forward in the recommendations for therapy development. Suggestions to focus on CS as soon as possible in the therapy mirrored the research team's interpretation of the results and stressed the transfer of skills to real communication outside the therapy. Most of the young people's suggestions for making the therapy fun and maintaining motivation had not been considered previously. Favourite activities had been used to elicit conversation towards the end of the therapy and in 'reward activities' such as picture reveals when the children had to guess what was behind uncovered squares. Greater application would develop personalised vocabulary sets of varying articulatory complexity [e.g. horse (complexity score = 4) saddle (complexity = 9) dressage (complexity = 13)] and may facilitate generalisation of skills. Using YouTube and TikToks to elicit CS is novel in SLT research and would provide access to wealth of material that would appeal to young people and increase the relevance of the therapy activities.

Young people and parents have agreed to continue their involvement and advise on dissemination of results to young people and their families.

Chapter 7 Discussion

Summary of findings

Our previous research showed that most children's speech intelligibility increased following the Speech Systems Approach and maintained this improvement in the short and medium term (6 and 12 weeks).[50,52,53] Children and their parents also reported that their speech was more understandable outside therapy sessions, and this had positive impacts on their family, education and social lives.[73] The current study analysed speech data collected at 1-week pre- and 1-week post-therapy to investigate the mechanisms underlying the intelligibility change. Our results suggest that intelligibility gains are due to a stronger vocal signal and an appropriate speech rate, which allows children to articulate individual sounds with greater precision. Following this intervention, listeners were able to identify initial and final singleton consonants and consonant clusters, across mono- and multi-syllabic words. Acoustic analyses suggest overall, that sounds within words were more intense (vocally stronger) and longer. There were marked individual differences in both the proportions of consonants identified correctly in children's speech pre-therapy and the degree of change in post-therapy scores. Individuals also differed in their patterns of acoustic change. Improvements in CS were more modest. There were greater differences between children in CS, with some children improving, some exhibiting no change, and some proving more difficult for listeners to understand after the therapy.

Change in single word speech

While previous research had demonstrated modest suprasegmental improvements in the children's voice quality and intonation,[16,37] this project was the first attempt to examine the potential impact of therapy on segmental aspects of the children's production. Our results demonstrate that therapy focussed on breath support can have a positive impact on the precision of articulation on individual sounds and their subsequent identification by listeners in SWs. This is because improvement to air initiation afforded by therapy based on breath support has a positive knock-on effect on phonation and articulation.

Phonation requires sufficient air pressure to either keep the vocal folds wide apart for the production of voiceless sounds or to activate the Bernoulli effect required for continuous vocal fold vibration. Improved breath support helps achieve the required air pressure, and our acoustic results show that this was manifested in increased intensity for both voiced and voiceless obstruents for the majority of children (18 out of 24). Improved respiration and louder voice also enable stronger articulation of individual words and sounds. Perceptually, this was manifested in overall improvement in the identification of initial and final consonants and clusters post-therapy. Acoustically, this was manifested through increased word duration and intensity for the majority of children whose acoustic data was available (17 out of 24), in line with other studies implementing a strong voice.[49] Increased duration and intensity were evident regardless of increase in intelligibility and did not show clear patterns for distinguishing subgroups of speakers. The same patterns were found when the manner of articulation was examined. The majority of children increased the intensity of obstruents post-therapy (16 out of 24), with maximum improvements being partly related to lower intensity pre-therapy. As we did not have the acoustic data for all participants in the study, these tentative patterns require further investigation. Future acoustic work on subphases of individual segments will also focus on whether improved respiration also aids improvements in the transitions between sounds, burst realisation for stops, timing of glottal and supraglottal events, and frication for fricatives and affricates. These acoustic characteristics allow listeners to better identify target sounds and subsequently the words that the children produce, as manifested in the improved identification of initial and final singletons at group level. Children whose intelligibility does not improve can then be offered more focussed therapy

on individual sound categories and transitions between sounds. Below we comment on patterns of identification for sounds/sound categories that were drawn from the perceptual analyses.

Initial position:

- It is not surprising that voiced labial obstruents remained the highest identified sounds pre- and post-therapy. Front articulations are relatively easier than articulations made further back in the oral cavity and benefit from added visual cues, and bilabial sounds have been observed to have highest levels of accuracy in other samples of children with CP and dysarthria.[74–76] Voiced obstruents in English do not typically involve vocal fold vibration during the closure duration, and are therefore phonetically voiceless, with a small and relatively lax glottal opening. This can be considered the 'default' state for phonation before aspiration for voiceless obstruents or full voicing (for languages that require it) are typically acquired. This, however, needs to be interpreted with caution as initial voiced plosives have also been observed to be pre-voiced in adults with CP,[35] with phonation starting before the plosive is articulated; this is unusual for English.[77]
- Similar to previous research,[74] voiceless coronal and labial obstruents, were the least identified due to the required co-ordination of laryngeal and supra-laryngeal events. They benefited the most from therapy, suggesting that breath control and rate intervention can aid the co-ordination and timing of phonation and articulation.
- The benefit of therapy for co-ordinating events in the vocal tract can also be seen in the relatively large improvement of labial sonorants /m/ and/w/, the first requiring coordination of airflow between the oral and nasal cavity, and second requiring double articulation at the labial and dorsal places of constriction.

Final position:

- Consistent with previous research, word final consonants were identified less frequently than those in word initial position.[74,78,79]
- The increase in odds of consonants being correctly identified was greater for those in word final position than word initial, suggesting better maintenance of intensity and articulatory precision across words post therapy.
- Our acoustic analysis also revealed that word final obstruents had lower intensity than word-initial obstruents despite an increase in their intensity post-therapy across all clusters of children.
- The improvement in dorsal /k/ identification in final (but not initial) position is an interesting one, as it follows developmental patterns in consonant acquisition, with labials and coronals appearing in syllable initial positions, and dorsals first appearing in syllable final position. It is clear, however, that dorsal articulation is still a struggle, with /ŋ/ and /g/ exhibiting no change in their identification after therapy while their labial counterparts benefited the most, across both obstruents and sonorants (m, w, p, b, f, v). Articulations that are further back in the oral cavity are generally acquired later than front articulation, and their production remained a challenge in word final position for the children with dysarthria in our study.
- No firm conclusions can be drawn about voicing patterns in final position. While there was an increase in intensity post-therapy, future investigation of VOT patterns in stop articulations pre- and post-therapy will enable us to ascertain improvements in the timing of voicing relative to the release of the stops in both initial and final position. This will require an examination of the timing of the start of voicing following obstruent release in order to achieve target-like patterns of voicing for English obstruents, which speakers with CP find challenging to differentiate,[5,35] voicing is expected to start shortly after stop releases for phonologically voiced sounds (e.g. b, d, or g) or following a period of aspiration for phonologically voiceless stops (e.g. p, t, or k). This mechanism is modulated by the size of the glottal opening, creating more tense articulation for voiceless and laxer articulation for voiced sounds of glottal and supraglottal events. In final position, vowel duration also plays a major role in cueing obstruent voicing, so investigation of timing of the vowel+final consonant will be important for understanding voicing patterns.

Word complexity played a role in correct identification in final consonants, supporting previous research that has shown that children with dysarthria and CP struggle to move between articulatory positions for target-like production of complex words.[42,76] This is not surprising given the required coordination of articulatory events in more complex words, and the fact that English phonotactic rules allow clusters in both initial and final position. Back sounds and multi-syllabic words add to the complexity. All aspects of word complexity improved post-therapy, including more cluster production in initial and final position, multi-syllabic word production, and initial and final consonant identification.

Our radar plots suggest that, expectedly, some children benefited more from therapy than others. This related both to their starting point pre-therapy, with a range of low to high performance in terms of segment identification, and to the degree of improvement, with two groups exhibiting clear improvement in most segmental features measured, two groups with small changes across some or all segmental measures, and two with little to no change post-therapy. That we see these differences is not surprising given the heterogeneity of CP in its varying impact on motor control and the fact that speech is one of the most finely tuned of all motor skills, depending on rapid, millimetre and millisecond precise movements within and across muscles from the diaphragm to the lips. Previous research has also shown variation between speakers with dysarthria associated with CP in terms of their acoustic patterns and listeners' identification of individual phonemes.[35,75,78] Our radar plots allow us to hone in on which individuals benefit the most from this universal-type therapy based on breath control and those who require further support with specific aspects of place, manner, and/or voicing coordination before any/further improvement can become evident. Given the multiple factors are involved in both speech production and speech perception, it is likely that detailed analyses of perceptual and acoustic measures for individuals is necessary to understand how speech is changing and if and when the change becomes overt and perceived by human listeners.

Change in connected speech

Initial and final consonant production also improved post-therapy in CS. The relatively lower number of initial and final clusters compared with the SW data and their uneven distribution across the participants did not allow us to test for improved identification post-therapy. Clusters are complex sounds which are acquired late by typically developing children.[80] In CS, initial and final clusters are likely to be adjacent to other consonants across word boundaries, making their production even more complex. This may lead to their avoidance in spontaneous speech.

Patterns of identification of VPM combinations in CS showed higher identification of dorsal obstruents relative to other consonant categories. This is in keeping with previous research involving speakers with CP and dysarthria,[74,76,78] and is reflecting the motor disorder rather than any developmental impact, with dorsals appearing late in speech development.[80] In contrast to SW speech, coronal consonants showed greater improvement in CS. Similar to previous research, obstruents were often misidentified pre-therapy.[74,76] Increased intensity and precision of movements may have allowed the children to vary the movement of the tongue tip to make finely graded movements to differentiate dentals, alveolar and post-alveolar (θ, s, ʃ) within the coronal place. Final labial consonants and voiced sonorants remained a challenge post-therapy. This suggests that the articulation of labials is relatively easier to achieve in initial than in final position. The voicing of sonorants also remains a challenge in final position; in contrast with voiced obstruents, which can be cued by the longer duration of the preceding vowel, the voicing for sonorants is mainly implemented in the consonant itself. Difficulty in sustaining vocal fold vibration till the end of the word can lead to voiceless realisations. Final devoicing is typical in acquisition and cross-linguistically in patterns of phonological variation and change.

