Quantifying covert contrast in children with Velar Fronting using Ultrasound Tongue Imaging

Charya Pallawala Kapurupastha Bandarage

School of Psychological Sciences and Health
University of Strathclyde

A thesis presented in fulfilment of the requirements for the degree of Master of Philosophy in Speech and Language Therapy in the college of Humanities & Social Sciences.

April 2023

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed: Charya Pallawala Kapurupastha Bandarage
Date: April 2023

Acknowledgements

I wish to thank my supervisor Dr. Joanne Cleland for her guidance and immense support given during my MPhil research. I also wish to thank my second supervisor Dr. Anja Kuschmann for her valuable input in guiding me in my thesis writing as well as her support during meetings. I am grateful to Dr. Ellie Sugden for her support and guidance given during the research phase of the project as well as the initial part of the thesis writing.

I want to thank my husband Hasanga and my son Heyan for giving me their time, strength and courage throughout this research. I want to give my sincere thanks to my father-in-law Rohan and mother-in-law Pushpa for giving me financial support for this course and their heartful blessings.

I want to thank my parents, Mahesh and Asoka who have been pillars of strength in supporting me throughout my life, and my sister Chathura and brother Hemnil for being there for me through the best and worst of times of my life.

I also wish to thank all my teachers in school and University lecturers in Sri Lanka who guided me thus far in my academics.

Abstract

The aim of the present study was to investigate whether school-aged children with SSDs of unknown origin and persistent Velar Fronting showed a covert contrast of a phonetic cause. The present study focused on analysing tongue shapes from UTI to assess motor skills in children with SSDs to determine whether those with persistent velar fronting have a phonetic cause for their errors. The study comprised of participants from a Control Group (CG) and a Study Group (SG), whose tongue shapes during production attempts of /k/ and /t/ were analysed.

We expected that evidence of a covert contrast between a child's /k/ and /t/ tongue shapes would suggest a phonetic cause, rather than a phonological cause for their SSDs. Therefore, perceptually neutralised tokens were compared using t-tests and a measure of spatial difference known as KTmax. The t-test criterion was $p \leq 0.001$ along a minimum of seven adjacent radii. The t-test function of AAA was used to test the distance from the origin of a measurement fan to the location of the tongue surface. The KTmax measure, which is also known as the maximum radial difference, indicates the highest degree of separation between tongue splines. Tongue shapes were also visually analysed.

Results showed that only one participant in SG (SG_4) demonstrated a covert contrast in non-lexical speech material data (repetitions to nonwords). This participant's KTmax measure fell within the range of typically developing children's. However, the same participant did not show a covert contrast in repetitions of single real words. The remaining perceptually persistent velar fronters (4/5) did not show a covert contrast either in non-lexical speech material data or in real word level.

The study concludes that some children with persistent velar fronting can show a covert contrast, suggesting phonetic difficulties and that such covert contrast can be measured using UTI.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

This thesis is an articulatory investigation of a common phonological process, known as velar fronting, in children with persistent Speech Sound Disorders (SSD) of unknown origin. SSD includes problems with articulatory output (deficits in phonetic structure due to poor motor skills) and phonological processes (failure to learn the pattern of sounds in a particular language) (ASHA, 2019). While the cause of some SSDs are known, causes for some SSDs are unknown. The estimated prevalence of persistent SSDs of unknown origin is 3.6% among children aged 8 to 9 years (Wren et al., 2016a).

Velar fronting is a rule-based pattern and hence is most often referred to in the literature as a phonological process. In fact, absence of velars in the phonetic inventory at 3 years of age has been claimed to be predictive of phonological disorder (Stoel-Gammon et al., 1996) Persistent Velar Fronting of unknown origin is defined as: children with the pattern of fronting velars who have passed the expected age for typically developing children's phonological process (e.g., 3 years), but have not shown progress using traditional auditory based intervention methods, and who do not have any genetic, structural or perceptual disorders (Wood and Scobbie, 2003). It has been debated in some literature whether persistent velar fronting of unknown origin is phonemic-based or articulatory-based. These studies were of both acoustic and instrumental nature.

