Ultrasound-Aided Versus Perception-Based Phonetic Transcription of Childhood Speech Sound Disorders

Eleanor Sugden and Joanne Cleland

University of Strathclyde, Glasgow, UK

Email: joanne.cleland@strath.ac.uk Twitter: @DrJoanneCleland @speechiellie

University of Strathclyde Humanities & Social Sciences

Accurate phonetic transcription is a key step in the assessment and differential diagnosis of childhood speech sound disorders[1]

Transcribers rely on perceptual information which can be unreliable, particularly if the disorder is severe or the errors are unusual[2] as is often the case in motor speech disorders

Adding a visual modality (ultrasound tongue imaging) increases the inter-rater reliability of the speech of children with cleft lip ± palate[3]. But what about other disorders?

Does ultrasound increase the accuracy or reliability of transcription of the speech of children with other speech disorders?

Method

We recorded speech from 23 English-speaking children aged 5 to 12 with speech sound disorders. Data was then transcribed in two different modalities:

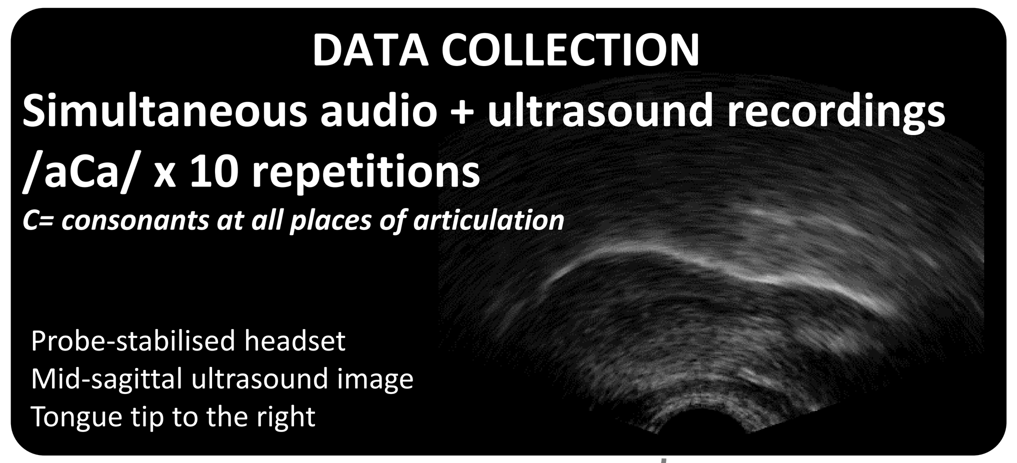
1. Ultrasound-aided transcription by two ultrasound-trained speech disorder specialist speech-language pathologists
2. Traditional audio-only phonetic transcription by two speech disorder specialist speech-language pathologists

We compared the number of consonants identified as in error by each transcriber, classified errors into 10 categories (see table), and calculated inter-rater reliability.

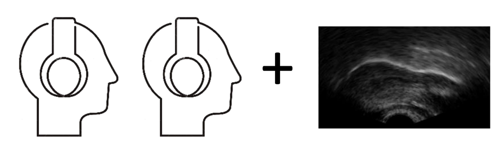
Participant (Speaker) Details

Age	• Mean 8.23 years (SD = 2.11) • Range 5;2 to 12;11
Sex	• 5 female • 18 male
Speech disorder*	• Articulation (n=11), inconsistency (n=4), phonological (n=3), other (n=5)
Accents	• All had a Scottish accent

*Reported by treating speech-language pathologist

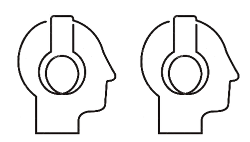


Ultrasound-Aided Transcription



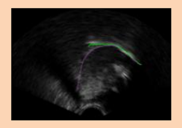
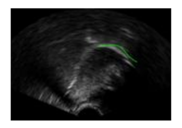
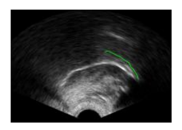
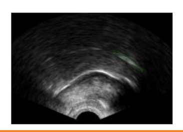
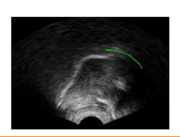
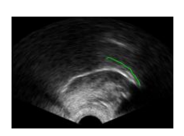
Classified into the error types based on ultrasound and audio data

