Investigating Variability in Child Speech (VariCS) – the VariCS corpus

Anja Kuschmann, Mridhula Murali, Lauren Taylor, Amy Smith, Joanne Cleland, Jane Stuart-Smith & David Young

Version 1.0, December 2024

Abstract

The Variability in Child Speech (VariCS) corpus is a collection of a range of speech data from 275 Scottish primary school children aged five to twelve years. The aim of the speech corpus is to systematically describe variability in child speech development from a cross-sectional and longitudinal perspective and to establish typical ranges for acoustic measures associated with the speech subsystems of respiration, phonation, resonance, and articulation.

The data are collected from three cohorts of children at four time points over the course of 18 months; we have currently collected three time points with another time point to follow. At each time point children are asked to complete a series of speech tasks including a story retell, a picture description, repetitions of sentences, single word productions, sound prolongations and rapid syllable repetitions. The data is complemented by ultrasound data, which was collected from a subgroup of children at time point 1 and 2. Data collection is taking place in primary schools in seven Scottish councils, representing rural and urban areas of Scotland. Detailed demographic information is gathered, including age, dialectal background, languages spoken, and relevant medical information. Ranges for key acoustic measures will be made publicly available via the VariCS website [varics.ac.uk], OSF and the University of Strathclyde repository PURE.

1 Introduction

The Variability in Child Speech (VariCS) corpus is a large longitudinal data set of speech production tasks by 275 children (sweep 1) aged five to twelve years attending mainstream primary school in Scotland. The specific aim of the VariCS project is to systematically chart variability in children's speech for a range of speech measures associated with the subsystems of respiration, phonation, resonance and articulation. Consideration of performances associated with each of these subsystems is common in clinical practice with children who have motor speech disorders (e.g. Allison & Hustad, 2018; Kuschmann & Neill, 2015; Kuschmann & van Brenk, 2019; Lee et al., 2014; Pennington, 2008). Using this approach to systematically chart the development of speech production in typically developing children allows us to gain a comprehensive understanding of typical variability in child speech from a cross-sectional as well as longitudinal perspective, that way providing the basis for:

i) a comprehensive description of how primary school-aged child speech patterns develop and mature over time, and
ii) establishing typical ranges for acoustic measures associated with speech subsystems that will be valuable in evaluating atypical child speech.

Beyond this, given the paucity of sociophonetic studies on children aged between 5 to 12 years old (e.g. Smith & Durham, 2019), the data is also useful for phoneticians and sociolinguists interested in exploring speech variation in primary school-aged children from a number of different Scottish regions. Similar sociophonetic research is currently underway in New Zealand looking at accent change in childhood (Wilson Black & Clark, under review). This sociolinguistic study investigates vocalic variation and change using a corpus of 131 children from Christchurch, New Zealand, aged between 3;11 years and 5;5 years. Comparable to our project, children were recorded at 6-months intervals over a time period of two years.

A review of other relevant existing child speech corpora showed that no speech data was available that would have allowed us to investigate how variability in child speech develops over time. Required key characteristics were the need for data to be longitudinal in nature to determine the developmental trajectories of speech for 5- to 12-year-old children, and to include connected speech as well as data from maximum performance tasks (MPTs), e.g. sustained vowel productions, for us to understand speech development across all four subsystems.

Existing datasets either focus on language development in infants and pre-school children, such as the ESRC International Centre for Language and Communicative Development corpus (LUCiD, www.lucid.ac.uk; ES/L008955/1) or are aimed at facilitating research into monolingual and bilingual language development (Escudero et al., 2024). Others were designed with testing speech recognition systems in mind, such as OGI Kids (Shobaki et al., 2000), UK PF STAR (Preparing for Future Multisensorial Interaction Research; Russell, 2006), or AusKidTalk (Ahmed et al., 2021). Whilst the latter corpora are relatively large, 1100, 158 and 600 children respectively, the datasets do not offer the required combination of connected speech data and samples of MPTs. The ALSPAC dataset (Avon Longitudinal Study of Parents and Children, http://www.bristol.ac.uk/alspac/) meets this criterion, however, speech samples were only collected at one time point (8-year olds), which does not allow for establishing developmental trajectories. Finally, the PHON-bank archive offers longitudinal speech samples of three children from the UK (https://phonbank.talkbank.org/), but does not lend itself to investigate patterns of variability. In conclusion, our speech data fills a data gap, providing a unique and comprehensive opportunity to determine how variability evolves in child speech over time, and to investigate to what extent the pattern reflects and predicts specific developmental trajectories.

2 Corpus description

2.1 Overview of new data

The VariCS data set is created through collection of speech recordings from Scottish primary school children aged five to twelve years. There is a paucity of information regarding the development of speech processes beyond the age of four (e.g. Beckman et al., 2017; Sadagopan & Smith, 2008), even though speech development continues until mid-adolescence (Sadagopan & Smith, 2008; Vorperian & Kent, 2007). Speech data are collected from three groups of children based on Scottish primary school years one, three and five (A = children in Primary 1; B = children in Primary 3; C = children in Primary 5). In Scotland, most children begin their first year of formal schooling (Primary 1) in mid-August between the ages of 4;6 and 5;6; children at the start of Primary 3 are usually between the ages of 6;6 to 7;6 and accordingly children in Primary 5 are 8;6 to 9;6 years old at the start of the school year. Also, in Scotland education policy is "presumption of mainstream" this means that most children with additional support needs are catered for in mainstream classrooms. Children with more complex needs often attend specialist provision and were not included in our sample.