Therefore, the present study mainly focused on tongue shape as a method for assessing motor skill in children with SSDs, to find out whether those who displayed persistent Velar Fronting had a motoric or a phonetic cause, by comparing metrics for quantification of children's tongue shape using ultrasound imaging (UTI).

Chapter 2 of this thesis reports on previous work on children's acquisition of speech,

speech perception and production before defining SSDs. It then introduces Velar Fronting to the reader and then goes on to report on earlier work investigating Covert Contrast. Four important studies pertaining to the present study are introduced and their outcomes are summarised next. Finally, this chapter presents the rationale for the present study, identifies the knowledge gap that it aims to address and declares the research aims, research questions and presents the hypothesis pertaining to each of them.

Chapter 3 presents to the reader the design of the research study and mentions briefly how the ethical consent was obtained and gives a brief introduction to ULTRAX2020, the project from which the secondary data for the present study was obtained. It then goes on to report the selection of participants and their categorisation into the Study Group (SG) and the Control Group (CG), the recording hardware setup of the UTI before describing the baseline measures used to validate the data collection. It concludes by introducing the UTI-based articulatory study and the data analysis methods.

Chapter 4 mainly presents a perceptual study's results and the results of the articulatory analysis including t-test, KTmax and direct visual inspection method. It concludes by presenting the tongue shapes of each participant of the SG and CG and summarises the findings.

In Chapter 5, the reader is presented a detailed discussion and analysis of the study's findings in relation to each Research Question presented at the end of Chapter 2 previously. It goes on to discuss its findings in light of relevant work in literature on covert contrast. This chapter then discusses the covert contrast found in one participant, namely SG_4, as a case study. It concludes by presenting the clinical implications of the present study's findings.

Chapter 6 which is the final chapter of this thesis, summarizes the present study's findings and identifies its limitations before presenting suggestions for future work to extend the present study.

Chapter 2

Literature Review

2.1 Children's acquisition of speech

It has been said that children acquire speech by mastering the perception and production of consonants, vowels, consonant clusters, tones, prosodic features, and phonological rules of their respective languages. According to McLeod and Crowe (2018), this leads to intelligible speech. The ability to correctly perceive/encode information in a speech signal, has been identified as an important requirement in children's speech acquisition (Klein et al., 2017) .

2.1.1 Speech perception

Speech perception is defined as the process by which the sounds of the language are heard, interpreted, and then understood (Manan et al., 2021). It employs two types of processing: active and passive speech perception. Active speech perception is concerned with modelling how humans generate lexical hypotheses from acoustic information which is alternatively called knowledge-based processing (Heald and Nusbaum, 2014). Active speech perception usually proceeds 'top-down', with syntactic and semantic knowledge.

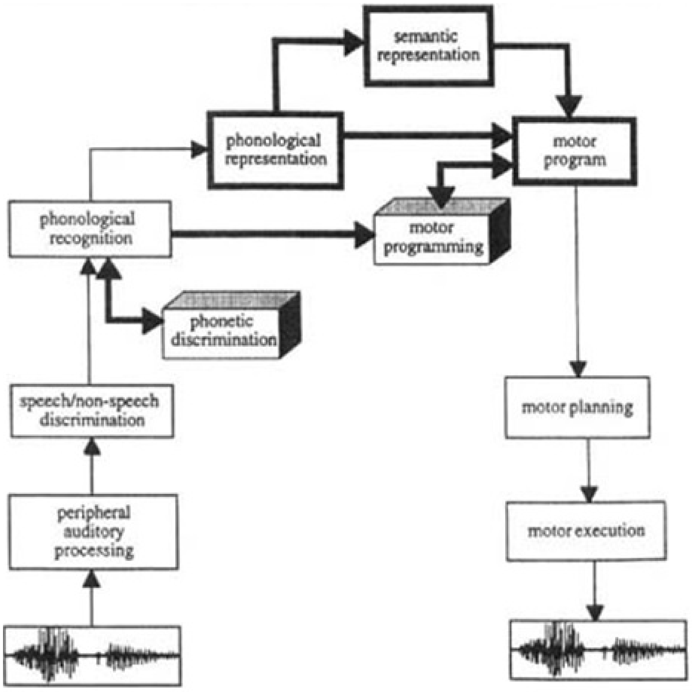
Passive speech perception on the other hand, is more related to automatised cognitive systems, having few demands on cognitive processing and proceeds 'bottom up'. Alternatively it is called data-based processing (Heald and Nusbaum, 2014). Similarly, Mackay (1956); Schneider and Chein (2003); Shiffrin and Schneider (1977) also claimed that a passive process of speech perception consist of an open loop sequence of transformations that are fixed, resulting in an invariant mapping of input to output, without active control of cognitive processes. In summary, active speech processing places demands on the system's limited cognitive resources to achieve cognitive and perceptual flexibility (Heald and Nusbaum, 2014).

