A PILOT FIELDWORK ULTRASOUND STUDY OF TONGUE SHAPE VARIABILITY IN CHILDREN WITH AND WITHOUT SPEECH SOUND DISORDER

Amy Smith, Maria Dokovova, Eleanor Lawson, Anja Kuschmann, Joanne Cleland

University of Strathclyde
amy.smith@strath.ac.uk, maria.dokovova@strath.ac.uk, eleanor.lawson@strath.ac.uk, anja.kuschmann@strath.ac.uk, joanne.cleland@strath.ac.uk

ABSTRACT

Children with Speech Sound Disorders (SSD) have been described as having increased tongue shape variability during speech. However, most studies do not compare this variability to typically developing children (TD) using ultrasound tongue imaging (UTI). Open access corpora suitable for answering this question are scarce. This pilot study addresses both the theoretical question of whether variability differs between TD children and children with SSDs; and the feasibility of a potential solution for acquiring ultrasound data quickly at a public science exhibition. We compare tongue shapes during multiple repetitions of various consonants in TD children across different ages, and children with SSD, using mean Nearest Neighbour Distance (NND). Results suggest no significant effect of age in the TD group. Children with SSD had significantly higher tongue shape variability than TD children. Field data collection was a viable method for collecting UTI data, despite some limitations.

Keywords: speech sound disorders, ultrasound, variability, articulation, field methods

1. INTRODUCTION

Speech Sound Disorder (SSD) is an umbrella term encompassing many subtypes that affect a child's ability to acquire clear or intelligible speech. With a prevalence of between 2.3% and 24.6% [1], SSDs make up a large proportion of Speech and Language therapy caseloads. Disagreement in how these disorders are classified creates challenges for Speech and Language Therapists (SLTs) to diagnose and treat SSD. However, SSDs can be broadly divided into those of a phonological nature (e.g., phonological disorders) and those of an articulatory or phonetic nature (articulation disorders, childhood apraxia of speech). Without understanding the true nature of the difficulty and due to limitations of current investigation practices, crucial information can be missed. Instrumental techniques such as Ultrasound Tongue Imaging (UTI) can negate this problem by providing detailed information about articulation, instead of relying on perceptual tools such as phonetic transcription, which have reliability issues [2].

In UTI, a probe is placed under the chin to image the midsagittal, or coronal tongue surface, displaying movement in real time. UTI is relatively inexpensive and portable, and the use of UTI in SSD research is growing [3]. Use of UTI is no longer limited to speech labs and hospital rooms and, as such, it is possible to collect data in a variety of settings. Despite the challenges field collection can present, it opens the possibility for larger N studies.

UTI has been shown to aid in the detection of covert errors in children with SSDs [4]. One of these is increased variability, meaning repetitions of the same sound in the same phonetic context are produced differently. Here we define variability as sub-phonemic, that is, repetitions are typically phonetically transcribed using the same symbol. This variability is therefore subtle and can only be detected using a modality such as UTI, in addition to e.g. auditory analysis [4]. Increased variability is suggestive of subtle motor difficulties [5]. This has strong implications for SLTs and could affect differential diagnosis and decisions surrounding therapy approaches. However, for SLTs to fully understand the articulatory variability associated with SSDs, it is essential to quantify articulatory variability in typically developing children of different ages.

In contrast to previous evidence, which suggested little to no difference in articulatory gestures made by children when compared to adults [6], recent literature suggests protracted development of speech motor control well into adolescence [7]. Younger typically developing children show greater articulatory variability than their older counterparts. What is not yet known is what this developmental trajectory looks like and what the typical range is. UTI has previously been used to measure lingual variability by comparing sets of tongue curves and supports the idea of protracted development [8]. However, as [9] points out, "there are still crucial gaps in our knowledge about the developmental paths taken by children to adult-like speech motor control".

Crucially for SLTs, we do not yet know how this compares to children with SSDs.

2. RESEARCH QUESTIONS

This is an exploratory pilot study, aiming to describe articulatory variability, as measured using mean Nearest Neighbour Distance (NND) [10] in typically developing (TD) children, and children with SSD across four key consonants: /t/, /k/, /ɹ/, /s/. Higher mean NND are suggestive of increased variability.