Each cohort is followed over 18 months with recordings made at six-month intervals to allow for longitudinal analyses. At the end of the data collection, speech data from children aged 5 through to 12 years will have been collected. Specifically, the data set will comprise:

- **Sound files**: For most children, each recording consists of a story retell, a picture description, a set of six repeated sentences, 22 single words produced three times, syllable repetitions (/pɛ/, /tɛ/, /kɛ/, /pɛtɛkɛ/) and elongated sound productions (/a/, /s/, /z/), each produced twice. The speech recordings are saved as .wav files. (For information on the specific speech tasks please see section 2.9 and appendices.)
- **Documentation:** Documentation includes a description of the dataset, information on collection and annotation procedures, as well as hardware and software used to collect and analyse the data.

Following annotation, ranges for key acoustic measures will be determined (cf. section 2.10). This will form the basis for an online look-up facility that will allow speech and language therapists (SLTs) working with children who have speech and voice disorders to look up reference norms for each measure. The look-up facility will allow SLTs to assess how a child's performance compares to that of typically developing children, to identify risk of atypical development, and to monitor treatment progress over time.

2.2 Ethical approval and Consent procedures

Ethical approval to conduct the study was granted by the departmental ethics committee at the University of Strathclyde. Ethical requirements stipulated an opt-in approach with parents consenting for their children to take part in the study. Consent was also sought from parents to store the speech data anonymously on a secure server and for speech samples to be made available on the VariCS website. In addition, written assent was sought from the children on the day of data collection. Due to the wide age range two different versions of the participant information sheet and consent form were created. The participant information sheets for parents/carers and children as well as the consent forms are available through OSF [https://osf.io/9dgkz/] and the University's repository [https://pureportal.strath.ac.uk/en/datasets/varics-data-collection-instruments].

The research team collecting the data are all members of the Protecting Vulnerable Groups (PVG) scheme (https://www.mygov.scot/pvg-scheme). This is a membership scheme for people who work with children and vulnerable adults in Scotland.

2.3 Access to schools

In Scotland, permission to conduct research in schools is given by each local authority following an application process specific to each council. Councils are local government authorities, which provide a range of services to people living in these areas including education. Following ethical approval, 18 local authorities in the central belt of Scotland were approached, of which 13 agreed to consider the application. Following submission of these applications, permission to approach primary schools was granted by eight local authorities' education services, allowing the team to approach headteachers in these areas (cf. section 2.4 for information on participating councils). As a result of this complex approval process, the project involved regional sampling across a wide geographical area.

Head teachers from schools in seven local authorities agreed to be part of the project. Interested schools were asked to electronically distribute information about the study to parents/carers of eligible children. As indicated above, parents/carers were asked to consent for their children to take part and to complete an online questionnaire providing information on their child's linguistic background, their own language/dialect, and any medical information that might affect speech production such as asthma. The questionnaire is available through OSF [https://osf.io/9dgkz/] and the University's Repository [https://pureportal.strath.ac.uk/en/datasets/varics-data-collection-instruments].

2.4 Recording Locations

Data is collected in 23 primary schools across seven Scottish councils. Participating councils include Glasgow City and City of Edinburgh, North and South Lanarkshire, West Dunbartonshire, Fife and Scottish Borders (cf. Figure 1). The councils represent rural and urban areas across the central belt of Scotland. Meta-data information on schools includes council and Scottish Index of Multiple Deprivation (SIMD) rank, which is an area-based measure of relative deprivation.