According to Heald and Nusbaum (2014), active speech perception proceeds 'top-down' while passive speech perception proceeds 'bottom up'. Therefore, the active, and passive speech perception processes differ in the cognitive and in the perceptual demands they place on the system. The present study mainly used both the top-down process in using real words of t/k singleton minimal pairs, as well as the bottom-up process when using speech material data as imitations of non-words.

Speech repetition or speech imitation occurs when the speaker maps the sounds they hear directly from another person's speech to similar manners of articulation/similar places of origin in their own vocal tract. To repeat a token, first the listener encodes what they hear using peripheral auditory processing. The listener then discriminates this as speech/non speech. Following that, speech production, a motor plan that corresponds to it is retrieved, and are expressed using the correct articulators (see 2.1.3). This is depicted in the Psycholinguistic Speech Processing model developed by Stackhouse and Wells (1997) in Figure 1.

Figure 1

The Psycholinguistic Speech Processing Model by Stackhouse and Wells (1997).



In this model of passive speech perception, accessing the motor plan can follow two pathways: the phonetic pathway and phonological pathway. Both these pathways can be from passive speech perception without semantic processing. This is because, as mentioned by Gibson and Carmichael (1966), bottom-up processing focuses on interpreting sensory information in real-time, as they are not entirely schema driven.

2.1.2 Development of speech perception

Some researchers have suggested that infants follow a set of universal speech perception milestones (Curtin et al., 2017b). In many research studies, it has been stated that speech perception starts before birth (Liu and Kager, 2018; May et al., 2011; Ramus et al., 2000). This was further evidenced by May et al. (2011) who showed that newborn infants listen longer to speech sounds than to non-speech sounds, which implies that infants possess some of the raw psychoacoustic and cognitive capacities necessary to perceive speech at or just before birth. Overall, by 12 months, infants categorically perceive speech sound (e.g., distinguish 'tea' and 'key') segment units from the speech stream, learn about legal sound combinations for their language (phonotactics), rhythm and stress (prosodic properties) and track statistical properties of the speech input. Infants then use this knowledge to begin extracting and learning words (Curtin et al., 2017a).

Although speech perception develops early in life, acquisition of speech sounds starts at a few months after birth and continues into late childhood with the continuing development of articulatory-motor skills. However, there is very little information on when various articulatory features achieve adult-like maturity (Singh and Singh, 2008).

2.1.3 Speech production

According to Yalcinkaya et al. (2010) , children's development of speech production (expression) can have two aspects: phonetic acquisition and phonemic acquisition. The term 'phonetic' refers to speech sound production (i.e., articulatory or motor skills). The term 'phonemic' refers to speech sound use (i.e., functions, behaviour, or organization of the speech sound system). Articulatory acquisition of speech development relies on the phonetic act (motor act) of producing sounds, while phonemic acquisition is more related to cognitive-linguistic ability and language rules (Dodd et al., 2003a).

2.1.3.1 Phonetic speech sound production

According to Reinstein et al. (2010), the process of speech sound production occurs at the phonetic level. Phonetics is the study of speech sound and consists of articulatory phonetics, auditory phonetics, and acoustic phonetics. Articulatory phonetics is the production of speech sound. The motor act of articulation produces vowels and consonants necessary to create an inventory of all sounds needed to speak our languages (AllAboutLinguistics.com, nd).

After perceiving the acoustic representations via active or passive speech perception described in section 2.1.1, a listener needs a representation of the sound sequence (i.e., several commands which will be executed by our speech organs for the utterance). According to Trujillo (2002), it involves the following processes: Initiation, phonation, oro-nasal process, and articulation.