Results suggest that improved identification of particular manners of articulation or voicing patterns (hypotheses 3 and 4) can therefore only be considered together with word position and place of articulation. There are VMP combinations that are optimal in particular word positions that facilitate

their target-like production pre-therapy and/or their improvement post-therapy. The patterns of identification in our data reflect these cross-linguistic tendencies and the multifactorial nature of speech change; it is therefore important not to consider each of the VPM components in isolation. This also explains why word position alone does not appear to show a therapy effect.

It is not surprising that, altogether, improvements in consonant identification in CS are much more modest for most children than for words in isolation. Coordinating articulatory gestures and timing is harder over longer stretches of speech and modulating air pressure and intensity to accommodate breath groups of various lengths and complexities may take longer to improve, particularly when the linguistic and speech planning demands are higher for CS than words in isolation. Children who make the most gains at this stage tend to be those whose pre-therapy identification is generally higher than those who do not, suggesting that a certain threshold for intelligibility is perhaps required before a therapy, based on breath support, can aid improvement in identification, or that earlier introduction of CS in therapy is warranted before improvement is evident.

Change in acoustic properties of CS are not clear cut, although this may be due in part to the small number of children whose CS was available for analysis. Similar to children who had received LSVT,[45] intensity seemed to be increasing but change was not stable, with only three of the 16 showing definite change. Unlike for SWs, where the children who had the greatest improvement in intelligibility produced slower and more intense speech, improvement in CS was less clearly defined. Comparing the amount of change in SWs and CS it is evident that acoustically a greater degree of change occurs in SWs. From the analysis of change in obstruent sounds in both SWs and CS the children were able to increase the intensity of these sounds in both word initial and word final positions to a greater extent in SWs. It may be that for some children the amount of change they experienced post-therapy in CS was evident acoustically but not sufficient to lead to increased perceptual accuracy.

Contrary to other studies involving young people with dysarthria associated with CP,[81] we found no consistent relationship between SW and CS intelligibility. Children in this study who scored high in SW identification did not necessarily do so in CS. Similarly, there was no clear relationship between change in speech conditions; children who made gains in word identification in CS did not always make similar improvements in SWs. There was little overlap between young people in the clusters and radar plot groups derived from SWs with those from CS. The disconnect between SW and CS we observed may be due to the methods used. Previous research has elicited speech in SWs and phrases through repetition.[81] Our studies sought to maximise ecological validity by eliciting spontaneous speech in picture description. Pictures used to elicit CS were selected to appeal to the participant children, therefore they were interested in describing them and so that subsequent utterances were unlikely to be 'guessed' by earlier content, that is, minimise top-down processing. Pictures were not selected on vocabulary they were likely to elicit, that is, complexity, frequency, length or neighbourhood density of words produced.

Age and cerebral palsy type

Change in perceived segmental characteristics were not predicted by age or CP type. Unlike studies of young children's intelligibility, where differences across ages were observed,[82] participants in our study did not appear to vary in their speech production. The lack of effect of age may be due to young people having completed their speech development, which usually occurs around 5 or 6 years of age.[83] The lack of influence of CP type supports the findings of previous research,[5,75] which suggested greater similarities than differences in perceptual characteristics across CP types. No effects of CP type were observed in either SWs or CS, and children with each different types were found in all clusters and radar plot groups. However, results should be interpreted with caution given the small numbers of children with each motor disorder type, especially Worster Drought and Ataxia.

Strengths and limitations

This secondary analysis involved 42 children and young people who had participated in previous research. The sample included children from across the north of England with a range of ages, types of CP and cognitive skills, who attended mainstream and special education and who represented the diverse speech and language characteristics of children receiving SLT in the NHS. To our knowledge, this study is the first to link perceptual and acoustic data to examine how and why speech changes following intervention.

Two of the three studies[52,53] that contributed data to this secondary analysis were interrupted time series and had no control group. The 42 children across the three studies, who had received the therapy, were included in this secondary data analysis. The estimates of the effect of therapy reported were therefore derived from pre-post therapy data and there were no pre-post estimates for children from a control group against which to compare the observed estimates; which limits the strength of conclusions we can draw about the effects of the therapy.

The number of children with each type of CP was low and although previous research has shown more similarities than differences in the speech patterns of people with different types of CP,[5,74] our failure to detect differences between children with different types of CP may have been due to insufficient power. Information on the site of neurological damage or maldevelopment was not available from the original studies but could help elucidate speech phenotypes in future studies. The acoustic analysis only included 24 children for SWs and 16 for CS, due to restricted access to the recordings for secondary data analysis and data incompatibility. This limited our ability to examine phonetic patterns within groups of children formed from the radar plots or by CP type. Acoustic data serve to support patterns of change for individuals better than groups. However, with the heterogeneity observed in this and other samples of speakers with CP and dysarthria an individual approach is recommended going forward to understand the mechanisms of action underlying intelligibility change for children with these disorders. The initial preparation work for acoustic analyses was very challenging for our acoustic models, which were trained on adult speech. The initial steps required checking gold transcriptions, which are word- and utterance-based, against actual phoneme realisation before automatic segmentation and labelling could be carried out. Following that another round of manual checking was required of the segmentation by trained phoneticians, which was labour intensive. As expected, the acoustic models struggled the most with CS due to lower accuracy/target-like realisation but also the presence of speech connected processes which challenge segmentation that is phoneme-based. As a result of these challenges, our acoustic analysis focussed more on word- and segment-based measures, while acoustic measures focussing on subcomponents of segments (e.g. burst amplitude, VOT, formant transitions) were left for future research. A potential limitation of the perceptual analysis may be that we did not include an additional random effect of listener. This was primarily due to the limitations of the data whereby listeners could not be uniquely identified. Despite this, a strong listener effect would not have been anticipated for two reasons related to the design of the source studies. Firstly, all participating listeners had no regular or specific experience of interacting with people with disabilities or speech disorders, thus limiting the between-listener variability in the ability to identify words spoken by children diagnosed with CP with moderate-severe dysarthria. Secondly, familiarisation effects were minimised in the original research by listeners hearing only three recordings, each from a different child.[84]

Another potential limitation of the perceptual analysis was that our regression models assumed a nested hierarchical structure; this assumption was made on the basis of the way the uttered words were elicited. However, a given target word could have been spoken by several children and so it would have been reasonable to consider a crossed random effects model to take account of this structure. We assessed the sensitivity of our results to the results from a model with crossed random effects, with children crossed with the target words.[85] We found that our conclusions remained qualitatively the same.

The choice of intelligibility measures has been debated[84,86,87] and all methods have notable limitations. The original research selected measures of word intelligibility, with a focus on children's success in conveying meaning. In this secondary analysis we have considered segmental intelligibility where the perception of constituent phonemes is measured. As we discussed above, the original methods met with limited success in capturing phonetic/segmental intelligibility. The original research used the CSIM[57] to elicit SW speech. The CSIM has 200 lists, each containing 50 words from a corpus of 600. Each item in the list is one of 12 phonetically similar words. For example, the first item is always one of the following: *torn*, *born*, *corn*, *door*, *floor*, *for*, *form*, *horn*, *sore*, *storm*, *swarm*, and *warm*. Listeners hear a word and select that word from the list of 12 phonetically similar words. This method of intelligibility testing has the advantage of allowing analysis of multiple types of errors, showing which features of sounds that are perceived correctly. For example, if listeners select 'born' for the target 'torn', their error suggests that the plosive sound has been recognised correctly but not the voicing or place of the initial consonant. This method may have greater ecological validity than minimal pairs, in which a choice of two words that differ in only one contrast presented.[88,89] However, the measure is not without its limitations. Firstly, not all feature contrasts appear in the CSIM foils. For example, to measure if all features of /t/ – voiceless alveolar (coronal) plosive – have been perceived correctly in 'torn', foils would need to include a voiced alveolar plosive ('dawn'), a voiceless labial and dorsal plosive ('pawn' and 'corn') and a voiceless alveolar fricative ('sawn') or affricate ('chorn' – although no such word exists in English), but only 'corn' is in the list of choices. Listeners may therefore have been forced to select a word that they did not perceive, and our estimates of sound identification may therefore be imprecise. Secondly, the original research allocated different lists from the CSIM to each child at each recording and allocation of the same list to multiple children in the studies was minimised. This limited learning effects for both the speakers and the listeners, but also meant that the target words differed across children and across recordings creating a source of between-child variability. Some words and consequently sounds appeared much less frequently than others in the data set. We were forced to combine places of articulation (e.g. bilabial and labiodental; dental, alveolar and post alveolar) and manners of articulation (e.g. plosives, fricatives and affricates in an obstruent class) to provide sufficient data for the regression models. A further restriction for the model is that sonorants cannot be voiceless in English. Consequently, we would not have been able to separate the main effects of voicing and manner had we included them as individual predictors in our GLMMs. Thus, we took the decision to combine the three characteristics to meet data analysis requirements. Although the combination prevented us from making examining the effect of the intervention on individual manners and places of articulation, our categories allowed us to examine the impact of therapy on sounds made with the lips, anterior and posterior tongue, and those that demand the tightest and finely modulated constriction (obstruents). Our method of analysis generates some nuance over what sort of consonants benefit from the therapy. For example, results demonstrate that consonants with dorsal place are more likely to be identified after therapy but under limited conditions with regards to manner and voicing and word position. Future acoustic work will enable us to look at transitions between sounds in order to understand if subchanges in the consonants at word edges or in the vowels surrounding them are changing even where identification is not, or whether these cues can be used in therapy to aid intelligibility.

Finally, some words in the CSIM appear infrequently in spoken British English. Low frequency words may not have been known to some children in the studies and in effect be 'nonwords' for these children. Production of non-words involves imitation, whereas words that are known to children will have a stored phonological representation which may be used to plan and execute the movements when they are spoken. The effect of lexical representation remains a confounding variable in the current study, as word knowledge was not assessed in the original research.