This study explores the following questions:

- Is it feasible to collect ultrasound data from children at a public event in a museum?
- Do older TD children have lower articulatory variability than younger TD children?
- Do TD children have lower articulatory variability than age-matched children with a Speech Sound Disorder?

3. METHODS

3.1. Participants and recruitment

TD children were recruited at a public engagement event (ethics approval from the University of Strathclyde). This free event was held at the Glasgow Riverside Museum on a Saturday afternoon. Recordings were made at a public stall, in a large exhibition room. Inclusion criteria were: spoken English, age 5-15 years, and no developmental conditions. Any child who approached the exhibit with their family was invited to be recorded and parents gave consent via an online survey on their mobile phone or a tablet, which also collected demographic information. They were recorded by one of five researchers present at the exhibit. SSD data was selected from the UltraSuite corpus [11]. These are one-off recordings from children with unresolved SSD attending SLT services. After the science event, the TD children were age-matched with 10 children with SSD (2 female). The children with SSD had a diagnosis of any SSD of unknown origin, given by their SLT. The exclusion criteria were: current or repaired cleft lip and/or palate, no spoken English, severe or profound hearing loss, severe or profound learning disability. They received the same instructions and produced the same set of consonants.

3.2. Data collection feasibility

To address the feasibility of data collection at a public event, the following topics were considered:

- How many parents and children would consent to recordings in a public venue?
- Would children tolerate the headset and duration of the recording?
- Would the recordings yield usable data?

3.3. Stimuli

Each child was instructed to produce 10 repetitions of all English consonants in an /aCa/ environment. Some children produced fewer than 10 repetitions. The consonants /s/ /t/ /k/ /ɹ/ were used for analysis in this pilot, as they represent high frequency segments in English, and therefore have a crucial impact on intelligibility [12]. Participants were given an orthographic representation and asked to read or were given a verbal demonstration and asked to imitate.

3.4. Ultrasound Set-Up

UTI data of the TD children were acquired using a Micro machine controlled via a Windows laptop running AAA v. 2.20.01 [13]. The echo return data were recorded at ~100 fps over a field of view of 162°. This field of view allowed the greatest view of the tongue, including both the hyoid and mandible shadows. A 5–8 MHz 10 mm radius microconvex ultrasound probe was stabilized with an Ultrafit lightweight plastic headset [14]. Audio data was recorded using an Audio Technica3350 microphone attached to the headset. The same set up and materials were used for both groups, however the SSD data were collected with Sonospeech [15] in a quiet clinical setting by SLTs trained in UTI data collection, rather than at a public event.

3.5. Analysis

3.5.1. Spline fitting

The tongue surface was identified by fitting splines at the point of maximal lingual gesture of each consonant, using AAA. The frame of interest, containing the maximal lingual gesture was identified manually, using the audio, waveform, and spectrogram. Regions of interest were observed frame by frame (stop consonant closure phases and whole durations of continuants). Splines were added using semi-automatic edge detection in AAA. Regions of low confidence were manually trimmed and the automatic AAA correction “snap-to-fit” applied.

3.5.2. Mean Nearest Neighbour Distance

The variability of tongue shapes was analysed using the mean Nearest Neighbour Distance (NND) procedure [10], implemented in AAA (version 220.04.01) at 80% minimum confidence. In this procedure splines are first analysed in pairs. The distance from each point on one spline to the nearest point on the other spline is measured then all distances between the two splines are averaged. Using this procedure, we compared each spline from one half of the repetitions of one consonant from one child to each of the splines in the other half of repetitions of that consonant from that child, resulting in a mean and standard deviation of NND per consonant per child.

3.5.3. Statistical analyses

Data processing and visualisation were performed with "tidyverse" [16] in RStudio (version 2022.02.1+461 for Windows) [17]. Statistical analyses were carried out using the lme4 package [18] (data and code available at: https://osf.io/3byvp/).