The listener needs a phonetic plan along with a motor plan which results in the physical production of sounds (Carmona et al., 1992). This is called initiation. The next step in speech sound production is phonation, an oro-nasal process. Speech is produced by the air stream coming from the lungs, which goes through the trachea and the oral and nasal cavities.

Then articulation occurs using active and passive articulators in the oral cavity. The active articulators are the tongue, the soft palate, the lower lip, the upper lip and the jaw. Of these, the main active articulator is the tongue, which is a key focus of the present study. The passive articulators are the maxilla, the teeth, the alveolar ridge, and the hard palate (AllAboutLinguistics.com, nd). Active articulators play an important role in producing speech sounds by providing a variety of articulatory motions to produce specific speech sounds (Aggarwal et al., 2021).

During articulation, an active speech organ (e.g., tongue) normally moves towards a passive articulator (e.g., alveolar ridge) and contacts it. This creates an obstruction that shapes and routes the airflow around it in a particular way (Knight, 2012, p. 36). This point where the maximum obstruction occurs in the vocal tract is called the 'place of articulation' (Anderson, 2022). The way the obstruction forms and releases or how close the active and passive articulators get to each other, to form the obstruction (e.g., stop, fricative, approximant), is called the "manner of articulation" or degree of constriction (Ashby, 2006, 2009).

For example, during the production of consonant /t/, the tip of the tongue touches the alveolar ridge. Thus, the alveolar ridge is the passive articulator, thereby making the place of articulation for /t/, 'alveolar'. In the production of consonant /k/, the back of the tongue (the dorsum) touches against the soft palate (velum), thereby making the place of articulation for /k/, 'velar'. The constriction degree or manner of articulation for /t/ and /k/ both are the same and they are stops or plosives. In a stop there is occlusion (i.e., blocking of the oral vocal tract resulting in no oral or nasal air flow). Therefore, the air flow stops completely.

The tongue is the major active articulator in all languages. This is because "the tongue can protrude forward, curl laterally as well as vertically, bunch at its centre or back or move its tip relatively independently of its body etc" (Knight, 2012, p. 27).

For example, velars which are consonants (e.g., /k/, /g/, /ŋ/) are articulated with the dorsum (i.e., back part of the tongue) making contact against the soft palate. On the other hand, alveolars such as /t/, /d/, /n/ are articulated when the tongue makes contact against or close to the superior alveolar ridge. The tongue plays an important role in both [k] and [t] productions, working as an active articulator. In the present study we mainly looked at tongue movements of alveolar and velar stops.

2.1.3.2 Phonetic speech production development

There are many examples in the literature which show that age effects contribute to acquisition of speech sounds. For example, MacNeilage et al. (2000) found that in the beginning children are constrained to mandibular movements without independent tongue control. As a result, labials become most frequently produced. At two years of age, due to children having stronger inter-lip coordination patterns, children are able to produce bilabial sounds (MacNeilage et al., 2000).

However, labiodental fricatives (e.g., /f/ and /v/) appear later in the development schedule as they need further differentiation to gain independent control of the functionally linked upper and lower lips (Green et al., 2000). Therefore, labiodentals occur between the ages of 2 and 3 years (Green et al., 2000; Stoel-Gammon and Dunn, 1985a).

Moreover, lingual coronal consonants emerge relatively late compared to velar consonants. This is probably because, since velar consonants are produced by the tongue dorsum, it requires fine adjustments by the tongue tip and blade to adapt to mandibular angles (Otomo and Stoel-Gammon, 1992; Wellman et al., 1931). In a similar manner, vowel productions in the first-year result from low, non-front, and non-rounded vowels. This may be due to the tongue barely elevating from the jaw (Buhr, 1980; Otomo and Stoel-Gammon, 1992). Front vowels, however, emerge later after children are able to produce refined tongue jaw movements. It has been stated that children develop finer control over tongue musculature with maturation and experience (Kent, 1992). Therefore, children begin to acquire rhotacized (retroflexed or bunched tongue) tongue movements after the age of four years (Namasivayam et al., 2020). This development of refined tongue movements at a later stage is not surprising given the fact that the tongue is considered a hydrostatic organ with distinct functional segments (e.g., tongue tip, tongue body) (Green and Wang, 2003; Noiray et al., 2013).