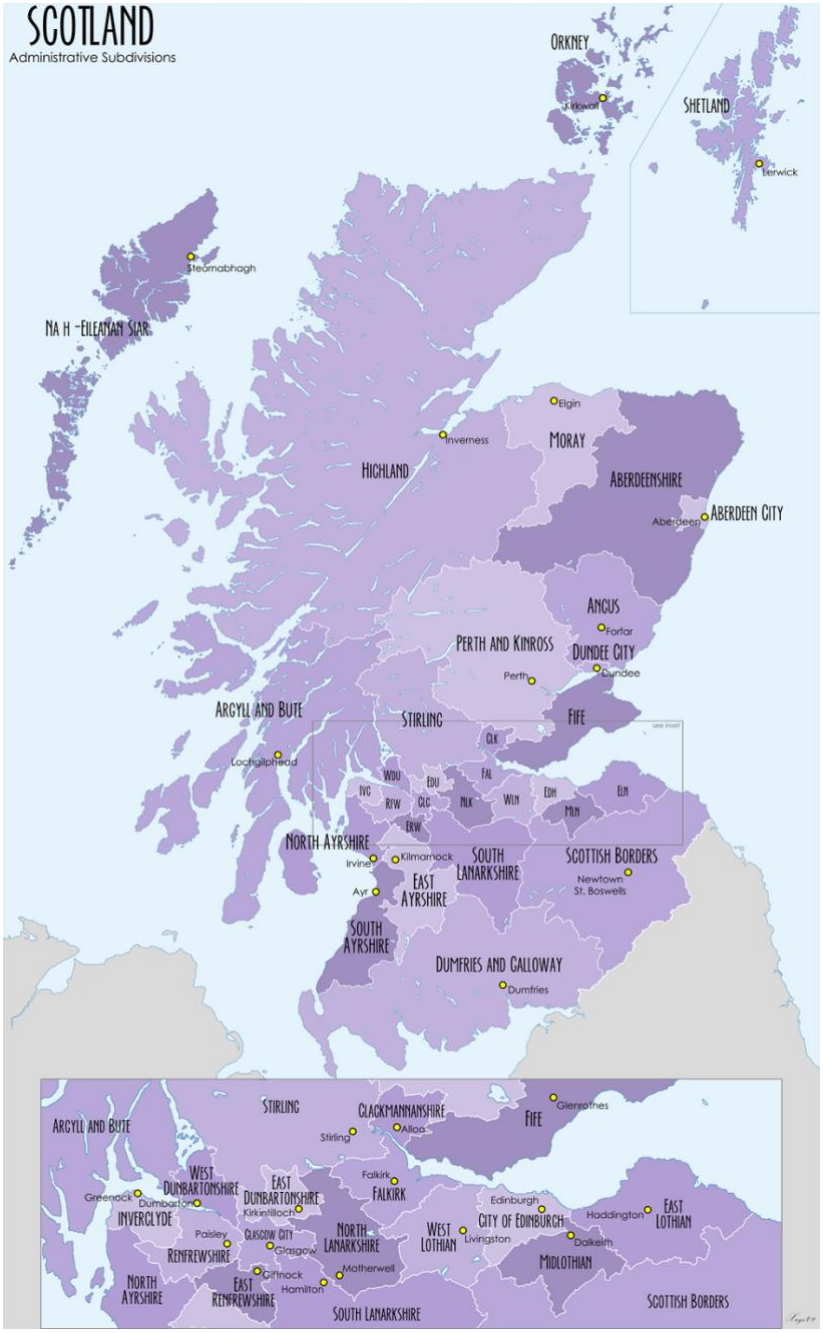

Figure 1: Map of the administrative councils in Scotland (*Scotland Administrative Subdivisions*, Wikipedia).

Recording locations are:

- 6 primary schools in Glasgow (GC)
- 5 primary schools in Fife (FI)
- 3 primary schools in the Scottish Borders (SB)
- 3 primary schools in West Dunbartonshire (WD)
- 3 primary schools in South Lanarkshire (SL)
- 2 primary schools in North Lanarkshire (NL)
- 1 primary school in Edinburgh (CoE)

2.5 Speaker demographics and sample size

As outlined in 2.1, the speech of children aged five to 12 years is recorded, with the three cohorts of children (A = children in Primary 1; B = children in Primary 3; C = children in Primary 5) being followed four times ('Sweeps') over a period of 18 months (cf. Figure 2 for timeline and dates of each Sweep's data collection). For each child, the following information is gathered: age, sex assigned at birth, languages and dialects spoken by the child as well as parents/carers, as well as information relevant to speech production such as asthma or any other known medical diagnoses. In addition, parents/carers were asked to report any communication-related issues and whether children were seen by speech and language therapy services.

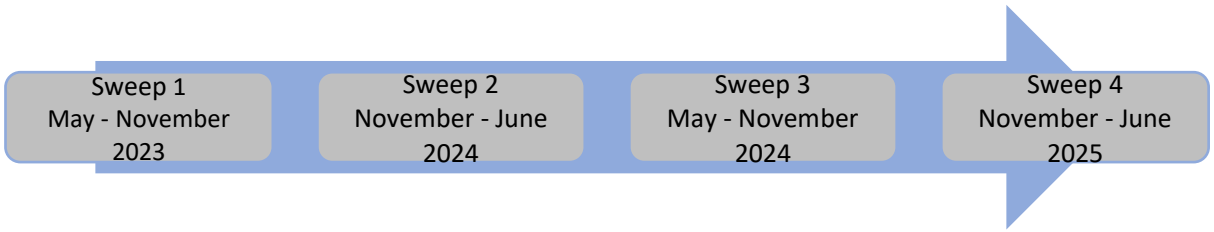

Figure 2: Timeline of each of the four sweeps of data collection.

In terms of sample size, the target number of children for the project was 500. At Sweep 1, speech recordings of 275 children were made. The overall number of participants was the result of a combination of factors including councils' decisions to either grant permission to access schools or not, interest from headteachers for the schools to be part of the study, as well as the opt-in nature of the consent procedures. At Sweep 2, 248 children were recorded. Attrition at Sweep 2 was due to withdrawing participation (1), relocating (9), or absence (17) during the data collection period. Information regarding the final number of children involved at each data collection point will be updated on conclusion of Sweeps 3 and 4 in Table 1, which provides information on the number of children recorded in each council per cohort and sex assigned at birth.

Table 1: Number of children at data collection point 1 with information on council, cohort and sex assigned at birth.