Word identification in CS can be influenced by phrase position or the identification of other words in the phrase, allowing listeners more guess work if they have identified other parts based on semantic and syntactic cues in their top-down processing. This makes is harder to pin particular improvements (or lack thereof) to therapy and is an expected challenge in the analysis of free transcription of CS. The pictures used to elicit CS were selected to appeal to a broad age range and to have sufficient complexity

that they would prompt children to produce multiple utterances containing different vocabulary. In this way words within the later phrases in the CS samples were unlikely to be aid by additional context.[81] However, the pictures were not selected on the linguistic features of the words they were likely to elicit, such as the presence of consonant clusters or neighbourhood density. Our observed lack of clusters in CS and lack of effect of neighbourhood density may be an artifact of the task rather than a true reflection of children's usual intelligibility in CS. The children may also have actively avoided producing clusters, which is under their control in CS but not in word-elicitation tasks. Other research has used sentence repetition, with increasingly longer sentences to stress the respiratory system.[42] However, repetition may encourage children to mimic speech patterns limiting ecological validity. There are also learning effects to consider – both for the children and listeners. If using the same sentences across children and recordings more listeners would be needed to avoid anticipation of phrases and priming word perception. Composite pictures have been developed to elicit words with all phonemes of English (Park Play),[90] and possibly for other languages. Questions could be developed to elicit phrases with these target words whilst still allowing children to generate spontaneous spoken language.

A further limitation of the CS task was that listeners were asked to transcribe the audio recording verbatim, and some may elect to omit words they could not identify entirely from their transcription. This is in contrast to the SW task where, by virtue of being a forced choice task, information on perceived consonants and clusters is available even if a listener could not fully identify the target word. Owing to this, information regarding the identification of components of a word such as its consonants and clusters were restricted to occasions where the listener was able to perceive a whole word. As a consequence, the proportion of target words with an accurate identification of consonants and clusters may actually be higher than what was recorded. This may explain why the estimate of the effect of therapy from the GLMMs in the CS analyses were smaller than those estimated from the SW analyses.

Lastly, children received different CSIM lists and pictures to elicit CS at pre- and post-therapy. Thus, the estimated effects of the therapy may be partially explained by the differences in the spoken language elicited. However, we may still be confident in our conclusions as the CSIM lists are phonetically balanced and an inspection of the median number of observations for the various initial and word final consonant characteristics in the CS data does not suggest a substantial difference in production from pre- and post-therapy.

Equality, diversity and inclusion

Equality, diversity and inclusion of the research is considered in terms of the research team and the PPI undertaken, as characteristics of participants other than presented here are unknown. The research team comprised four established academics and two early career researchers; three male and three female; three from the UK and three from overseas. One of the statisticians is an early career researcher and was part of the team throughout. They worked closely with a senior academic to develop a statistical analysis plan and undertake the analysis of perceptual data. They were an integral part of the team, took part in all regular team meetings and were key in the decisions the team made throughout the project. They have written parts of the report and will be co-authors on papers and presentations leading from it. Internationally there are few biostatisticians who work on speech data and understand the multifactorial nature of speech and language. Statistician input is vital in developing rigorous analytical research in this developing field and future applications are planned by the team to progress this collaboration to develop expertise. The second early career researcher is a phonetician, who worked closely with Cunningham to prepare recordings for automated acoustic analysis and manage the work of the student transcribers on the project. They joined the meetings in the latter part of the research, observed the decision-making and will work on acoustic papers arising from the project. They are also part of work to develop further bids for funding as a potential co-investigator.

The PPI undertaken included young people, three male and two female, with a range of neurodisabilities and from diverse ethnic and socioeconomic backgrounds. The parents were also from diverse backgrounds, some spoke languages other than English. The speech and language therapists, like the majority of the profession in the UK, were all female and White British. The PPI garnered views of the results of the current study and advised on plans for individualising the therapy. The young people thought that use of social media, such as YouTube clips and TikToks, would appeal to their peers and reported that most young people they knew had access to and used smart phones or computers for their social lives. However, digital poverty has potential to exclude some young people and families from future studies. It was notable that access to IT was inequitable during the pandemic and children from poorer neighbourhoods were less likely to have access to computers for their schoolwork.[91] It is important that any future research provides hardware if necessary to engage in the therapy and young people and families are supported in their digital literacy.

Chapter 8 Conclusions and recommendations

Conclusions

The Speech Systems Approach has led to gains in intelligibility that have been maintained for up to 12 weeks without further intervention. This secondary analysis shows that improvements in intelligibility of SWs may be achieved by the generation of a stronger vocal signal and greater articulatory precision. Listeners were able to identify most categories of singleton consonants better in both word initial and final position. Similarly, clusters of consonants were better identified at both the start and end of words. We did not find a differential effect for obstruents, which demand the most precise constriction of the vocal tract, or for word final sounds, suggesting a general effect on speech intensity and articulatory precision in SW speech. Changes in CS were more modest and a small number of children were more difficult to understand after therapy in CS. This suggests that some participants continued to experience difficulties controlling breath supply and the precision of speech movements at phrase level. The marked individual differences suggest a varying response to therapy between children but also within children across speech segments. Given the heterogenous nature of CP and its diverse impacts on voluntary movement children with CP may benefit from individualised instruction to control speech intensity and rate to maximise their intelligibility. Testing the personalisation of the therapy is now warranted.

Recommendations for clinical practice

1. Given the overall response to the Speech Systems Approach and in accordance with the NICE guidance on CP in the 25's,[12] children and young people with dysarthria who use speech as their main means of communication should be offered the therapy as part of their care pathway. Augmentative and alternative communication (AAC) should be offered to ensure young people can communicate independently when their speech is not understood.
2. Practice should move quickly to CS to promote generalisation of techniques in everyday interaction. Additional methods of eliciting CS, such as TikToks and YouTube videos, should be considered, especially for older children.
3. Feedback should be individualised and easy to understand, for example, emojis. Summary feedback at the end of each week of therapy could aid motivation.

Recommendations for research

A programme of research is required to determine the most effective SLT pathway for children with nonprogressive dysarthria, whose communication challenges seriously impact their social participation, education and life chances in adulthood. This should include further research on the Speech Systems Approach, which is showing promise in early trials:

1. Personalisation of the Speech Systems Approach intervention, with cues adapted to the child's performance, should be investigated in a phase II study to investigate if the intervention can be enhanced prior to a randomised controlled trial. Acoustic analysis of speech during the intervention could aid personalisation, showing how speech changes in response to individual cues.
2. Long term effects of the one-off burst of intervention on intelligibility and social participation should be examined.
3. Differential effects of the approach should be examined for subgroups of children (e.g. by age, cognitive skills, motor disorder, sire of underlying neurological damage/maldevelopment) to determine who the intervention works best for and when it should be offered.

Specific recommendations for measuring speech outcomes include:

4. Intelligibility, through listener identification of words and constituent sounds, should continue to be assessed in SWs and CS to investigate speech clarity when breath support is minimised and in utterances closely reflecting conversational speech.

5. To test SWs, two matched lists of high frequency words all containing phoneme contrasts and word initial and word final clusters should be developed to facilitate comparison across children and time. Free choice response would allow listeners to report the sounds they thought they heard, including parts of words if they did not perceive the whole. Using one at 6-weeks pre-therapy and immediately post-therapy and the second list immediately pre-therapy and at follow-up would minimise learning effects. Such lists would allow both word and segmental intelligibility to be tested. Such data would also allow the use of algorithmic methods to estimate the similarity of sounds to their intended target sound.[92]

6. Methods of eliciting CS samples should be tested. Sets of pictures depicting words that have all speech sounds of a language may elicit the most natural speech but bring variation in actual words spoken and grammar. Repeating sentences of increasing complexity after a lag to avoid mimicking the modeller's speech pattern. The two methods should be compared in terms of the word and segmental intelligibility testing they confer.

7. Acoustic change should be measured to help understand how perceptual change is achieved by individuals and whether some changes are evident in acoustics before they are perceived.

Acknowledgements

We thank the following for their contributions to the study: The young people who participated in the original research that was analysed in this study, as well as their parents, speech and language therapists and school staff who supported them and the studies. The young people, parents and speech and language therapists who took part in the advisory groups, for their help in interpreting the results of the study and advice on tailoring the intervention methods and procedures in future trials. Dr Bruce Wang for his assistance with the preparation and analysis of the acoustic data. Newcastle University students for their assistance in measuring acoustic data and in coding and entering perceptual data; in particular Carol-Ann McConnellogue and Holly Fender. The External Steering Committee members: Dr Leah Li (Chair and Associate Professor Population, Policy and Practice Department, UCL GOS Institute of Child Health), Dr Anja Kuschmann (Senior Lecturer in SLT, University of Strathclyde), Ms Mel Ogle, Lay Member; Professor Johan Verhoeven (Professor of Experimental Phonetics, School of Health and Psychological Sciences, Department of Language and Communication Science, City University of London). Mr Alan Marshall, Research Manager from NIHR for his support.

Contributions of authors

Lindsay Pennington (https://orcid.org/0000-0002-4540-2586) (Professor of Childhood Neurodisability), Chief Investigator, led the design and delivery of the project; participated in all project stages; wrote sections of the report; and finalised the report.

Stuart Cunningham (https://orcid.org/0000-0001-9418-8726) (Senior Lecturer in Human Communication Sciences) led the analysis of the acoustic data and prepared the results for publication.

Shaun Hiu (https://orcid.org/0000-0003-1699-4348) (Research Assistant, Biostatistics) co-wrote the statistical analysis plan for the perceptual data; conducted the analyses; prepared the results for publication; and contributed to all sections of the report.

Ghada Khattab (https://orcid.org/0000-0002-8451-8135) (Professor of Phonetics and Phonology) contributed to the description of the phonetic and phonological concepts and notations, the analysis and interpretation of perceptual and acoustic data and co-wrote the report.

Vicky Ryan (https://orcid.org/0000-0002-7008-3193) (Lecturer in Biostatistics) was a co-applicant; co-wrote the statistical analysis plan for the perceptual data; reviewed and contributed to all sections of the report.

Funding

Ethics statement

Data-sharing statement

References

1. Rosenbaum P, Paneth N, Leviton A, Goldstein M, Bax M, Damiano D, *et al.* A report: the definition and classification of cerebral palsy April 2006. *Devel Med Child Neurol* 2007;**49**(suppl 109):8–14.