There is a relatively high number of reported studies demonstrating the order of acquisition of consonants, particularly in English. For example, a recent cross-sectional linguistic review by McLeod and Crowe (2018), which involved examining single word production combining data from 27 languages, identified that most of the world's consonants were acquired by children by the time they are 5;0 years. By this age, children produced at least 93% of consonants correctly (McLeod and Crowe, 2018). According to McLeod and Crowe (2018), children acquired plosives, nasals, and nonpulmonic consonants (e.g., clicks) much earlier than consonants involving trills, flaps, fricatives and affricates. Most labial, pharyngeal, and posterior lingual consonants were acquired earlier than consonants with anterior tongue placement. However, there was an interaction between place and manner where plosives and nasals produced with anterior tongue placement (plosives and nasals, e.g., /t, d, n/) were acquired earlier than anterior trills (e.g., /r/), fricatives (e.g., /s, z, ʃ, ʒ/), and affricates (e.g.,/ʧ,ʤ/) (McLeod and Crowe, 2018).

The review carried out by McLeod and Crowe (2018) on the topic of acquisition of English pulmonic consonants, which was based on 15 studies, found that the average age that children learn to pronounce English consonant phonemes were as follows:

- by 2 to 3 years: /p, b, m, d, n, h, t, k, g, w, ŋ, f, j/
- by 4 years: /l, ʤ, ʧ, s, v, ʃ, z/
- by 5 years: /r, ʒ, ð/
- by 6 years: /θ/

McLeod and Crowe (2018) also reported that velar /k/ and alveolar /t/ are accurately produced by 2 to 3-year old typically developing children. It is to be noted that, both [k] and [t] are plosives, although [k] is a posterior lingual consonant and [t] is a consonant with anterior tongue placement.

The above studies show that that there can be some differences in what is considered as the 'typical age' to master the production of different sounds by children. Therefore, an exact age for which speech sounds should be fully mastered, cannot be determined for certain (Yeh, 2011). Therefore, apart from age of acquisition, Speech and Language Therapists (SLTs) frequently use other constructs such as percentage-of-consonants-correct (PCC) and early-middle-late (EML), to describe children's speech acquisition of consonants during assessment and diagnosis in clinical settings, when they work with children (Bernthal et al., 2016; Crowe and McLeod, 2020; McLeod and Baker, 2016).

2.1.3.3 Phonemic/Phonological speech sound system

Phonological development refers to how young children learn to organise individual speech sounds into meanings and then form sound patterns into words, to form a language and to communicate clearly. Phonological development refers to the gradual acquisition of an adult-like system of speech sounds that are used to convey meaning in a language.

Many researchers have used phonological processes to describe speech production simplification patterns in typical children, to gain an insight into the phonological system (Donegan and Stampe, 2009; Grunwell, 1981; Smith et al., 1991; Stoel-Gammon et al., 1996). Phonological processes are defined as a set of "mental operations that change or omit phonological units as the result of the natural limitations and capacities of human vocal production and perception" (Hua, 2002, p. 28).

According to Smith (1989), the cause(s) for the majority of phonological processes are unclear. Some other reported work (Donegan and Stampe, 2009; Fey and Gandour, 1982; Macken et al., 1983; Vihman, 1978) attribute occurrence of most, but not all phonological processes to high level linguistic or cognitive factors, whereas others (Grunwell, 1986; Ohala, 1974; Smith, 1989) believe that low-level physiological and aerodynamic factors are responsible for many processes. The reason for causes for phonological processes being difficult to clarify was explained by Menn (2013) and Nittrouer and Studdert-Kennedy (1987) who state that, finding actual sources of most phonological processes is a difficult task since the speech 'output' which children produce, is the end-product of a variety of contributing factors. This is because linguistic, sensory, motoric, biomechanical and other sources are involved in the complex process of speech production (Smith et al., 1991).