Council	**Sweep 1 total**	**Male**	**Female**	**A1** (5;0-6;5)	**B1** (5;9-8;3)	**C1** (8;7-11;9)	**Sweep 2 total**	**Male**	**Female**	**A2** (5;6-6;9)	**B2** (6;4-8;8)	**C2** (9;1-12;3)
Fife	64	29	35	20	18	26	59	27	32	18	17	24
Scottish Borders	47	21	26	16	20	11	41	19	22	14	16	11
West Dunbartonshire	36	18	18	11	15	10	34	17	17	11	13	10
Glasgow City	62	29	33	16	31	15	58	26	32	16	29	13
North Lanarkshire	26	11	15	4	11	11	21	9	12	2	9	10
South Lanarkshire	33	16	17	15	13	5	28	13	15	13	11	4
City of Edinburgh	7	4	3	6	1	0	7	4	3	6	1	0
Totals	**275**	**128**	**147**	**88**	**109**	**78**	**248**	**115**	**133**	**80**	**96**	**72**

2.6 Data collection

The speech data is collected using an app run from an iPad that moves through the different speech activities. The app was specifically designed by *Therapy Box*^{LTD} for the purpose of the study to ensure a unified data presentation and collection approach that uses child-friendly and motivating materials (cf. Figure 3). Speech recordings are made through the app, with the audio signal being stored in .wav format.

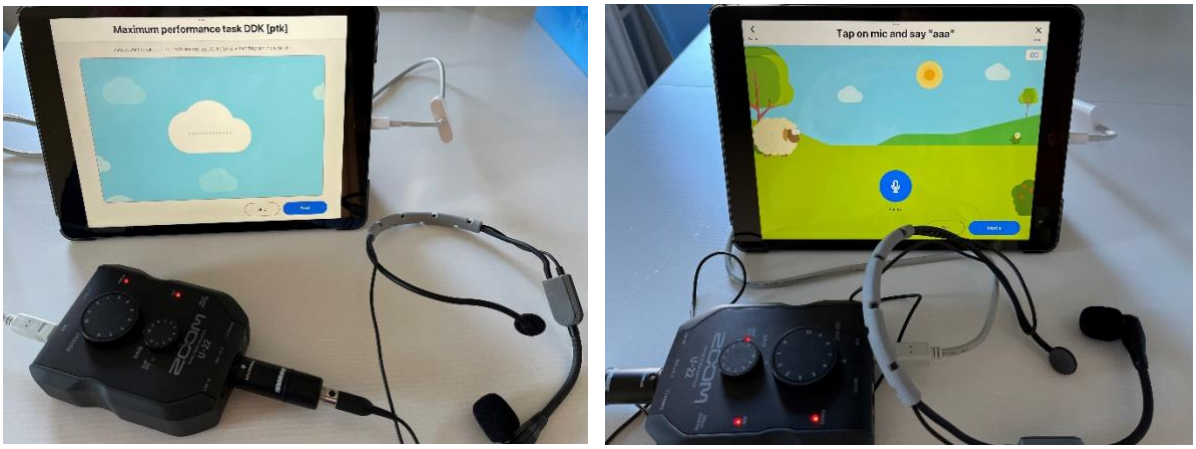

Figure 3: VariCS app showing the interface for some speech tasks, and the equipment setup.

Researchers from the Department of Psychological Sciences and Health, University of Strathclyde, collect the data in schools. The research team requests access to a closed room to minimise background noise, that way ensuring an appropriate recording quality. Once children provided written assent to participate, children were guided through six tasks, with one task split into three parts, in the following order: story retell, word naming task (1), maximum sound prolongation, maximum repetition rate, word naming task (2), sentence repetition, word naming task (3), and picture description (cf. section 2.9 for details). The Diagnostic Evaluation of Articulation and Phonology Screen (DEAP, Dodd et al., 2002) was administered in the first Sweep to detect speech sound disorders and to record common articulatory distortions. Children with suspected speech sound disorders or other speech, language or communication difficulties were not excluded from the sample. The entire procedure took 15-20 minutes per child.

2.7 Recording equipment

A unidirectional head-mounted microphone (Shure SM35) is used to minimise background noise and to ensure a constant mouth-to-microphone distance (7 cm), which is required for accurate sound pressure level and vowel formant measures, especially higher formants, such as F3 and F4. Recordings are made in 32 bit mono, sampled at 44.1KHz. Figure 3 shows some of the equipment used to make the speech recordings.

2.8 File naming conventions

Data for each child is stored in individual files. Naming conventions will be updated following completion of Sweep 3 and 4.

2.9 Recorded speech data

At each data collection point, children are asked to complete a series of speech production tasks that will allow analyses of within- and between-speaker variability across the four time points to determine developmental trajectories of child speech development and maturation patterns. Specifically, the different tasks elicit unstructured and structured connected speech, single words, syllables and prolonged sounds:

i) Unstructured connected speech is elicited through:
   - Story retell
   - Picture description

ii) Structured connected speech is elicited through:
   - Repetition of 6 sentences

iii) Single words are elicited through:

- Naming of 22 words (repeated three times at each data point)

iv) Syllables and prolonged sounds are elicited through:

maximum sound prolongation of /a/, /s/ and /z/

maximum repetition rate of /pɛ/, /tɛ/, /kɛ/, /pɛtɛkɛ/

This combination of speech tasks provides us with ecologically valid speech in the form of connected speech as well as samples that can assess the maximum performance capacity of speech subsystems. In the following sections, the different type of speech data and their elicitation method are described in greater detail.