2. Sellier E, Platt MJ, Andersen GL, Krägeloh-Mann I, De La Cruz J, Cans C, *et al.*. Decreasing prevalence in cerebral palsy: a multi-site European population-based study, 1980 to 2003. *Devel Med Child Neurol* 2016;**58**(1):85–92.

3. Parkes J, Hill N, Platt MJ, Donnelly C. Oromotor dysfunction and communication impairments in children with cerebral palsy: a register study. *Devel Med Child Neurol* 2010;**52**(12):1113–9.

4. Nordberg A, Miniscalco C, Lohmander A, Himmelmann K. Speech problems affect more than one in two children with cerebral palsy: Swedish population-based study. *Acta Paediatr Int J Paediatr* 2013;**102**(2):161–6.

5. Workinger MS, Kent RD. Perceptual analysis of the dysarthrias in children with athetoid and spastic cerebral palsy. In Moore CA, Yorkston KM, Beukelman DR, editors. *Dysarthria and Apraxia of Speech: Perspectives on Management*. Baltimore: Paul Brookes; 1991. pp. 109–26.

6. Colver A, Rapp M, Eisemann N, Ehlinger V, Thyen U, Dickinson HO, *et al.* Self-reported quality of life of adolescents with cerebral palsy: a cross-sectional and longitudinal analysis. *Lancet* 2014;**385**(9969):705–16.

7. Parkes J, White-Koning M, Dickinson HO, Thyen U, Arnaud C, Beckung E, *et al.* Psychological problems in children with cerebral palsy: a cross-sectional European study. *J Child Psychol Psychiatry* 2008;**49**(4):405–13.

8. Michelsen S, Uldall P, Kejs AMT, Madsen M. Education and employment prospects in cerebral palsy. *Devel Med Child Neurol* 2005;**47**(08):511–7.

9. Parkes J, Caravale B, Marcelli M, Franco F, Colver A. Parenting stress and children with cerebral palsy: A European cross-sectional survey. *Devel Med Child Neurol*. 2011;**53**(9):815–21.

10. Morris C, Simkiss D, Busk M, Morris M, Allard A, Denness J, *et al.* Setting research priorities to improve the health of children and young people with neurodisability: a British Academy of Childhood Disability-James Lind Alliance Research Priority Setting Partnership. *BMJ Open* 2015;**5**(1):e006233.

11. Pennington L, Parker NK, Kelly H, Miller N. Speech therapy for children with dysarthria acquired before three years of age. *Cochrane Database Syst Rev* 2016;**2016**(7):CD006937.

12. National Institute of Health and Care Excellence (NICE). *Cerebral Palsy in under 25s: Assessment and Management – NICE Guideline [NG62]*. London: NICE; 2017.

13. Watson RM, Pennington L. Assessment and management of the communication difficulties of children with cerebral palsy: a UK survey of SLT practice. *Int J Lang Commun Disord* 2015;**50**(2):241–59.

14. Children's Commissioner for England. *We Need to Talk. Access to Speech and Language Therapy*. 2019. URL: https://www.childrenscommissioner.gov.uk/report/we-need-to-talk/ (accessed 10 July 2022).

15. I CAN and RCSLT. *Bercow 10 Years On*. London: I CAN and RCSLT; 2019. URL: https://www.bercow10yearson.com/i-can-and-rcslt/ (accessed 10 July 2022).

16. Kuschmann A, Miller N, Lowit A, Pennington L. Intonation patterns in older children with cerebral palsy before and after speech intervention. *Int J Speech-Lang Pathol* 2017;**19**(4):370–80.

17. Kent RD, Read C. *The Acoustic Analysis of Speech*. San Diego, CA: Singular Publishing Group; 1992.

18. Fant G. *Acoustic Theory of Speech Production*. The Hague, Netherlands: Mouton; 1960.

19. Kent RD, Weismer G, Kent JF, Vorperian HK, Duffy JR. Acoustic studies of dysarthric speech: Methods, progress, and potential. *J Comm Disord* 1999;**32**(3):141–86.

20. Iskarous K. The nonlinear relation between articulation and acoustics: multiple constrictions. *J Acous Soc Am* 2007;**121**(3):3136.

21. Schleef E, Ramsammy M. Labiodental fronting of/θ/ in London and Edinburgh: a cross-dialectal study. *English Lang Linguistics* 2013;**17**(1):25–54.

22. Cleland J, Scobbie JM, Heyde C, Roxburgh Z, Wrench AA. Covert contrast and covert errors in persistent velar fronting. *Clin Ling Phon* 2017;**30**(3–5):249–76.

23. McAllister Byun T, Buchwald A, Mizoguchi A. Covert contrast in velar fronting: an acoustic and ultrasound study. *Clin Ling Phon* 2016;**30**(3–5):249–76.

24. Gibbon FE, Lee A. Electropalatographic (EPG) evidence of covert contrasts in disordered speech. *Clin Ling Phon*. 2017;**31**(1):4–20.

25. Surveillance of Cerebral Palsy in Europe. Surveillance of cerebral palsy in Europe: a collaboration of cerebral palsy surveys and registers. Surveillance of Cerebral Palsy in Europe (SCPE). *Devel Med Child Neurol* 2000;**42**(12):816–24.

26. Sanger TD. Pediatric movement disorders. *Curr Opin Neurol* 2003;**16**(4):529–35.

27. Darley FL, Aronson AE, Brown JR. Differential diagnostic patterns of dysarthria. *J Speech Hearing Res* 1969;**12**(2):246–69.

28. Kim Y, Kent RD, Weismer G. An acoustic study of the relationships among neurologic disease, dysarthria type, and severity of dysarthria. *J Speech Lang Hear Res* 2011;**54**(2):417–29.

29. Lansford KL, Liss JM, Norton RE. Free-classification of perceptually-similar speakers with dysarthria. *J Speech Lang Hear Res* 2014;**57**(6):2051–64.

30. Morgan AT, Liegeois F. Re-thinking diagnostic classification of the dysarthrias: a developmental perspective. *Folia Phon Logo* 2010;**62**:120–6.

31. Love RJ. *Childhood Motor Speech Disability*. 2nd ed. Boston, MA: Allyn & Bacon; 2000.

32. Solomon NP, Charron S. Speech breathing in able-bodied children and children with cerebral palsy: a review of the literature and implications for clinical intervention. *Am J Speech Lang Pathol* 1998;**7**(2):61–78.

33. Bax M, Tydeman C, Flodmark O. Clinical and MRI correlates of cerebral palsy: the European Cerebral Palsy Study. *JAMA* 2006;**296**(13):1602–8.

34. Allison KM, Hustad KC. Acoustic predictors of pediatric dysarthria in cerebral palsy. *J Speech Lang Hear Res* 2018;**61**(3):462–78.

35. Ansel BM, Kent RD. Acoustic-phonetic contrasts and intelligibility in the dysarthria associated with mixed cerebral palsy. *J Speech Lang Hear Res* 1992;**35**(2):296–308.

36. Lee J, Hustad KC, Weismer G. Predicting speech intelligibility with a multiple speech subsystems approach in children with cerebral palsy. *J Speech Lang Hear Res* 2014;**57**(5):1666–78.

37. Pennington L, Lombardo E, Steen N, Miller N. Acoustic changes in the speech of children with cerebral palsy following an intensive program of dysarthria therapy. *Int J Lang Commun Disord* 2018;**53**(1):182–95.

38. Clarke WM, Hoops HR. Predictive measures of speech proficiency in cerebral palsied speakers. *J Comm Disord* 1980;**13**(5):385–94.

39. Schölderle T, Staiger A, Lampe R, Strecker K, Ziegler W. Dysarthria in adults with cerebral palsy: clinical presentation and impacts on communication. *J Speech Lang Hear Res* 2016;**59**:216–29.

40. Kuschmann A, Neill R. *Developmental Dysarthria in a Young Adult with Cerebral Palsy: A speech Subsystems Analysis*. 18th International Congress of Phonetic Sciences, Glasgow, 10–14 August 2015.

41. Darling-White M, Sakash A, Hustad KC. Characteristics of speech rate in children with cerebral palsy: a longitudinal study. *J Speech Lang Hear Res* 2018;**61**(10):2502–15.

42. Allison KM, Hustad KC. Impact of sentence length and phonetic complexity on intelligibility of 5-year-old children with cerebral palsy. *Int J Speech Lang Pathol.* 2014;**16**(4):396–407.

43. Patel R. Acoustic characteristics of the question-statement contrast in severe dysarthria due to cerebral palsy. *J Speech Lang Hear Res* 2003;**46**(6):1401–15.

44. Kuschmann A, Lowit A. Sentence stress in children with dysarthria and cerebral palsy. *Int J Speech Lang Pathol* 2019;**21**(4):336–46.

45. Fox C, Boliek C. Intensive voice treatment (LSVT LOUD) for children with spastic cerebral palsy and dysarthria. *J Speech Lang Hear Res* 2012;**55**(3):930–45.

46. Jeng JY, Weismer G, Kent RD. Production and perception of mandarin tone in adults with cerebral palsy. *Clin Ling Phon* 2006;**20**(1):67–87.

47. Hustad KC, Beukelman DR. Listener comprehension of severely dysarthric speech: effects of linguistic cues and stimulus cohesion. *J Speech Lang Hear Res* 2002;**45**(3):545–58.

48. Boliek C, Fox C. Therapeutic effects of intensive voice treatment (LSVT LOUD®) for children with spastic cerebral palsy and dysarthria: a phase I treatment validation study. *Int J Speech Lang Pathol* 2017;**19**(6):601–15.

49. Levy ES, Chang YM, Ancelle JA, McAuliffe MJ. Acoustic and perceptual consequences of speech cues for children with dysarthria. *J Speech Lang Hear Res* 2017;**60**(6S):1766–79.

50. Pennington L, Stamp E, Smith J, Kelly H, Parker N, Stockwell K, *et al.* Internet delivery of intensive speech and language therapy for children with cerebral palsy: a pilot randomised controlled trial. *BMJ Open* 2019;**9**:e024233.

51. Pennington L, Rauch R, Smith J, Brittain K. Intensive speech therapy for children with cerebral palsy: the views of children and parents on effectiveness and acceptability. *Disab Rehabil* 2019;**42**(20):2935–43.