Children in their early years show deviations in their speech compared to adult speech (i.e., error patterns). These error patterns are characterised by various phonological processes related to structural simplifications and systemic simplifications. Error patterns in English-speaking children's speech have been identified in the literature and they have been categorized at several levels (Dodd et al., 2003b). The first level of error patterns contain syllable error patterns (i.e., errors affecting the syllabic structure of the target words) or substitution error patterns (i.e., substituting one sound for another) (Bernthal and Bankson, 1997). According to Bernthal and Bankson (1997) the next level of error patterns (second level) contain syllable error patterns which are divided into eight sub-categories: deletion; weak syllable deletion, reduplication, consonant cluster reduction, assimilation, epenthesis, metatheses and coalescence (Stoel-Gammon and Dunn, 1985b). According to various work (Bernthal and Bankson, 1997; Stoel-Gammon and Dunn, 1985b), substitution error patterns can also be classified into eight sub-categories. They are velar fronting, backing, stopping, liquid gliding, affrication, deaffrication, vocalization and voicing.

It has been claimed that, by the age of four years most phonological processes disappear. Roberts et al. (1990) conducted a quasi-longitudinal study during which they assessed error patterns in the speech of 145 children aged 2;5 to 8;0 years, annually. In this study a standardized articulation test was used which showed that error patterns resolved rapidly between 2;5 and 4;0 years. Among children under four years error patterns of cluster reduction, deletion of final/medial consonants, liquid gliding, fronting, stopping and deaffrication were dominant. However, the study showed that only error patterns of cluster reduction, liquid gliding and deaffrication persisted beyond four years (Roberts et al., 1990).

Velar fronting is a substitution pattern error. During velar fronting, a velar consonant (i.e., a sound that is usually made while the back of the tongue makes contact with the soft palate), is replaced with a consonant that is produced at the front of the mouth. For example, /k/ can be substituted with [t], /g/ substituted by [d], and /ŋ/ is substituted with [n]. In other words, a child with a phonological pattern of velar fronting may be perceived to neutralise /k/ and /t/ sounds so that words such as ”tea” and ”key” sound the same: [ti] (Byun et al., 2016a). Grunwell (1981) states that in typically developing children, velar fronting is likely to be eliminated by 3,6 years. However, when these children pass this age, if certain phonological processes (3;6 years), and velar fronting or other processes still exist, then it can be classified as a Speech Sound Disorder (SSD) (Grunwell, 1981).

2.2 Speech Sound Disorders (SSD)

SSD is a term used to describe a range of difficulties in producing speech sounds in children (McLeod and Baker, 2016). McLeod et al. (2013) defines speech sound disorders (SSD) as:

> any combination of difficulties with perception, articulation/motor produc-

> tion, and/or phonological representation of speech segments (consonants and vowels), phonotactics (syllable and word shapes), and prosody (lexical and grammatical tones, rhythm, stress and intonation), which may impact on speech intelligibility and acceptability.(p. 1)

Since children master certain sounds and words at each age, a child diagnosed with SSD may have trouble saying certain sounds past this expected age. An articulation (phonetic) disorder is a type of SSD that affects the phonetic/motoric level, whereas a phonological disorder is a type of SSD that affects the phonological (phonemic) level (i.e., a cognitive or linguistic difficulty) (ASHA, 2019). Phonetic disorders occur when the child lacks the articulation or motor act to produce the target sound thereby resulting in an alteration of the place and at times the manner of articulation (Alighieri et al., 2020).