2.9.1 Unstructured connected speech

A story retell is used to elicit unstructured connected speech. Children are asked to watch the *Treasure Story*, a short animated video developed by *Therapy Box*^{LTD} as part of the *Language Explorer* project (https://www.languageexplorer.app/; Bright et al., 2023), and are subsequently asked to retell the story with the support of a series of pictures. The story is used as licensed by Therapy Box with the exception that we re-recorded the narrative with a central Scottish English speaker. In addition, children are asked to look at a picture (Playpark picture; Patel & Connaghan, 2014) and describe what they see in as much detail as possible. The picture, which is available as part of the above publication, was originally designed for children with motor speech disorders in mind, containing items to elicit a range of speech sounds and words with e.g., a varying number of syllables.

2.9.2 Sentences

Structured connected speech is elicited through a sentence repetition task. This allows the elicitation of speech samples that are the same across all children. Children hear a sentence and are asked to repeat it. The sentence is also presented in written form. Six sentences containing a high number of different plosives were selected from the speech assessment protocol for speech disorders associated with cleft palate and/or velopharyngeal dysfunction (GOS.SP.ASS.; Sell, Harding & Grunwell, 1997). The sentences are balanced regarding the voicing status of the plosive and its place of articulation. They vary in terms of number of plosives, number of words and number of syllables. Most of the sentences contain personal names, some of these were changed to reflect more contemporary names in a Scottish context (e.g. Tim was changed to Tom). The final list of sentences can be found in appendix A.

2.9.3 Single words

Single words are elicited through a picture naming task. At each data collection point, the picture naming task is repeated three times to allow for variability analyses. A list containing 22 words was specifically created to capture the Scottish corner vowels /i/ (FLEECE), /a/ (CAT) and /o/ (GOAT) /) in one-syllable words (CVC structure). Note that Scottish English has only three corner vowels, differing from Southern British English in that the Scottish equivalent to GOOSE and FOOT is a single close front/central vowel, BOOT, /ʉ/, and the Scottish equivalent to BATH, TRAP, and PALM, is a single open central vowel, CAT, /a/ (see Figure 4; Stuart-Smith, 2008). Specifically, three words were selected for each of the Scottish corner vowels, prioritising those with the highest imageability and an age of acquisition under the age of 5. In addition, one word with the highest imageability rating was selected for each of the remaining Scottish monophthongs (/e/ FACE, /ɪ/ KIT, /ɛ/ DRESS, /ʌ/ STRUT, /ɔ/ COT, and /ʉ/, BOOT). The single words also include words with specific consonantal sounds to investigate realisation of /r/ in onset and coda, potential l-vocalisation and sibilants in specific vowel contexts. Finally, two non-words, *meem, beeb,* were included to allow investigation of nasalisation through acoustic measures (e.g. A1–P1; Lee et al., 2014). The complete wordlist and pictures used (created by *Therapy Box*) are included in appendix B.

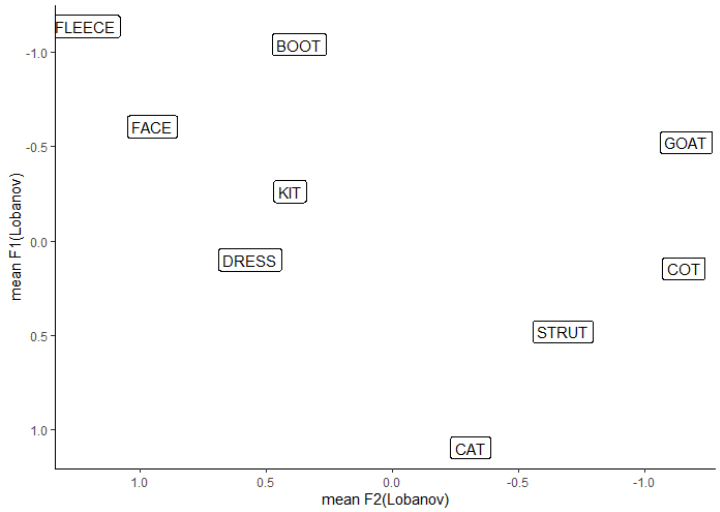

Figure 4: Acoustic quality of Scottish English monophthongs as mean Lobanov normalized formant values for lexical keywords, for 1159 speakers of Scottish Standard English and regional Scots, from the Scottish SPeech Across Dialects of English (SPADE) corpora (Smith et al., 2024).

2.9.4 Syllables and single sounds

Maximum performance tasks including maximum sound prolongation of /a/, /s/ and /z/ and maximum repetition rate of /pɛ/, /tɛ/, /kɛ/, /pɛtɛkɛ/ are also elicited at each data collection point. These tasks

are also referred to as non-speech or para-speech tasks (Kent et al., 2015), and are used to quantify the physiological functioning of the different speech subsystems (e.g. Kent et al., 2021; Speyer et al., 2010; Wit et al., 1993). For the sound prolongation task, children are asked to take a deep breath and produce the sounds "as long as you can in one breath". Similarly, for the task eliciting maximum repetition rate, children are asked to take a deep breath and to repeat the syllables and syllable combinations "as fast as you can in one breath." Each sound and syllable are produced twice.