Audio-Only Transcription



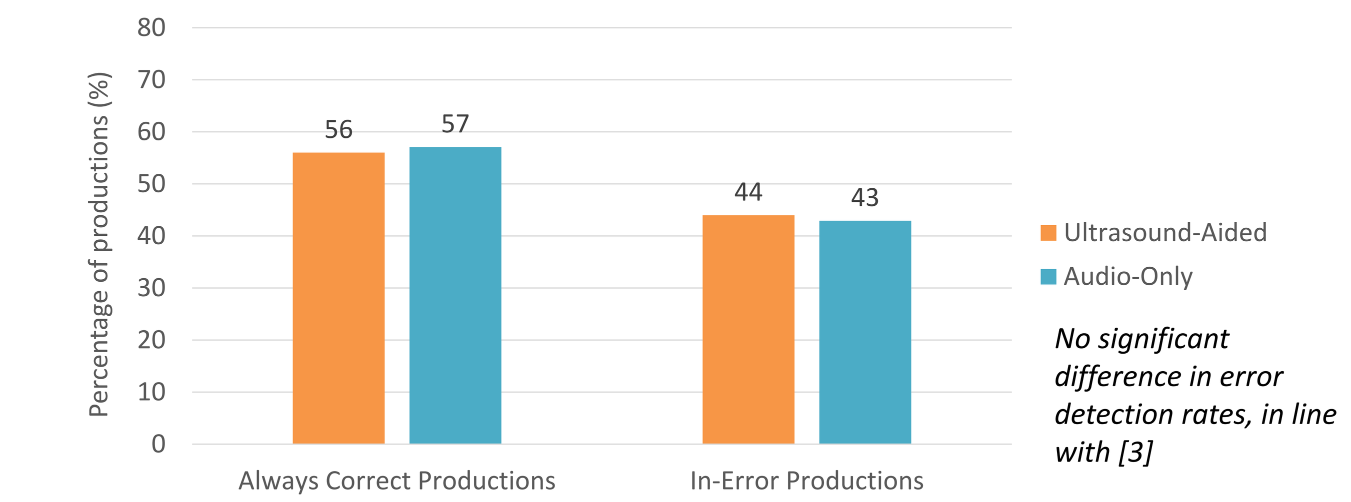
Later classified into the error types based on transcriptions

Error Types (based on Gibbon, 2004[4] and Cleland et al., 2019[3])