4. RESULTS

4.1. Feasibility results and descriptive statistics

Age group	Consonant	Mean NND	SD NND	n participants
[7-8]	/k/	0.10	0.04	4
	/ɹ/	0.10	0.05	4
	/s/	0.07	0.02	4
	/t/	0.08	0.03	5
[9-11]	/k/	0.07	0.02	5
	/ɹ/	0.07	0.03	5
	/s/	0.05	0.03	5
	/t/	0.07	0.02	5
[14]	/k/	0.07	0.04	1
	/ɹ/	0.07	0.02	1
	/s/	0.07	0.02	1
	/t/	0.05	0.02	1

Table 1: Mean and SD of Nearest Neighbour Distance (cm) per consonant per age group in TD children.

Eighteen children and their parents consented to participate at the public event, over a four-hour period. Of those, 7 datasets were excluded due to: corrupt files (n=1), missing audio (n=5), and 1 who had autism (identified after the event from the survey), leaving 11 TD participants (5 female). There was no age match for a 14-year-old TD participant in the SSD sample, so the analyses comparing both groups were run with and without that participant. Anecdotally, we consider this a higher rate of data wastage than previous lab-based recordings of children. Their mean age was 9 years (range 7-14). All children who consented to the recording completed the tasks, were enthusiastic, and tolerated the headset and duration of the recording. Duration of recordings ranged from 3-9 minutes (Mean=6 minutes). Recording varied in length as not all participants completed 10 repetitions of each stimulus. Table 1 provides mean and SD NND (cm) per consonant per age for the TD children.

4.2. Age

A linear mixed effects model was run with mean NND as an outcome variable for TD children and age (in years, centred around the mean) as a continuous predictor. The model included random intercepts per participant. Age was not a significant predictor for NND (Table 1).

Predictor	*β*	*t*- value	*p*- value
Intercept	0.07	12.10	< 0.001
Age	-0.002	-0.61	0.56

Table 1: Model results from the statistical analysis of age and mean NND in TD children.

4.2. Diagnosis

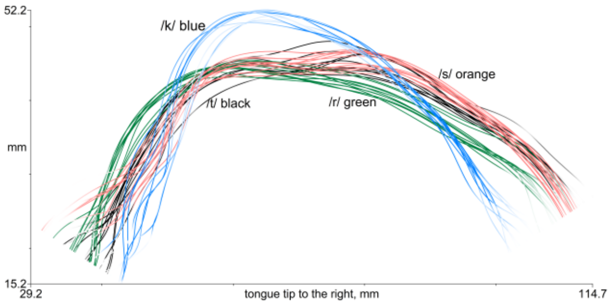

Figure 1: Tongue splines for /k/, /t/, /ɹ/, /s/ of a 10-year-old male child with SSD.

Lastly, we explored whether TD children had lower articulatory variability than age-matched children with SSD. We ran a linear mixed effects model with mean NND as an outcome variable and Diagnosis (TD as reference) as a predictor. The model included random intercepts per participant. Results suggest that SSD diagnosis is associated with increased mean NND per consonant (Table 1). Excluding the 14-year-old TD participant did not change direction or significance of the results. Fig. 1 and 2 show all tongue splines per consonant from a TD child and a child with SSD. Fig. 3 shows the differences in variability by group.

Predictor	*β*	*t*- value	*p*- value
Intercept	0.14	8.79	< 0.001
Diagnosis (TD vs SSD)	0.10	4.41	< 0.001

Table 2: Model results from the statistical analysis of diagnosis and mean NND.

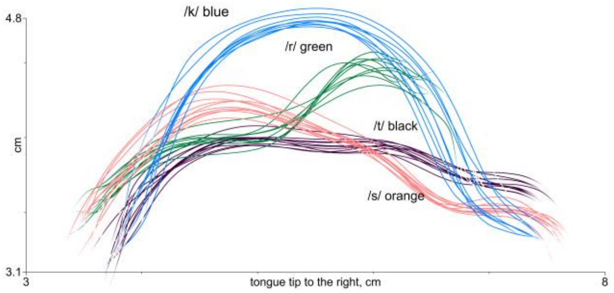

Figure 2: Tongue splines for /k/, /t/, /ɹ/, /s/ of a 10-year-old male TD child.