2.10 Measures

Based on the above speech samples, a range of acoustic parameters and derived speech measures associated with each speech subsystem is measured to chart developmental trajectories and variability patterns (cf. Table 2). Ranges for these measures will be made available through OSF and the University's Repository.

Table 2: Key measures taken from the data.

Subsystem	**Acoustic and speech measures**
Respiration	Maximum sound prolongation, Sound pressure level, s/z ratio
Phonation	Fundamental frequency (F0) (mean, minimum, maximum, range), perturbation measures including shimmer, jitter, harmonic to noise (HNR) ratio, (Smoothed) Cepstral Peak Prominence (CPP/CPPS)
Resonance	Formant frequency difference between amplitudes of F1 and extra peak (A1–P1)
Articulation	Maximum repetition rate, mean vowel and word duration, formant frequencies (F1, F2, F3, F4) for corner vowels, vowel space area, speech rate, articulation rate

3 Nested Ultrasound study

3.1 Overview

A nested PhD study collected additional Ultrasound Tongue Imaging (UTI) data from a subset of participants. Children were invited to additional consent to collection of UTI over the first two time points with a recruitment target of 60 children distributed evenly across the different age groups.

The study focuses on just one of the subsystems of speech: articulation. The aim of this project is to describe sub phonemic variability, using UTI, in child speech development from both a cross-sectional and longitudinal perspective using established articulatory measures. The additional study was subject to the same ethical and consent procedures as described above. Due to the target of 60 participants only two councils (Fife and Glasgow) were given participant information sheets and

consent forms that contained the ultrasound study [https://osf.io/9dgkz/]. Parents in five schools in Fife and four in Glasgow chose to opt-in for the nested study.

3.2 Participant demographics

At data collection point 1, recordings of 51 participants were made from 59 participants who consented and are shown in Table 3 below. Attrition was due to absence (n=2), missing data (n=1), headset not fitting (n=1) and no consent from child on day (n=4). A further three participants were lost at point 2 due to no consent on the day, resulting in 48 participant recordings at time point 2.

Table 3: Participant demographics total number by council, gender and cohort for Sweeps 1 and 2.

Council	Sweep 1 Total	Male	Female	A1	B1	C1	Sweep 2 Total	Male	Female	A2	B2	C2
Fife	31	9	22	10	12	9	30	9	21	9	12	9
Glasgow	20	10	10	8	8	4	18	8	10	7	8	3

3.3 Data collection

Before the recordings, the children were fitted with the Ultrafit lightweight plastic headset, which held a 5-8 MHz 10mm radius microconvex ultrasound probe (Articulate Instruments Ltd., 2008). The probe was positioned in the midline under the chin to collect mid-sagittal ultrasound images and simultaneous audio recording (cf. Figure 5). The headset ensures minimal movement across the planes in relation to the head (Pucher et al., 2020). Recordings for the main data were made during the same session either before or after the ultrasound activities.

Figure 5: The Ultrafit headset worn by a 5-year-old child. Green arrow shows position of microphone.

3.4 Recording setup

Synchronised Audio and ultrasound video were recorded using a portable laptop based set up. Audio was captured using an Audio Technica3350 clip on microphone at a sample rate of 44.1kHz. The raw ultrasound was recorded using a micro machine controlled via a windows laptop running Articulate Assistant Advanced (AAA v. 221.2.0) (Articulate Instruments Ltd » AAA, 2014). The frame rate was set at 132 fps with a field of view of 123°.

3.5 Ultrasound materials

In order to study the within-speaker sub phonemic variability of English consonants, the children were asked to produce repetitions of selected English consonants within an /a/ vowel context. A smaller number of consonants were selected to be produced as real words within a carrier phrase to imitate a more natural speech sample, an approach used in previous work (Lenoci et al., 2021; Lenoci & Ricci, 2018; Zharkova, 2017; Zharkova et al., 2011, 2012, 2015). The full list of consonants recorded are listed below:

Table 4: List of consonants and carrier phrases elicited from children.

Prompt	Repetitions	Target Consonant
ata	10	/t/
ana	10	/n/
ara	10	/ɹ/
asa	10	/s/
asha	10	/ʃ/
ala	10	/l/
aka	10	/k/
aga	10	/g/
aya	10	/j/
acha	10	/tʃ/
atha	10	/θ/
afa	10	/f/
awa	10	/w/
apa	10	/p/
ama	10	/m/
bang	10	/ŋ/
It's a Tan	6	/t/
It's a Can	6	/k/
It's a Ram	6	/ɹ/
It's a Sham	6	/ʃ/
It's a Sam	6	/s/

3.6 Measures

A scoping review was carried out to identify articulatory measures used for investigating within speaker variability using UTI (Smith et al., under review). The mostly commonly used measure for analysing within-speaker token-to-token variability is the mean Nearest Neighbour Distance (Zharkova et al., 2012), which can be used to quantify the average distance between tongue curves. Lower mNND indicates a lower level of variability and more mature speech motor control. We use this measure to calculate the average distance between each tongue curve of the repetitions for each sound studied.

4 Summary

The VariCS speech corpus fills a data gap, combining a range of desired large data base features including a relatively large number of speakers, a variety of speech production tasks per child, multiple recording sessions, and high-quality audio recordings. This is complemented by a sizeable data set for ultrasound in a subgroup of children. The VariCS corpus is unique in its format and is expected to contribute valuable insights into child speech development and its variability from a clinical and sociolinguistic perspective.