52. Pennington L, Miller N, Robson S, Steen N. Intensive speech and language therapy for older children with cerebral palsy: a systems approach. *Devel Med Child Neurol* 2010;**52**(4):337–44.

53. Pennington L, Roelant E, Thompson V, Robson S, Steen N, Miller N. Intensive dysarthria therapy for younger children with cerebral palsy. *Devel Med Child Neurol* 2013;**55**(5):464–71.

54. Hirano M. *Clinical Examination of Voice*. Vienna: Springer; 1981.

55. Miller N, Pennington L, Robson S, Roelant E, Steen N, Lombardo E. Changes in voice quality after speech-language therapy intervention in older children with cerebral palsy. *Fol Phon Logop* 2013;**65**(4):200–7.

56. Palisano RJ, Rosenbaum P, Walter S, Russell D, Wood E, Galuppi B. Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Devel Med Child Neurol* 1997;**39**:214–23.

57. Wilcox K, Morris S. *Children's Speech Intelligibility Measure*. San Antonio, TX: Harcourt Assessment; 1999.

58. Sehgal S. *Dysarthric Speech Analysis and Automatic Recognition Using Phase Based Representations*. Sheffield: University of Sheffield; 2018.

59. Morton J. Interaction of information in word recognition. *Psychol Rev* 1969;**76**(2):165–78.

60. BNC Consortium. *The British National Corpus*. Oxford: Oxford Text Archive; 2007.

61. van Bergem DR. Acoustic vowel reduction as a function of sentence accent, word stress, and word class. *Speech Comm* 1993;**12**(1):1–23.

62. Hustad KC. A closer look at transcription intelligibility for speakers with dysarthria: evaluation of scoring paradigms and linguistic errors made by listeners. *Am J Speech Lang Pathol* 2006;**15**(3):268–77.

63. Baayen RH, Piepenbrock R, Gulikers L. *CELEX2 LDC96L14. Web Download*. Philadelphia, PA: Linguistic Data Consortium; 1995. URL: https://catalog.ldc.upenn.edu/LDC96L14 (accessed 10 July 2022).

64. Leyland AH, Groenewegen PP. *Multilevel Modelling for Public Health and Health Services Research*. Chamonix, Switzerland: Springer; 2020.

65. Snijders TAB, Bosker RJ. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Los Angeles, CA: Sage; 2021.

66. Cappelleri JC, Lundy JJ, Hays RD. Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clin Therap* 2014;**1**(5):648–62.

67. Makles A. How to get the optimal k-means cluster solution. *STATA J* 2012;**12**(2):347–51.

68. Johnson K. *Acoustic and Auditory Phonetics*. Oxford: Blackwell; 1997.

69. Stevens KN. *Acoustic Phonetics*. Cambridge, CT: MIT Press; 1998.

70. Khattab G, Al-Tamimi J, Alsiraih W. Nasalisation in the production of Iraqi Arabic pharyngeals. *Phonetica* 2018;**75**(4):310–48.

71. Sussman HM, Hoemeke KA, McCaffrey HA. Locus equations as an index of coarticulation for place of articulation distinctions in children. *J Speech Lang Hear Res* 1992;**35**(4):769–81.

72. Boersma P, Weenink D. *Praat: Doing Phonetics by Computer*. 2010. Published 25 February 2010. 5.1.26. URL: http://www.praat.org (accessed 10 July 2022).

73. Pennington L, Rauch R, Smith J, Brittain K. Intensive speech therapy for children with cerebral palsy: the views of children and parents on effectiveness and acceptability. *Disab Rehab* 2020;**42**:2935–43.

74. Platt LJ, Andrews G, Young M, Quinn PT. Dysarthria of adult cerebral palsy: I – intelligibility and articulatory impairment. *J Speech Hear Res* 1980;**23**(1):28–40.

75. Platt LJ, Andrews G, Howie PM. Dysarthria of adult cerebral palsy: II – phonemic analysis of articulation errors. *J Speech Hear Res* 1980;**23**(1):41–55.

76. Kim H, Martin K, Hasegawa-Johnson M, Perlman A. Frequency of consonant articulation errors in dysarthric speech. *Clin Ling Phon* 2010;**24**(10):759–70.

77. Docherty G. *The Timing of Voicing in British English Obstruents*. Berlin; New York: De Gruyter Mouton; 2011.

78. Andrews G, Platt LJ, Young M. Factors affecting the intelligibility of cerebral palsied speech to the average listener. *Fol Phoniat* 1977;**29**(4):292–301.

79. Byrne M. Speech and language development in athetoid and spastic children. *J Speech Hear Dis* 1959;**24**:231–40.

80. Dodd B, Holm A, Hua Z, Crosbie S. Phonological development: a normative study of British English-speaking children. *Clin Ling Phon* 2003;**17**(8):617–43.

81. Hustad KC. Effects of speech stimuli and dysarthria severity on intelligibility scores and listener confidence ratings for speakers with cerebral palsy. *Folia Phon Logop* 2007;**59**(6):306–17.

82. Braza MD, Sakash A, Natzke P, Hustad KC. Longitudinal change in speech rate and intelligibility between 5 and 7 years in children with cerebral palsy. *Am J Speech Lang Pathol* 2019;**28**(3):1139–51.

83. Hustad KC, Kent RD, Beukelman DR. DECTalk and MacinTalk speech synthesizers: intelligibility differences for three listener groups. *J Speech Lang Hear Res* 1998;**41**(4):744–52.

84. Borrie SA, McAuliffe MJ, Liss JM. Perceptual learning of dysarthric speech: a review of experimental studies. *J Speech Lang Hear Res* 2012;**55**(1):290–305.

85. Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. *J Mem Lang* 2008;**59**(4):390–412.

86. Miller N. Measuring up to speech intelligibility. *Int J Lang Commun Disord* 2013;**48**(6):601–12.

87. Kent RD, Miolo G, Bloedel ST. The intelligibility of children's speech: a review of evaluation procedures. *Am J Speech Lang Pathol* 1994;**2**:81–95.

88. Whitehill TL, Ciocca V. Perceptual-phonetic predictors of single-word intelligibility: a study of Cantonese dysarthria. *J Speech Lang Hear Res* 2000;**43**(6):1451–65.

89. Schiavetti N. Scaling procedures for the measurement of speech intelligibility. In Kent R, editor. *Intelligibility in Speech Disorders: Theory, Measurement and Management*. Amsterdam: Benjamins. 1992. pp. 11–34.

90. Patel R, Connaghan K. Park play: a picture description task for assessing childhood motor speech disorders. *Int J Speech Lang Pathol* 2014;**16**(4):337–43.

91. Children's Commissioner for England. *Child Poverty: The Crisis We Can't Keep Ignoring*. London: Children's Commissioner for England; 2021. URL: https://www.childrenscommissioner.gov.uk/report/child-poverty/ (accessed 10 July 2022).

92. Witt SM, Young S. Phone-level pronunciation scoring and assessment for interactive language learning. *Speech Commun* 2000;**30**:95–108.

Appendix 1 Full GLM models

TABLE 18 Effect of therapy on identification of initial singleton consonants in SWs (n = 17,323 observations over N = 42 children)

	Model 1			Model 2			Model 3			Model 4		
	OR (95% CI)	Wald z	p	OR (95% CI)	Wald z	p	OR (95% CI)	Wald z	p	OR (95% CI)	Wald z	p
Treatment (ref: pre-therapy)												
Post-therapy	1.51 (1.41 to 1.61)	11.85	< 0.01	1.51 (1.41 to 1.61)	11.85	< 0.01	1.51 (1.41 to 1.62)	11.94	< 0.01	1.54 (1.44 to 1.65)	12.15	< 0.01
Age, years				1.07 (0.95 to 1.21)	1.05	0.29	1.07 (0.95 to 1.21)	1.06	0.29	1.07 (0.94 to 1.22)	1.07	0.29
Sex (ref: Male)												
Female				1.07 (0.63 to 1.82)	0.24	0.81	1.06 (0.62 to 1.81)	0.22	0.83	1.07 (0.61 to 1.88)	0.25	0.80
Study source (ref: Pennington 2010)[52]												
Pennington 2013[53]				2.81 (1.13 to 7.02)	2.22	0.03	2.83 (1.13 to 7.09)	2.23	0.03	2.98 (1.14 to 7.79)	2.23	0.03
Pennington 2019[50]				2.04 (0.82 to 5.06)	1.54	0.12	2.05 (0.82 to 5.09)	1.54	0.12	2.11 (0.81 to 5.48)	1.54	0.13
CP type (ref: Spastic)												
Dyskinetic				1.07 (0.61 to 1.86)	0.22	0.83	1.07 (0.61 to 1.87)	0.24	0.81	1.09 (0.60 to 1.95)	0.27	0.78
Other				0.51 (0.21 to 1.27)	−1.45	0.15	0.51 (0.21 to 1.26)	−1.45	0.15	0.49 (0.19 to 1.26)	−1.49	0.14
Target word complexity							0.96 (0.95 to 0.98)	−4.86	< 0.01	1.00 (0.98 to 1.02)	0.14	0.89
Target word syllables (ref: Monosyllabic)												
Polysyllabic							0.80 (0.74 to 0.88)	−5.03	< 0.01	0.75 (0.68 to 0.81)	−6.55	< 0.01
Consonant VPM (ref: Voiced coronal obstruent)												
Voiced coronal sonorant										0.62 (0.54 to 0.72)	−6.52	< 0.01
Voiced dorsal obstruent										0.83 (0.65 to 1.05)	−1.57	0.12
Voiced labial obstruent										1.52 (1.29 to 1.79)	5.04	< 0.01

TABLE 18 Effect of therapy on identification of initial singleton consonants in SWs (n = 17,323 observations over N = 42 children) (*continued*)