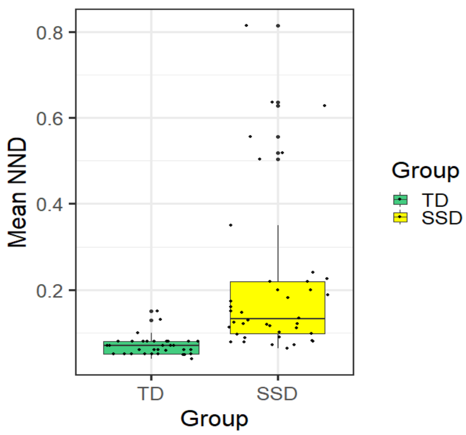

Figure 3: Differences in mean NND between TD children and children with SSD.

5. DISCUSSION

This paper provides preliminary normative data on tongue shape variability in TD children; compares this to children with SSD and describes the feasibility of collecting data at a public engagement event. Despite greater losses of data, it was feasible to collect recordings that can be used for analysis at these kinds of events. For future research we suggest a recording protocol be established to avoid discrepancies between researchers undertaking recordings and minimise experimenter error. In addition, future field research should consider recording the age of the children in years and months, for increased precision in investigating age-effects.

Tongue shape variability was measured using mean Nearest Neighbour Distance (NND) per consonant, across /s/, /t/, /k/, /ɹ/. In addition, the data was explored to investigate the relationship between age and tongue shape variability, and the relationship between having an SSD diagnosis and variability in consonant production. The results were inconclusive about the relationship between age and mean NND in TD children. However, there was a significant effect of SSD diagnosis. Children with SSD had greater tongue shape variability than children without a diagnosis.

These results differ from some previous findings. For example, Zharkova, et al. [19] report greater within-speaker variability in children than in adults, opposite to their expectation of greater variability associated with larger oral cavities. Namasivayam, et al. [20] discuss evidence demonstrating that between-articulator coordination tends to complete its development in child speech before within-articulator coordination, specifically the coordination of multiple tongue gestures. The present study has a limited number of participants (potentially contributing to the null result for the effect of age on consonant variability), which prevents a more detailed by-consonant analysis. Future research could explore differences in age-related variability of tongue gestures with only a primary lingual gesture e.g. /t/, compared with those with a primary and secondary lingual gestures such as /r/ or /l/.

The greater tongue shape variability in the SSD group is consistent with an account of variability being linked to motor skill immaturity or disorder. While earlier research such as [6] did not observe increased variability during repetitions in most children with SSD, this might also be related to differences in instrumentation. Electropalatography, which was used by Hardcastle, et al. [6] only shows which part of the palate the tongue is in contact with. The advantage of using UTI is that it allows measurements of the whole midsagittal surface of the tongue, from the root to near the tip. Future research could explore whether any specific area of the tongue contributes disproportionately to increased variability in the SSD group.

In conclusion, it was feasible to collect data from children at a public event. The likelihood of larger than usual amounts of data wastage can be mitigated by developing a recording protocol before the event. Results of this pilot suggest increased variability in children with SSD compared to TD children. Larger studies are needed to examine the effects of age.

6. ACKNOWLEDGMENTS

This research was funded by grants from the Engineering & Physical Sciences Research Council [EP/P02338X/1] and the Economic and Social Research Council [ES/W003244/1]. For Open Access, the author has applied a CC licence to any Author Accepted Manuscript (AAM) version arising from this submission. We would like to thank the Glasgow Riverside Museum and the children, for participating.

7. REFERENCES

[1] J. Law, J. Boyle, F. Harris, A. Harkness, and C. Nye, "Prevalence and natural history of primary speech and language delay: findings from a systematic review of the literature," *International Journal of Language & Communication Disorders*, vol. 35, no. 2, pp. 165-88, 2000, doi: 10.1080/136828200247133.

[2] D. Sell, "Issues in perceptual speech analysis in cleft palate and related disorders: a review," *International Journal of Language & Communication Disorders*, vol. 40, no. 2, pp. 103-121, 2005, doi: https://doi.org/10.1080/13682820400016522.

[3] E. Sugden, S. Lloyd, J. Lam, and J. Cleland, "Systematic review of ultrasound visual biofeedback in intervention for speech sound disorders," *International Journal of Language & Communication Disorders*, vol. 54, no. 5, pp. 705-728, 2019, doi: 10.1111/1460-6984.12478.