Acknowledgements

The corpus data is collected as part of the UK Economic and Social Research Council Grant [ES/W003244/1] on 'Variability in Child Speech (VariCS)'. We thank the head teachers of the 23 participating schools for their interest in and commitment to our longitudinal study and the children for being part of this speech project.

How to cite this work

Kuschmann, A., Murali, M., Taylor, L., Smith, A., Cleland, J., Stuart-Smith, J. & Young, D. (2024), Investigating Variability in Child Speech (VariCS) – the VariCS corpus. Version 1.0.

OSF link to project

https://osf.io/xv4u6/

References

Ahmed, B., Ballard, K.J., Burnham, D., Sirojan, T., Mehmood, H., Estival, D., Baker, E., Cox, F., Arciuli, J., Benders, T., Demuth, K., Kelly, B., Diskin-Holdaway, C., Shahin, M., Sethu, V., Epps, J., Lee, C.B., Ambikairajah, E. (2021) AusKidTalk: An Auditory-Visual Corpus of 3- to 12-Year-Old Australian Children's Speech. *Proc. Interspeech* 2021, 3680-3684, doi: 10.21437/Interspeech.2021-2000

Allison, K. M., & Hustad, K. C. (2018). Acoustic predictors of pediatric dysarthria in cerebral palsy. *Journal of Speech, Language, and Hearing Research, 61*, 462–478.

Articulate Instruments. (2014). Articulate Assistant Advanved Ultrasound Module user manual revision 2.16.

Articulate Instruments Ltd. (2008). Ultrasound Stabilisation Headset—Users Manual, revision 1.5. http://Www.Articulateinstruments.Com/.

Beckman, M. E., Plummer, A. R., Munson, B., & Reidy, P. F. (2017). Methods for eliciting, annotating, and analyzing databases for child speech development. *Computer Speech & Language, 45*, 278–299.

Bright, R., Ashton, E., McKean, C. & Wren, Y. (2023). The development of a digital story-retell elicitation and analysis tool through citizen science data collection, software development and machine learning. *Frontiers in Psychology, 14*.

Dodd, B., Zhu, H., Crosbie, S., Holm, A., & Ozanne, A. (2002). *Diagnostic evaluation of articulation and phonology (DEAP)*. London: Psychological Corporation.

Escudero, Paola, Pino Escobar, Gloria, Hernandez Gallego, Milena, Diskin-Holdaway, Chloé, & Hajek, John. (2024). A database of multilingual child speech with recordings from a longitudinal project for multilingual education. In Olga Maxwell & Rikke Bundgaard-Nielsen (Eds.), *Proceedings of the Nineteenth Australasian International Conference on Speech Science and Technology* (pp. 97-101). Australasian Speech Science and Technology Association. [pdf]

Kent R. D. (2015). Nonspeech Oral Movements and Oral Motor Disorders: A Narrative Review. *American Journal of Speech-Language Pathology, 24*(4), 763-789.

Kent, R. D., Yunjung Kim, & Chen, L.-M. (2021). Oral and Laryngeal Diadochokinesis Across the Life Span: A Scoping Review of Methods, Reference Data, and Clinical Applications. *Journal of Speech Language and Hearing Research*, *65*(2), 574–623.

Kuschmann, A., & Neill, R. (2015). Developmental dysarthria in a young adult with cerebral palsy: A speech subsystems analysis. In: The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences. Glasgow, UK*.

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

Lee, J., Hustad, K. C., & Weismer, G. (2014). Predicting speech intelligibility with multiple speech subsystem approach in children with cerebral palsy. *Journal of Speech, Language, and Hearing Research, 57*, 1666–1678.

Lenoci, G., Celata, C., Ricci, I., Chilosi, A., & Barone, V. (2021). Vowel variability and contrast in Childhood Apraxia of Speech: acoustics and articulation. Clinical Linguistics & Phonetics, 35(11), 1011–1035. https://doi.org/10.1080/02699206.2020.1853811

Lenoci, G., & Ricci, I. (2018). An ultrasound investigation of the speech motor skills of stuttering Italian children. Clinical Linguistics & Phonetics, 32(12), 1126–1144. https://doi.org/10.1080/02699206.2018.1510983

Patel, R. & Connaghan, K. (2014). Park Play: A picture description task for assessing childhood motor speech disorders. *International Journal of Speech-Language Pathology, 16*(4), 337-343.

Pennington, L. (2008). Cerebral palsy and communication. *Paediatrics and Child Health*, *18*(9), 405-409.

Pucher, M., Klingler, N., Luttenberger, J., & Spreafico, L. (2020). Accuracy, recording interference, and articulatory quality of headsets for ultrasound recordings. *Speech Communication*, *123*, 83–97. https://doi.org/10.1016/j.specom.2020.07.001

Russell, M. (2006). *The PF-STAR British English Children's Speech Corpus*. Version 1. The Speech Ark Limited.

Sadagopan, N., & Smith, A. (2008). Developmental changes in the effects of utterance length and complexity on speech movement variability. *Journal of Speech Language & Hearing Research, 51*(5), 1138–1151.