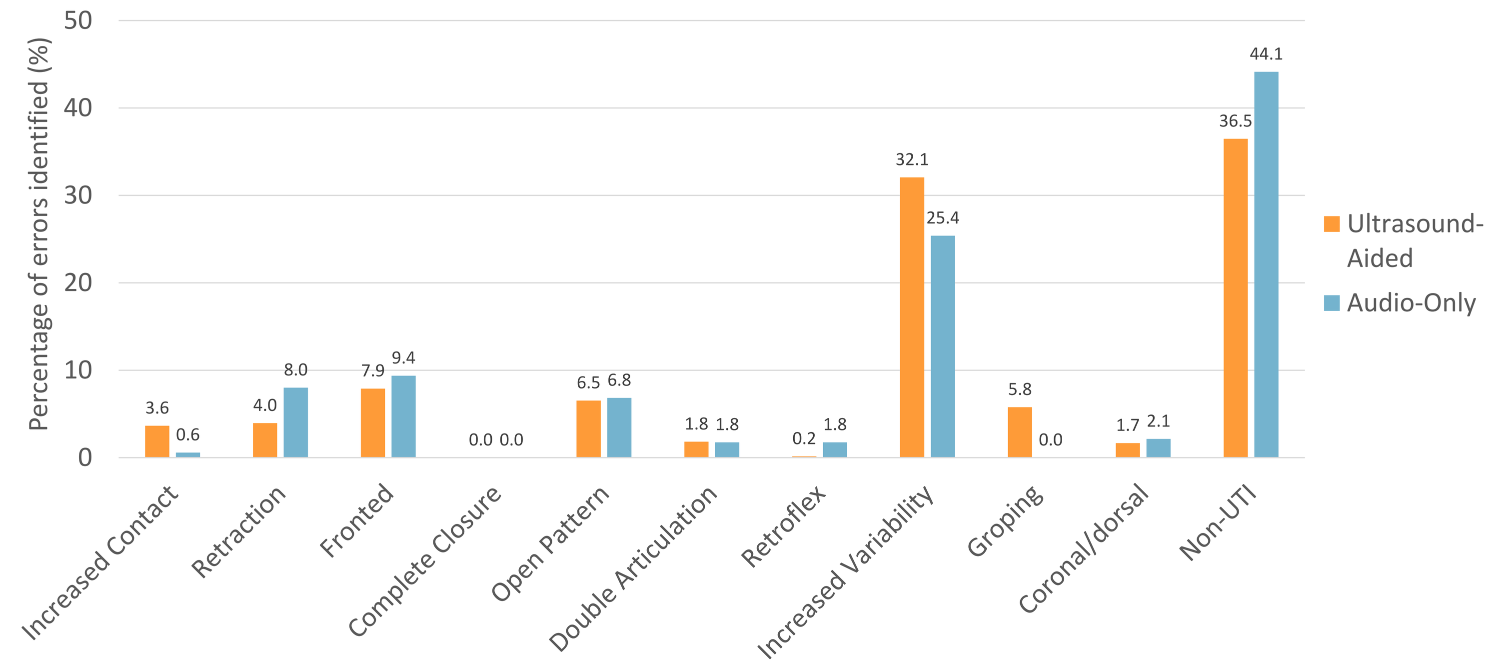
Error type	Description	Likely Cause (motoric/linguistic)	Ultrasound example	Expected transcription
Increased contact	Raising of tongue body and tip/blade towards hard palate	Motoric: Undifferentiated lingual gesture		Simultaneous alveolar + postalveolar + palatal
Retraction	Alveolar target retracted to velar or palatal	Either: backing may be due to phonological disorder or speech motor constraint		Velar or palatal
Fronted	Posterior target fronted to palatal, post-alveolar, or alveolar	Linguistic: velar fronting		Alveolar, post-alveolar, or palatal
Open pattern	Uvular or pharyngeal articulation OR undershoot	Either: backing due to phonological disorder or problems with motor execution		Uvular, pharyngeal or "lowered" diacritic
Double articulation	Simultaneous production of two consonants	Motoric		Any double articulation e.g. [k͡t] or [p͡t]
Increased variability	Different tongue-shapes per repetition	Motoric instability	*Dynamic analysis required*	Different transcriptions across repetitions
Abnormal timing/Groping	Mis-directed articulatory gestures or release of articulations with abnormal timing	Motoric	*Dynamic analysis required*	Any diacritic denoting timing such as lengthening marks
Retroflexion	Tongue tip retroflexion during any non-retroflex target	Unknown (reported in single case of CAS)		Any retroflex consonant
Coronal/dorsal production	Labial targets produced as coronal or dorsal consonants	Either: backing may be due to phonological disorder or speech motor constraint		Alveolar, post-alveolar, palatal or velar production
Non-ultrasound error (Non-UTI)	An error not visible using ultrasound	Either	*N/A*	Examples include voicing errors, lateralisation

Shaded rows are errors identified more often by the ultrasound-aided transcribers. These errors are likely to result from constraints in the speech motor control system and are more likely to be missed using audio only transcription.

Rating of Correct versus Incorrect Productions



Classification of Errors into Categories



Inter-Rater Reliability (Cohen's Kappa, 95% CI)

Transcriber Group	Rating of correctness		Error type classification	
	κ	Interpretation	κ	Interpretation
Ultrasound-aided transcribers	0.858	*Almost perfect*	0.704	*Substantial*
Audio-only transcribers	0.575	*Moderate*	0.473	*Moderate*

Key Findings

Ultrasound-aided transcription does not affect the overall number of errors identified in children's speech but it is **more reliable than traditional transcription** for deciding the type of error. **Errors due to constraints in speech motor control are more likely to be identified with ultrasound**

Acknowledgements

This study was funded by the UK Engineering and Physical Sciences Research Council (grant reference EP/P02338X/1). Thank you to the speech and language therapists who collected the data to Dr Zoe Roxburgh and Sarah White for completing transcriptions.

UKRI Engineering and Physical Sciences Research Council

References

1. Child Speech Disorder research Network. (2017). *Good Practice Guidelines for Transcription of Children's Speech Samples in Clinical Practice and Research*. Available from: https://www.nbt.nhs.uk/bristol-speech-language-therapy-research-unit/bsltruresearch/child-speech-disorder-research-0
2. Peterson-Falzone, S., Hardin-Jones, M., & Karnell, M. (2010). *Cleft Palate Speech* (4th Ed.). St. Louis: Mosby.
3. Cleland, J., Lloyd, S., Campbell, L., Crampin, L., Palo, J. P., Sugden, E., . . . Zharkova, N. (2019). The Impact of Real-Time Articulatory Information on Phonetic Transcription: Ultrasound-Aided Transcription in Cleft Lip and Palate Speech. *Folia Phoniatrica et Logopaedica*. doi:10.1159/000499753
4. Gibbon, F. E. (2004). Abnormal patterns of tongue-palate contact in the speech of individuals with cleft palate. *Clinical Linguistics & Phonetics, 18*(4-5), 285-311. doi:10.1080/02699200410001663362