	Model 1			Model 2			Model 3			Model 4		
	OR (95% CI)	**Wald z**	**p**	**OR (95% CI)**	**Wald z**	**p**	**OR (95% CI)**	**Wald z**	**p**	**OR (95% CI)**	**Wald z**	**p**
Voiced labial sonorant										0.87 (0.74 to 1.02)	−1.77	0.08
Voiceless coronal obstruent										0.33 (0.28 to 0.37)	−16.31	<0.01
Voiceless dorsal obstruent										0.82 (0.69 to 0.97)	−2.35	0.02
Voiceless labial obstruent										0.45 (0.39 to 0.52)	−10.81	<0.01
LR test for therapy effect	X^2	degrees of freedom (df)	p	X^2	df	p	X^2	df	p	X^2	df	p
Therapy	141.47	1	<0.01	141.44	1	<0.01	143.70	1	<0.01	148.94	1	<0.01
	Coefficient	SE		Coefficient	SE		Coefficient	SE		Coefficient	SE	
Between-child variance estimate	0.69	0.15		0.58	0.13		0.59	0.13		0.65	0.14	

TABLE 19 Effect of therapy on identification of final singleton consonants in SWs (n = 16,623 observations over N = 42 children)

	Model 1			Model 2			Model 3			Model 4		
	OR (95% CI)	Wald z	p	OR (95% CI)	Wald z	p	OR (95% CI)	Wald z	p	OR (95% CI)	Wald z	p
Therapy (ref: pre-therapy)												
Post-therapy	1.59 (1.49, 1.71)	13.46	< 0.01	1.59 (1.49 to 1.71)	13.45	< 0.01	1.60 (1.49 to 1.71)	13.45	< 0.01	1.61 (1.51 to 1.73)	13.63	< 0.01
Age, years				1.05 (0.94 to 1.17)	0.81	0.42	1.05 (0.94 to 1.17)	0.81	0.42	1.05 (0.94 to 1.18)	0.9	0.37
Sex (ref: Male)												
Female				1.00 (0.62 to 1.62)	< 0.01	0.99	0.99 (0.61 to 1.62)	−0.03	0.98	1.02 (0.62 to 1.66)	0.07	0.95
Study source (ref: Pennington 2010)[52]												
Pennington 2013[53]				2.63 (1.15 to 6.04)	2.28	0.02	2.62 (1.14 to 6.06)	2.26	0.02	2.80 (1.20 to 6.51)	2.38	0.02
Pennington 2019[50]				1.70 (0.75 to 3.89)	1.26	0.21	1.70 (0.74 to 3.90)	1.25	0.21	1.78 (0.77 to 4.12)	1.35	0.18
CP type (ref: Spastic)												
Dyskinetic				1.15 (0.69 to 1.92)	0.55	0.58	1.15 (0.69 to 1.92)	0.55	0.58	1.14 (0.68 to 1.91)	0.5	0.62
Other				0.72 (0.32 to 1.65)	−0.77	0.44	0.72 (0.31 to 1.65)	−0.77	0.44	0.70 (0.30 to 1.61)	−0.85	0.40
Target word complexity							0.95 (0.94 to 0.97)	−5.9	< 0.01	0.97 (0.96 to 0.99)	−2.91	< 0.01
Target word syllables (ref: Monosyllabic)												
Polysyllabic							1.58 (1.42 to 1.76)	8.42	< 0.01	1.46 (1.30 to 1.64)	6.28	< 0.001
Consonant VPM (ref: Voiced coronal obstruent)												
Voiced coronal sonorant										2.40 (2.14 to 2.69)	14.84	< 0.01
Voiced dorsal obstruent										0.91 (0.51 to 1.63)	−0.31	0.76
Voiced dorsal sonorant										0.61 (0.47 to 0.80)	−3.56	< 0.01

TABLE 19 Effect of therapy on identification of final singleton consonants in SWs (n = 16,623 observations over N = 42 children) *(continued)*

	Model 1			Model 2			Model 3			Model 4		
	OR (95% CI)	Wald z	p	OR (95% CI)	Wald z	p	OR (95% CI)	Wald z	p	OR (95% CI)	Wald z	p
Voiced labial obstruent										1.20 (1.00 to 1.44)	1.97	0.05
Voiced labial sonorant										1.40 (1.19 to 1.64)	4.05	<0.01
Voiceless coronal obstruent										1.45 (1.31 to 1.61)	7.1	<0.01
Voiceless dorsal obstruent										2.21 (1.87 to 2.61)	9.28	<0.01
Voiceless labial obstruent										1.31 (1.15 to 1.50)	3.91	<0.01
LR test for therapy effect	X^2	df	p	X^2	df	p	X^2	df	p	X^2	df	p
Therapy	182.74	1	<0.01	182.46	1	<0.01	182.55	1	<0.01	187.59	1	<0.01
	Coefficient	SE		Coefficient	SE		Coefficient	SE		Coefficient	SE	
Between-child variance estimate	0.56	0.13		0.48	0.11		0.49	0.11		0.5	0.11	

TABLE 20 Effect of therapy on identification of initial consonant clusters in SWs (n = 4261 observations over N = 42 children)

	Model 1			Model 2			Model 3			Model 4		
	OR (95% CI)	Wald z	p	OR (95% CI)	Wald z	p	OR (95% CI)	Wald z	p	OR (95% CI)	Wald z	p
Therapy (ref: pre-therapy)												
Post-therapy	1.77 (1.55, 2.03)	8.25	< 0.01	1.77 (1.55 to 2.03)	8.26	< 0.01	1.79 (1.56 to 2.05)	8.38	< 0.01	1.84 (1.60 to 2.12)	8.59	< 0.01
Age, years				1.08 (0.94 to 1.23)	1.06	0.29	1.08 (0.94 to 1.23)	1.05	0.30	1.08 (0.94 to 1.25)	1.08	0.28
Sex (ref: Male)												
Female				1.41 (0.78 to 2.56)	1.13	0.26	1.43 (0.79 to 2.60)	1.17	0.24	1.40 (0.74 to 2.62)	1.04	0.30
Study source (ref: Pennington 2010)[52]												
Pennington 2013[53]				3.67 (1.33 to 10.18)	2.50	0.01	3.67 (1.32 to 10.24)	2.49	0.01	4.23 (1.44 to 12.48)	2.62	0.01
Pennington 2019[50]				2.68 (0.97 to 7.39)	1.90	0.06	2.67 (0.96 to 7.40)	1.88	0.06	3.05 (1.04 to 8.93)	2.03	0.04
CP type (ref: Spastic)												
Dyskinetic				1.06 (0.57 to 1.98)	0.20	0.85	1.06 (0.57 to 1.99)	0.19	0.85	1.12 (0.58 to 2.16)	0.33	0.74
Other				0.44 (0.16 to 1.22)	−1.58	0.12	0.45 (0.16 to 1.24)	−1.54	0.12	0.42 (0.15 to 1.24)	−1.57	0.12
Target word complexity							1.01 (0.97 to 1.05)	0.43	0.67	1.07 (1.02 to 1.12)	2.93	< 0.01
Target word syllables (ref: Monosyllabic)												
Polysyllabic							0.67 (0.54 to 0.84)	−3.50	< 0.01	0.77 (0.60 to 0.99)	−2.06	0.04
Cluster VPM (ref: VVCCOS)												
VVDCOS										2.17 (1.54 to 3.07)	4.40	< 0.01
VVLCOS										2.65 (1.72 to 4.08)	4.40	< 0.01
VIVICCOO										0.73 (0.50 to 1.08)	−1.59	0.11
VIVICLOO										0.42 (0.27 to 0.66)	−3.86	< 0.01
VIVICDOO										0.60 (0.38 to 0.94)	−2.23	0.03
VIVCCOS										0.85 (0.60 to 1.21)	−0.91	0.37

TABLE 20 Effect of therapy on identification of initial consonant clusters in SWs ($n = 4261$ observations over $N = 42$ children) (continued)

	Model 1			Model 2			Model 3			Model 4		
	OR (95% CI)	Wald z	p	OR (95% CI)	Wald z	p	OR (95% CI)	Wald z	p	OR (95% CI)	Wald z	p
VIVCLOS										1.43 (0.99 to 2.06)	1.91	0.06
VIVDCOS										1.23 (0.84 to 1.81)	1.07	0.29
VIVDLOS										0.91 (0.54 to 1.54)	−0.35	0.73
VIVLCOS										1.09 (0.73 to 1.62)	0.43	0.67
LR test for therapy effect	X^2	df	p	X^2	df	p	X^2	df	p	X^2	df	p
Therapy	68.98	1	<0.01	69.21	1	<0.01	71.23	1	<0.01	75.02	1	<0.01
	Coefficient	SE		Coefficient	SE		Coefficient	SE		Coefficient	SE	
Between-child variance estimate	0.91	0.21		0.69	0.17		0.70	0.17		0.78	0.19	

Note
V, voiced; Vl, voiceless; C, coronal; L, labial; D, dorsal; O, obstruent; S, sonorant.