Scotland administrative map. [Map]. Retrieved 9 December 2024, from https://simple.wikipedia.org/wiki/Council_areas_of_Scotland#/media/File:Scotland_Administrative_Map_2009.png

Sell D., Harding, A., & Grunwell P. (1997). A screening assessment of cleft palate speech (Great Ormond Street Speech Assessment). International Journal of Language and Communication Disorders, 29(1), 1-15.

Shobaki, K., Hosom, J. P., & Cole, R. A. (2000). The OGI Kids' speech corpus and recognizer. In: *Proceedings of ICSLP 2000*, 258–261.

Smith, A., Kuschmann, A., Lawson, E., Dokovova, M., & Cleland, J. Instrumental articulatory techniques investigating lingual variability in typically developing children: A scoping review. [manuscript under review]. Department of Psychological Sciences & Health. University of Strathclyde.

Smith, J., Stuart-Smith, J., Macdonald, R., & Jamieson, E. (2024). Scots and Scottish Standard English. In S. Fox (ed.). *Language in Britain and Ireland*. Third Edition: 151–177. Cambridge University Press. https://doi.org/10.1017/9781108769617.009

Smith, J., & Durham, M. (2019). *Sociolinguistic Variation in Children's Language: Acquiring Community Norms* (Studies in Language Variation and Change). Cambridge University Press.

Speyer, R., Bogaardt, H. C. A., Passos, V. L., Roodenburg, N. P. H. D., Zumach, A., Heijnen, M. A. M., Baijens, L. W. J., Fleskens, S. J. H. M., & Brunings, J. W. (2010). Maximum phonation time: Variability and reliability. *Journal of Voice*, *24*(3), 281–284.

Stuart-Smith, J. (2008). Scottish English: phonology. *The British Isles*. Volume 1. Kortmann, B., Upton, C. and Schneider, E. W. (eds): 48-70. De Gruyter Mouton.

Vorperian, H. K., & Kent, R. D. (2007). Vowel Acoustic Space Development in Children: A Synthesis of Acoustic and Anatomic Data. *Journal of Speech, Language & Hearing Research, 50*(6), 1510-1545.

Wilson Black J. & Clark, L. (under review). In search of the onset of vernacular reorganisation.

Wit, J., Maassen, B. & Gabreels, F., & Thoonen, G. (1993). Maximum performance tests in children with developmental spastic dysarthria. *Journal of Speech & Hearing Research, 36*, 452–459.

Zharkova, N. (2017). Voiceless alveolar stop coarticulation in typically developing 5-year-olds and 13-year-olds. *Clinical Linguistics & Phonetics*, *31*(7–9), 503–513. https://doi.org/10.1080/02699206.2016.1268209

Zharkova, N., Gibbon, F. E., & Hardcastle, W. J. (2015). Quantifying lingual coarticulation using ultrasound imaging data collected with and without head stabilisation. *Clinical Linguistics & Phonetics*, *29*(4), 249–265. https://doi.org/10.3109/02699206.2015.1007528

Zharkova, N., Hewlett, N., & Hardcastle, W. J. (2011). Coarticulation as an indicator of speech motor control development in children: An ultrasound study. *Motor Control*, *15*(1), 118–140. https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2011-03921-007&authtype=shib&site=ehost-live&authtype=ip,shib&custid=ns075038

Zharkova, N., Hewlett, N., & Hardcastle, W. J. (2012). An ultrasound study of lingual coarticulation in /sV/ syllables produced by adults and typically developing children. *Journal of the International Phonetic Association*, *42*(2), 193–208. https://doi.org/10.1017/S0025100312000060

Appendices

A GOS.SP.ASS. sentences

Sentence	Consonant	Place	Manner	Voicing	No. of plosives	No. of words	No. of syllables
Tiny Tom is putting a hat on.	t	alveolar	plosive	n	5	7	9
My dad mended a door.	d	alveolar	plosive	y	5	5	7
Happy Cara is making a cake.	k	velar	plosive	n	6	6	9
Baby Gaby's got a bag of Legos.	b/g	bilabial, velar	plosive	y	7	7	10
The puppy is playing with a rope.	p	bilabial	plosive	n	4	7	9
Bouncy Ben is a baby boy.	b	bilabial	plosive	y	5	6	8

Nb.: Selected names changed from original sentences – Tim > Tom; Karen > Cara; Bob > Ben.

B single word list and pictures

Vowels/ consonants	Words	AOA (years)	Imageability
i	teeth	3.61	611
	feet	3.44	597
	sheep	4.25	596
a	bat	4.85	586
	cat	3.68	617
	hat	3.33	562
o	coat	3.58	572
	boat	3.84	631
	soap	3.17	600
e	cake	3.26	624
ɪ	fish	4.05	615
ɛ	bed	2.89	635
ʌ	cup	3.57	558
ɔ	sock	2.94	553
~~ʉ~~	boot	3.89	251
r	car	3.37	638
	ring	4.53	601
l	hill	3.47	607
	lip	3.79	619
s	sing	3.47	527
i (nonsense words)	meem		
	beeb		

Nb.: AOA = age of acquisition; Imageability index and AOA based on the MRC Psycholinguistic database. Rated on a scale of 100 (minimum) to 700 (maximum), with 100 = 100 x 1 [0-2 years] and 7 [13 years on] for AOA.

Word list pictures – in order of the list above