Gradient Change in Lingual Gestures in Children with Speech Motor Delay

Joanne Cleland[1] and James M. Scobbie[2]

[1]University of Strathclyde, Glasgow and [2]Queen Margaret University, Edinburgh, UK

Email: joanne.cleland@strath.ac.uk   Twitter: @DrJoanneCleland

University of Strathclyde Humanities & Social Sciences

Background

- Children with persistent speech disorders have erroneous motor plans for specific speech gestures.
- One method for establishing new speech motor plans is Ultrasound Visual Biofeedback (U-VBF). This technique uses ultrasound to image tongue movements providing knowledge of performance.
- Effectiveness evidence is promising[1], however few studies investigate how the gestures are acquired and stabilised.
- This study looks at the acquisition of new motor plans in children with persistent velar fronting.
- Velar fronting is of interest in understanding Speech Motor Delay because children who fail to differentiate coronal and dorsal articulations may present with motoric deficits[2] despite a surface-level phonemic error.
- This study compares coronal/dorsal separation in children with persistent velar fronting to typically developing children.

Participants

Group	Speech Motor Delay with persistent velar fronting
n	4
Age	5 to 8

Group	Typically developing children
n	30
Age	5 to 13

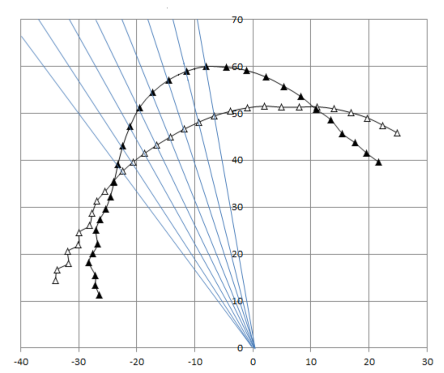
KEY FINDINGS

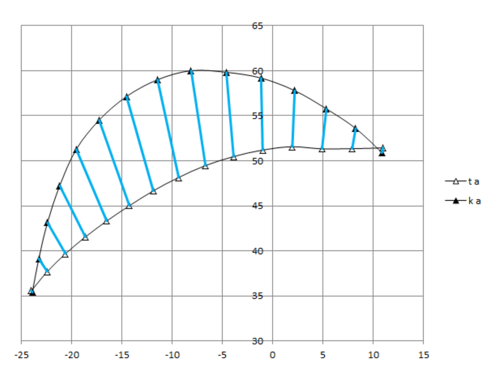
Children receiving motor-based ultrasound visual biofeedback therapy for persistent velar fronting show gradient change in tongue-shape during intervention in line with theories of motor-speech disorder

Typically Developing Children: Spatial Norms: k t Crescent

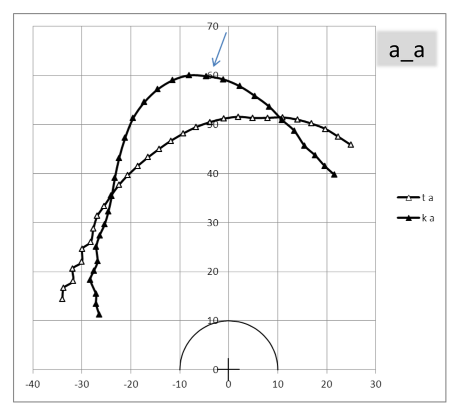
- 30 children (Ultrasuite Corpus[3]) aged 5;7 to 12;8
- 1 token of /k, t/ in 3 VCV contexts: /a, i, o/
- Synchronised audio and ultrasound data at 121.5 frames per second
- A fan-shaped grid (origin at probe centre) giving 42 radial sectors
- Stops annotated at burst
- Export of radial values (origin to surface)

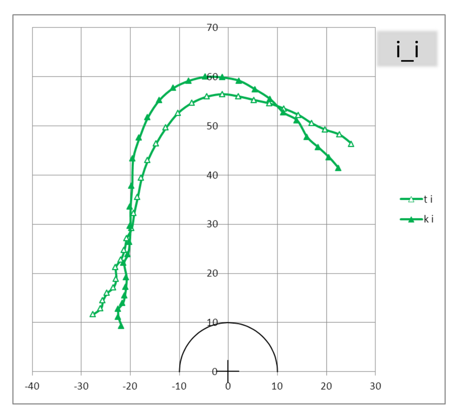
KT crescent depth = max radial /k/-/t/ (mm)