TABLE 21 Effect of therapy on identification of final consonant clusters in SWs (n = 1949 observations over N = 42 children)

	Model 1			Model 2			Model 3			Model 4		
	OR (95% CI)	Wald z	p	OR (95% CI)	Wald z	p	OR (95% CI)	Wald z	p	OR (95% CI)	Wald z	p
Therapy (ref: pre-therapy)												
Post-therapy	1.38 (1.13 to 1.70)	3.10	< 0.01	1.38 (1.13 to 1.70)	3.11	< 0.01	1.37 (1.11 to 1.68)	2.96	< 0.01	1.42 (1.15 to 1.75)	3.26	< 0.01
Age, years				1.12 (0.96 to 1.30)	1.42	0.15	1.11 (0.95 to 1.30)	1.30	0.19	1.10 (0.94 to 1.29)	1.24	0.22
Sex (ref: Male)												
Female				0.85 (0.44 to 1.63)	−0.48	0.63	0.85 (0.43 to 1.67)	−0.48	0.63	0.87 (0.44 to 1.72)	−0.40	0.69
Study source (ref: Pennington 2010)[52]												
Pennington 2013[53]				5.14 (1.67 to 15.80)	2.86	< 0.01	5.27 (1.64 to 16.92)	2.80	< 0.01	5.16 (1.60 to 16.68)	2.74	0.01
Pennington 2019[50]				2.05 (0.95 to 9.13)	1.88	0.06	2.98 (0.92 to 9.63)	1.82	0.07	2.93 (0.90 to 9.54)	1.79	0.07
CP type (ref: Spastic)												
Dyskinetic				1.01 (0.51 to 1.99)	0.03	0.98	1.04 (0.51 to 2.12)	0.12	0.91	1.02 (0.50 to 2.08)	0.06	0.96
Other				0.44 (0.15 to 1.34)	−1.44	0.15	0.47 (0.15 to 1.48)	−1.29	0.20	0.51 (0.16 to 1.61)	−1.15	0.25
Target word complexity							0.97 (0.93 to 1.01)	−1.32	0.19	0.94 (0.89 to 0.99)	−2.28	0.02
Target word syllables (ref: Monosyllabic)												
Polysyllabic							2.92 (2.07 to 4.11)	6.14	< 0.01	2.09 (1.29 to 3.38)	3.01	< 0.01
Cluster VPM (ref: VVCCSO)												
VIVICCOO										0.84 (0.62 to 1.13)	−1.17	0.24
VIVICLOO										0.39 (0.18 to 0.85)	−2.36	0.02
VIVIDCOO										1.80 (0.92 to 3.50)	1.72	0.09
VIVILCOO										0.44 (0.28 to 0.69)	−3.60	< 0.01
VIVICDOO										0.36 (0.21 to 0.64)	−3.50	< 0.01
VVICCSO										0.42 (0.26 to 0.68)	−3.56	< 0.01
VVILLSO										1.32 (0.77 to 2.28)	1.01	0.31

TABLE 21 Effect of therapy on identification of final consonant clusters in SWs (*n* = 1949 observations over *N* = 42 children) (*continued*)

	Model 1			Model 2			Model 3			Model 4		
	OR (95% CI)	Wald z	*p*	OR (95% CI)	Wald z	*p*	OR (95% CI)	Wald z	*p*	OR (95% CI)	Wald z	*p*
LR test for therapy effect	X^2	*df*	*p*	X^2	*df*	*p*	X^2	*df*	*p*	X^2	*df*	*p*
Therapy	9.69	1	<0.01	9.73	1	<0.01	8.81	1	<0.01	10.71	1	<0.01
	Coefficient	SE		Coefficient	SE		Coefficient	SE		Coefficient	SE	
Between-child variance estimate	1.01	0.26		0.76	0.20		0.84	0.22		0.85	0.22	

Note
V, voiced; VI, voiceless; C, coronal; L, labial; D, dorsal; O, obstruent; S, sonorant.

TABLE 22 Effect of therapy on identification of initial singleton consonants in CS data (n = 8457 observations over N = 42 children)

	Model 1			Model 2			Model 3			Model 4		
	OR (95% CI)	Wald z	*p*	OR (95% CI)	Wald z	*p*	OR (95% CI)	Wald z	*p*	OR (95% CI)	Wald z	*p*
Therapy (ref: pre-therapy)												
Post-therapy	1.26 (1.15 to 1.39)	4.84	< 0.01	1.26 (1.15 to 1.39)	4.84	< 0.01	1.27 (1.15 to 1.39)	4.87	< 0.01	1.26 (1.15 to 1.39)	4.80	< 0.01
Age, years				1.07 (0.89 to 1.28)	0.75	0.45	1.07 (0.89 to 1.29)	0.74	0.46	1.07 (0.89 to 1.28)	0.71	0.48
Sex (ref: Male)												
Female				1.30 (0.59 to 2.88)	0.65	0.52	1.30 (0.58 to 2.9)	0.64	0.52	1.29 (0.58 to 2.89)	0.62	0.53
Study source (ref: Pennington 2010)[52]												
Pennington 2013[53]				2.56 (0.66 to 9.96)	1.35	0.18	2.55 (0.65 to 10.09)	1.34	0.18	2.47 (0.62 to 9.79)	1.29	0.20
Pennington 2019[50]				1.62 (0.42 to 6.25)	0.70	0.49	1.58 (0.40 to 6.17)	0.65	0.51	1.54 (0.39 to 6.07)	0.62	0.54
CP type (ref: Spastic)												
Dyskinetic				1.21 (0.53 to 2.78)	0.45	0.65	1.22 (0.53 to 2.81)	0.46	0.65	1.20 (0.52 to 2.78)	0.42	0.67
Other				0.33 (0.08 to 1.28)	−1.60	0.11	0.32 (0.08 to 1.27)	−1.62	0.11	0.33 (0.08 to 1.3)	−1.59	0.11
Word position in phrase (ref: Phrase initial)												
Phrase final							0.87 (0.78 to 0.97)	−2.45	0.01	0.88 (0.78 to 0.98)	−2.28	0.02
Log of Target word density							0.97 (0.92 to 1.01)	−1.51	0.13	0.98 (0.94 to 1.03)	−0.72	0.47
Target word syllables (ref: Monosyllabic)												
Polysyllabic							0.96 (0.86 to 1.08)	−0.67	0.50	0.99 (0.88 to 1.11)	−0.24	0.81
Consonant VPM (ref: Voiced coronal obstruent)												
Voiced coronal sonorant										0.94 (0.80 to 1.11)	−0.71	0.48
Voiced dorsal obstruent										1.32 (1.03 to 1.7)	2.16	0.03
Voiced labial obstruent										1.02 (0.85 to 1.23)	0.24	0.81

TABLE 22 Effect of therapy on identification of initial singleton consonants in CS data (n = 8457 observations over N = 42 children) *(continued)*

	Model 1			Model 2			Model 3			Model 4		
	OR (95% CI)	Wald z	p	OR (95% CI)	Wald z	p	OR (95% CI)	Wald z	p	OR (95% CI)	Wald z	p
Voiced labial sonorant										1.4 (1.20 to 1.62)	4.34	<0.01
Voiceless coronal obstruent										0.93 (0.78 to 1.09)	−0.90	0.37
Voiceless dorsal obstruent										1.39 (1.10 to 1.76)	2.79	<0.01
Voiceless labial obstruent										1.19 (0.97 to 1.47)	1.65	0.10
LR test for therapy effect	X^2	df	p	X^2	df	p	X^2	df	p	X^2	df	p
Therapy	23.51	1	<0.01	23.44	1	<0.01	23.80	1	<0.01	23.10	1	<0.01
Between-child variance estimate	Coefficient	SE		Coefficient	SE		Coefficient	SE		Coefficient	SE	
	1.45	0.34		1.28	0.30		1.30	0.31		1.31	0.31	

TABLE 23 Effect of therapy on identification of final singleton consonants in CS data (n = 7205 observations over N = 42 children)

	Model 1			Model 2			Model 3			Model 4		
	OR (95% CI)	Wald z	p	OR (95% CI)	Wald z	p	OR (95% CI)	Wald z	p	OR (95% CI)	Wald z	p
Therapy (ref: pre-therapy)												
Post-therapy	1.29 (1.17 to 1.43)	4.86	< 0.01	1.29 (1.16 to 1.43)	4.83	< 0.01	1.29 (1.16 to 1.43)	4.79	< 0.01	1.27 (1.15 to 1.41)	4.54	< 0.01
Age, years				1.03 (0.87 to 1.22)	0.33	0.74	1.03 (0.87 to 1.23)	0.37	0.71	1.04 (0.87 to 1.23)	0.39	0.69
Sex (ref: Male)												
Female				1.05 (0.49 to 2.24)	0.12	0.90	1.04 (0.49 to 2.22)	0.10	0.92	1.03 (0.48 to 2.23)	0.08	0.93
Study source (ref: Pennington 2010)												
Pennington 2013				2.21 (0.60 to 8.07)	1.20	0.23	2.30 (0.63 to 8.41)	1.26	0.21	2.33 (0.63 to 8.64)	1.26	0.21
Pennington 2019				1.08 (0.30 to 3.92)	0.12	0.91	1.12 (0.31 to 4.07)	0.18	0.86	1.15 (0.31 to 4.24)	0.21	0.84
CP type (ref: Spastic)												
Dyskinetic				1.35 (0.61 to 2.99)	0.74	0.46	1.36 (0.61 to 3.01)	0.76	0.45	1.38 (0.62 to 3.07)	0.78	0.44
Other				0.45 (0.12 to 1.66)	−1.19	0.23	0.45 (0.12 to 1.65)	−1.20	0.23	0.43 (0.12 to 1.61)	−1.24	0.21
Word position in phrase (ref: Phrase initial)												
Phrase final							1.04 (0.91 to 1.18)	0.59	0.55	1.04 (0.92 to 1.19)	0.64	0.52
Log of Target word density							1.07 (1.03 to 1.11)	3.56	< 0.01	1.07 (1.02 to 1.11)	2.94	< 0.01
Target word syllables (ref: Monosyllabic)												
Polysyllabic							0.92 (0.81 to 1.05)	−1.19	0.23	0.99 (0.85 to 1.15)	−0.14	0.87
Consonant VPM (ref: Voiced coronal obstruent)												
Voiced coronal sonorant										1.59 (1.34 to 1.87)	5.45	< 0.01
Voiced dorsal obstruent										1.96 (1.36 to 2.81)	3.64	< 0.01
Voiced dorsal sonorant										1.05 (0.82 to 1.35)	0.39	0.70

TABLE 23 Effect of therapy on identification of final singleton consonants in CS data (n = 7205 observations over N = 42 children) (continued)

	Model 1			Model 2			Model 3			Model 4		
	OR (95% CI)	Wald z	p	OR (95% CI)	Wald z	p	OR (95% CI)	Wald z	p	OR (95% CI)	Wald z	p
Voiced labial obstruent										1.26 (0.94 to 1.7)	1.55	0.12
Voiced labial sonorant										0.79 (0.60 to 1.05)	−1.65	0.10
Voiceless coronal obstruent										1.27 (1.07 to 1.51)	2.76	0.01
Voiceless dorsal obstruent										1.48 (1.16 to 1.89)	3.12	<0.01
Voiceless labial obstruent										0.97 (0.72 to 1.3)	−0.21	0.83
LR test for therapy effect	X^2	df	p	X^2	df	p	X^2	df	p	X^2	df	p
Therapy	23.65	1	<0.01	23.35	1	<0.01	23.02	1	<0.01	20.64	1	<0.01
	Coefficient	SE		Coefficient	SE		Coefficient	SE		Coefficient	SE	
Between-child variance estimate	1.30	0.31		1.16	0.28		1.15	0.28		1.19	0.29	