[4] E. Sugden and J. Cleland, "Using ultrasound tongue imaging to support the phonetic transcription of childhood speech sound disorders," *Clinical Linguistics & Phonetics*, pp. 1-20, 2021, doi: 10.1080/02699206.2021.1966101.

[5] J. C. Vick *et al.*, "Data-Driven Subclassification of Speech Sound Disorders in Preschool Children," *Journal of Speech, Language, and Hearing Research*, vol. 57, no. 6, pp. 2033-2050, 2014, doi:10.1044/2014_JSLHR-S-12-0193.

[6] W. J. Hardcastle, R. A. M. Barry, and C. J. Clark, "An Instrumental Phonetic Study of Lingual Activity in Articulation-Disordered Children," *Journal of Speech, Language, and Hearing Research*, vol. 30, no. 2, pp. 171-184, 1987, doi: doi:10.1044/jshr.3002.171.

[7] B. Walsh and A. Smith, "Articulatory Movements in Adolescents," *Journal of Speech, Language, and Hearing Research*, vol. 45, no. 6, pp. 1119-1133, 2002, doi:10.1044/1092-4388(2002/090).

[8] D. Abakarova, S. Fuchs, and A. Noiray, "Developmental Changes in Coarticulation Degree Relate to Differences in Articulatory Patterns: An Empirically Grounded Modeling Approach," *Journal of Speech, Language, and Hearing Research*, vol. 65, no. 9, pp. 3276-3299, 2022, doi:10.1044/2022_JSLHR-21-00212.

[9] N. Zharkova, N. Hewlett, and W. J. Hardcastle, "Coarticulation as an Indicator of Speech Motor Control Development in Children: An Ultrasound Study," *Motor Control*, vol. 15, no. 1, pp. 118-140, 2011.

[10] N. Zharkova and N. Hewlett, "Measuring lingual coarticulation from midsagittal tongue contours: Description and example calculations using English /t/ and /ɑ," *Journal of Phonetics*, vol. 37, no. 2, pp. 248-256, 2009/04/01/ 2009, doi: https://doi.org/10.1016/j.wocn.2008.10.005.

[11] A. Eshky *et al.*, "UltraSuite : a repository of ultrasound and acoustic data from child speech therapy sessions," presented at the Interspeech 2018, 2018. [Online]. Available: https://strathprints.strath.ac.uk/64824/.

[12] L. Gilner and F. Morales, "Functional load: Transcription and analysis of the 10,000 most frequent words in spoken English," *The Buckingham Journal of Language and Linguistics*, vol. 3, pp. 135-162, 2010.

[13] *Articulate Assistant Advanced Ultrasound Module user manual, revision 2.16.*, Articulate Instruments, 2014.

[14] M. Pucher, N. Klingler, J. Luttenberger, and L. Spreafico, "Accuracy, recording interference, and articulatory quality of headsets for ultrasound recordings," *Speech Communication*, vol. 123, pp. 83-97, 2020/10/01/ 2020, doi: https://doi.org/10.1016/j.specom.2020.07.001.

[15] *Sonospeech user manual*, Articulate Instruments, 2023.

[16] H. Wickham, Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H., "Welcome to the Tidyverse," *Journal of Open Source Software*, vol. 4, no. 43, p. 1686, 2019, doi: 10.21105/joss.01686.

[17] *R: A Language and environment for statistical computing*. (2010). R Foundations for Statistical Computing, Vienna, Austria.

[18] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting Linear Mixed-Effects Models Using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1 - 48, 10/07 2015, doi: 10.18637/jss.v067.i01.

[19] N. Zharkova, N. Hewlett, and W. J. Hardcastle, "An ultrasound study of lingual coarticulation in /sV/ syllables produced by adults and typically developing children," *Journal of the International Phonetic Association*, vol. 42, no. 2, pp. 193-208, 2012, doi: 10.1017/S0025100312000060.

[20] A. K. Namasivayam, D. Coleman, A. O'Dwyer, and P. van Lieshout, "Speech Sound Disorders in Children: An Articulatory Phonology Perspective," (in English), *Frontiers in Psychology*, Review vol. 10, 2020-January-28 2020, doi: 10.3389/fpsyg.2019.02998.