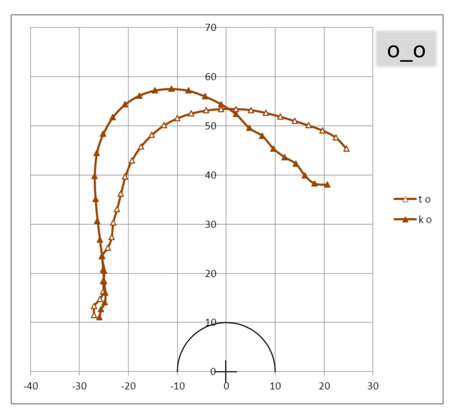


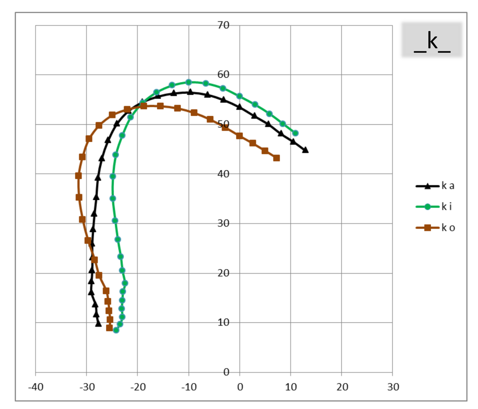


A contrastive measure that can quantify dorsal velar constriction spatially in relation to a same-speaker alveolar baseline. Two crossing points demarcate the dorsal gesture.





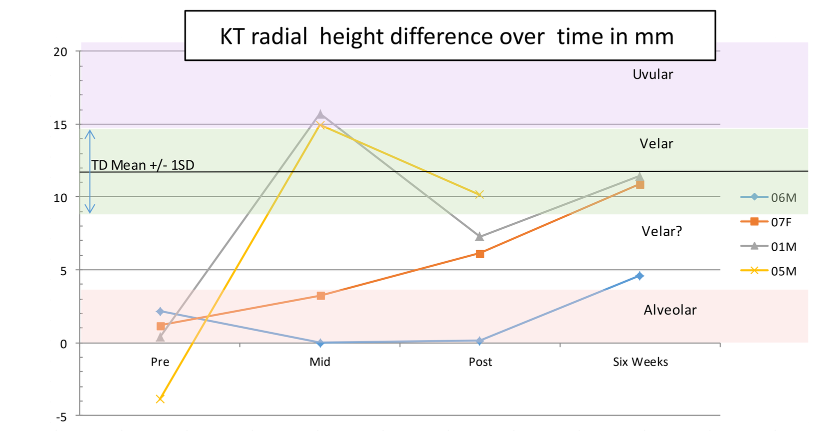




Max rad diff (mm)			
	a	i	o
Mean	11.9	7.5	12.1
SD	3.0	3.4	3.8
Min	7.3	3.3	5.9
Max	18.0	16.0	22.0

- Average tongue shapes (above) show clear co-articulatory effects
- Clear differentiation in tongue body in low vowel contexts
- Taking a_a as our comparison for SSD children, we expect a height difference of 11.9mm +/- 1SD as normal.
- Normal range = 8.9mm to 14.9mm