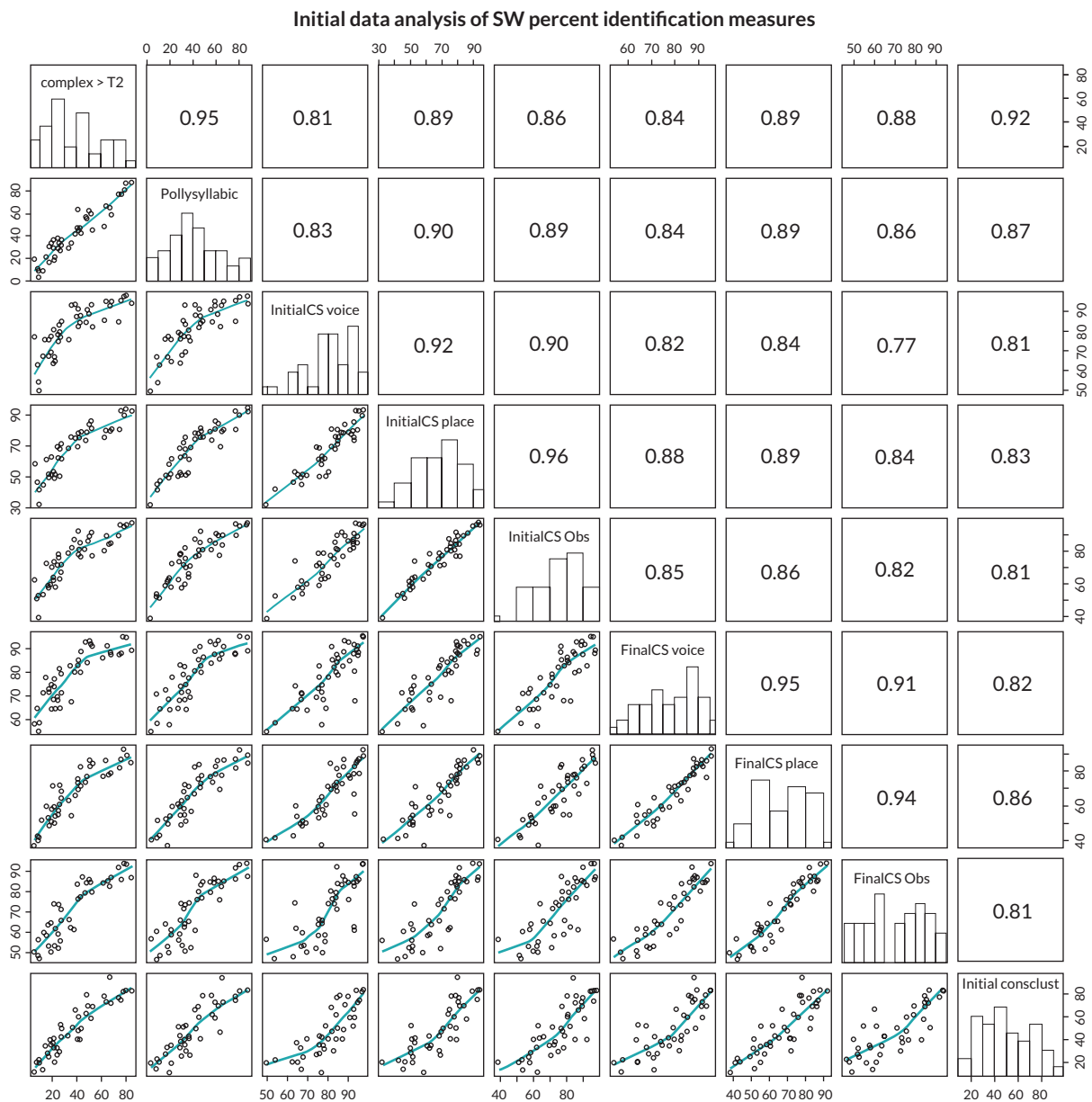# Appendix 2 Supporting information for principal components analysis in SW data

FIGURE 18 Bivariate correlation between percent identification measures in SW data.

TABLE 24 Descriptive statistics of PI measures at pre-therapy and their loadings on component 1

PI measure at pre-therapy	Mean	SD	Loading
SWs of complexity > tertile 2 (upper third)	37.95	22.34	0.34
Polysyllabic words	42.01	22.05	0.34
Initial consonant voicing in words with initial singleton consonants	80.85	11.90	0.32
Initial consonant place in words with initial singleton consonants	68.09	15.54	0.34
Initial consonant manner in words with initial singleton consonants that have obstruent manner	75.18	14.28	0.33
Final consonant voicing in words with final singleton consonants	78.46	11.44	0.33
Final consonant place in words with final singleton consonants	67.03	14.79	0.34
Final consonant manner in words with final singleton consonants that have obstruent manner	70.55	14.42	0.33
Initial consonant cluster (two consonant length) in words with initial clusters	49.68	22.21	0.33

Appendix 3   Frequency of combined voice, place, and manner categories for consonant clusters

TABLE 25   Initial consonant clusters (two consonants)

| Acronym | VPM combination | | | Frequency |
	Voicing	Place	Manner	n = 1072 observations (%)
VVCCOS	Voiced-Voiced	Coronal-Coronal	Obstruent-Sonorant	47 (4.38)
VVCCSS	Voiced-Voiced	Coronal-Coronal	Sonorant-Sonorant	15 (1.40)
VVDCOS	Voiced-Voiced	Dorsal-Coronal	Obstruent-Sonorant	95 (8.86)
VVLCOS	Voiced-Voiced	Labial-Coronal	Obstruent-Sonorant	118 (11.01)
VVLCSS	Voiced-Voiced	Labial-Coronal	Sonorant-Sonorant	6 (0.56)
VlVlCCOO	Voiceless-Voiceless	Coronal-Coronal	Obstruent-Obstruent	110 (10.26)
VlVlCLOO	Voiceless-Voiceless	Coronal-Labial	Obstruent-Obstruent	65 (6.06)
VlVlCDOO	Voiceless-Voiceless	Coronal-Dorsal	Obstruent-Obstruent	66 (6.16)
VlVCCOS	Voiceless-Voiced	Coronal-Coronal	Obstruent-Sonorant	142 (13.25)
VlVCLOS	Voiceless-Voiced	Coronal-Labial	Obstruent-Sonorant	87 (8.12)
VlVDCOS	Voiceless-Voiced	Dorsal-Coronal	Obstruent-Sonorant	110 (10.26)
VlVDLOS	Voiceless-Voiced	Dorsal-Labial	Obstruent-Sonorant	9 (0.84)
VlVLCOS	Voiceless-Voiced	Labial-Coronal	Obstruent-Sonorant	202 (18.84)

Note
V, voiced; Vl, voiceless; C, coronal; L, labial; D, dorsal; O, obstruent; S, sonorant. For consonant clusters, the VPM combinations are reported as acronyms for ease of reporting whose structure goes as follows: voice of first consonant, voice of second consonant, place of first consonant, place of second consonant, manner of first consonant, and manner of second consonant in the cluster. For example, the initial consonant cluster/st/ has/s/ as the first consonant and/t/ as the second consonant.

TABLE 26 Final consonant clusters (two consonants)

| Acronym | VPM combination | | | Frequency |
	Voicing	Place	Manner	*n* = 1900 observations (%)
VVCCOO	Voiced-Voiced	Coronal-Coronal	Obstruent-Obstruent	3 (0.16)
VVCCSO	Voiced-Voiced	Coronal-Coronal	Sonorant-Obstruent	1119 (58.89)
VVCLSS	Voiced-Voiced	Coronal-Labial	Sonorant-Sonorant	3 (0.16)
VVDCOO	Voiced-Voiced	Dorsal-Coronal	Obstruent-Obstruent	5 (0.26)
VVLCSO	Voiced-Voiced	Labial-Coronal	Sonorant-Obstruent	30 (1.58)
VIVICCOO	Voiceless-Voiceless	Coronal-Coronal	Obstruent-Obstruent	373 (19.63)
VIVIDCOO	Voiceless-Voiceless	Dorsal-Coronal	Obstruent-Obstruent	69 (3.63)
VIVILCOO	Voiceless-Voiceless	Labial-Coronal	Obstruent-Obstruent	5 (0.26)
VVICCSO	Voiced-Voiceless	Coronal-Coronal	Sonorant-Obstruent	259 (13.63)
VVICLSO	Voiced-Voiceless	Coronal-Labial	Sonorant-Obstruent	10 (0.53)
VVICDSO	Voiced-Voiceless	Coronal-Dorsal	Sonorant-Obstruent	6 (0.32)
VVILLSO	Voiced-Voiceless	Labial-Labial	Sonorant-Obstruent	18 (0.95)

Note
V, voiced; VI, voiceless; C, coronal; L, labial; D, dorsal; O, obstruent; S, sonorant. For consonant clusters, the VPM combinations are reported as acronyms for ease of reporting whose structure goes as follows: voice of first consonant, voice of second consonant, place of first consonant, place of second consonant, manner of first consonant, and manner of second consonant in the cluster. For example, the final consonant cluster/st/ has/s/ as the first consonant and/t/ as the second consonant.

Appendix 4 Supporting information for principal components analysis in CS data

FIGURE 19 Bivariate correlation between percent identification measures in CS data.

TABLE 27 Descriptive statistics of percentage identification measures at pre-therapy and their loadings on component 1

Percent identification measure at pre-therapy	Mean	SD	Component 1 loading
Phrase initial words	32.07	22.53	0.35
Phrase final words	31.93	24.08	0.36
Initial singleton consonant voicing	43.43	22.66	0.36
Initial singleton consonant place	39.99	23.45	0.36
Obstruent initial singleton consonants	41.90	23.64	0.35
Final singleton consonant voicing	39.79	22.81	0.35
Final singleton consonant place	38.74	22.44	0.35
Obstruent final singleton consonants	36.46	23.23	0.35

EME
HSDR
HTA
PGfAR
PHR