Children with SSD Tongue Shapes over Time

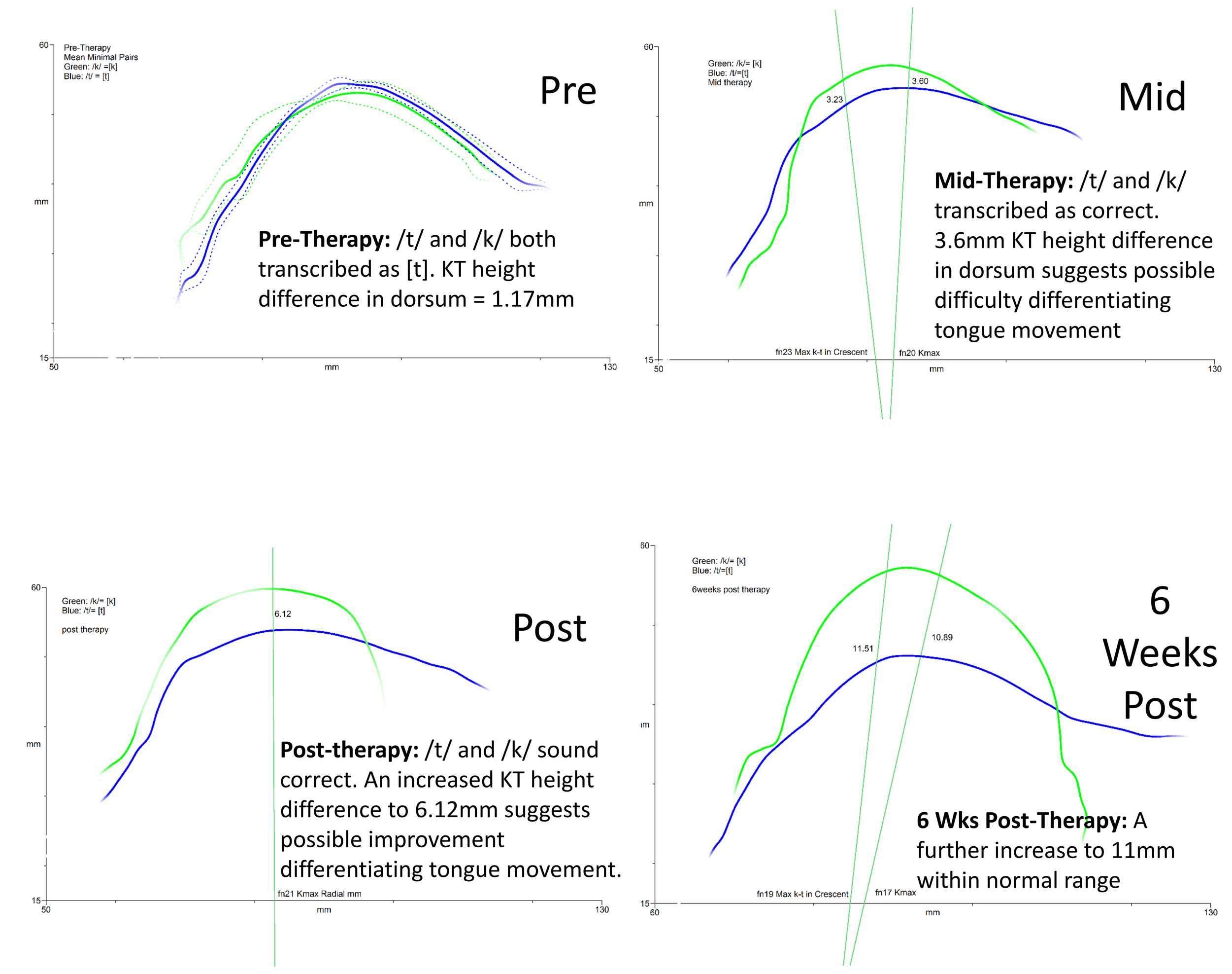


	KT Radial Height Difference (mm)			
Child	Pre	Mid	Post	Post 6 Weeks
06M	2.17	-0.01	0.15	4.62
07F	1.17	3.23	6.14	10.89
01M	0.38	15.72	7.29	11.44
05M	-3.89	14.93	10.16	

[k] transcriptions
[t] transcriptions
[q] transcriptions

- All children begin with negligible difference in tongue height between /t/ and /k/ at pre-therapy
- Two children (07F & 06M) show slowly increasing differentiation
- Two children (01M & 05M) show "overshoot" with abnormally large differences in tongue height at mid-therapy: transcribed as [q]
- All children move towards tongue-height differences in line with TD children

Example: 07F



Acknowledgements

This study was funded by the UK Engineering and Physical Sciences Research Council

UKRI Engineering and Physical Sciences Research Council

References

1. Sugden, E., Lloyd, S., Lam, J., & Cleland, J. (2019). Systematic review of ultrasound visual biofeedback in intervention for speech sound disorders. *International journal of language & communication disorders*, *54*(5), 705-728.
2. Gibbon, F. E. (1999). Undifferentiated lingual gestures in children with articulation/phonological disorders. *Journal of Speech, Language, and Hearing Research*, *42*(2), 382-397.
3. Eshky, A., Ribeiro, M. S., Cleland, J., Richmond, K., Roxburgh, Z., Scobbie, J., & Wrench, A. (2019). UltraSuite: a repository of ultrasound and acoustic data from child speech therapy sessions. *arXiv preprint arXiv:1907.00835*.