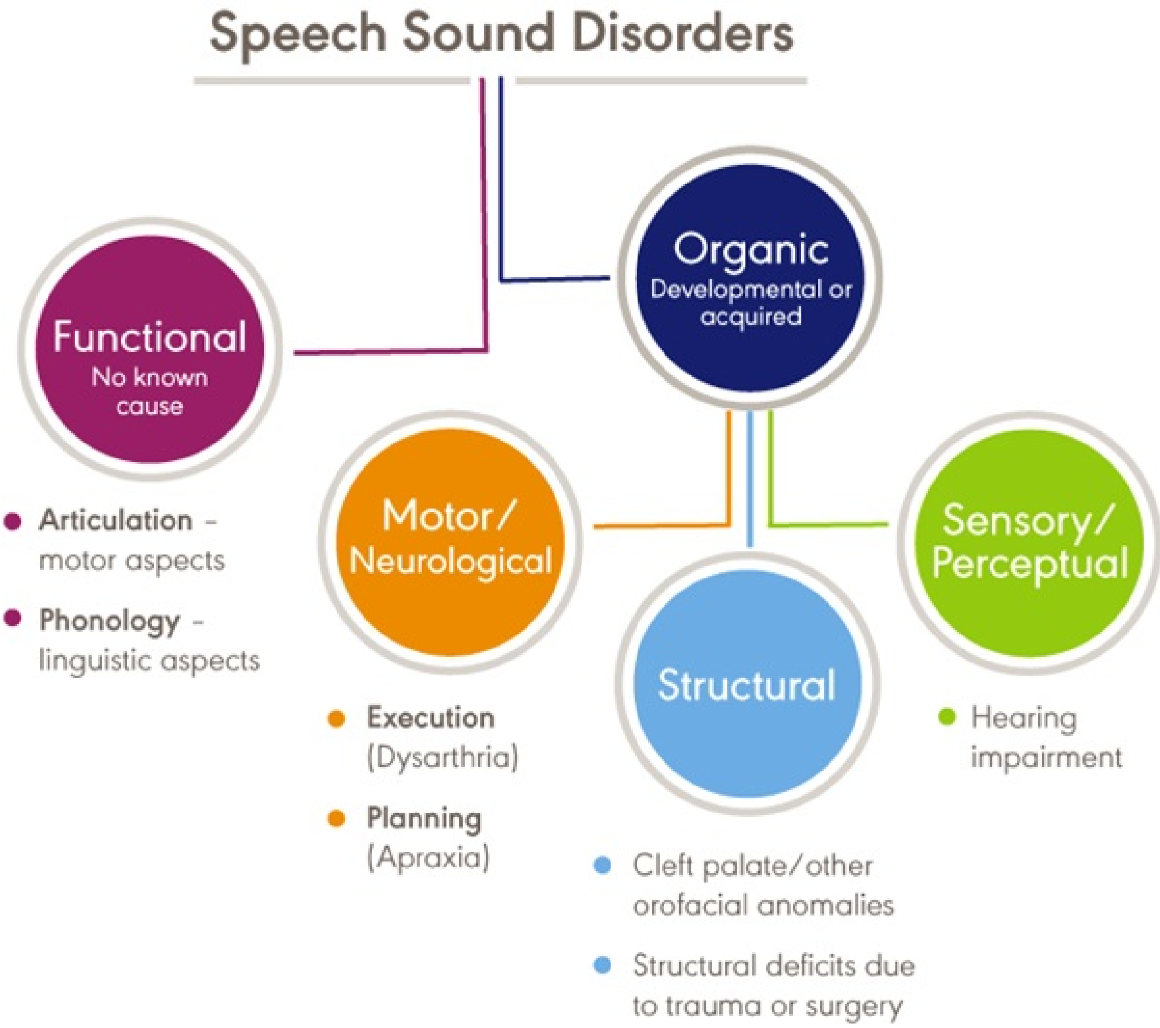
In some literature, phonetic (motoric) errors have been generally regarded as indicative of lower-level speech motor control difficulties while phonological (phonemic) errors have been interpreted as reflecting high-level cognitive/linguistic deficits (Bernthal et al., 2016; Grunwell, 1988). Therefore, phonetic level speech production difficulties happen due to articulatory difficulties or deficits with speech motor act. The phonemic difficulties are due to phonological representation of speech segments of consonants and vowels or cognitive -linguistic difficulty. Therefore, it is uncertain whether SSDs arise from motor-based articulation problems or language/grammar-based phonological problems. This has been well debated in past literature (Dodd, 2014; Shriberg et al., 2010; Terband et al., 2019).

Children with articulation disorders are described as having difficulties with the production of specific classes of speech sounds. Some have suggested that a child with an articulation impairment, although assumed to have correct phoneme selection, may be imprecise in the speech motor specification and implication of the sounds (McLeod and Baker, 2016; Preston et al., 2016). According to Shriberg et al. (2010) and McLeod and Baker (2016) the definition for SSD covers a range of difficulties in producing speech sounds in children. These can be due to limitations related to perceptual, speech motoric, linguistic processes or a combination of known (e.g., Down syndrome, cleft lip and palate) and unknown origins (Namasivayam et al., 2020). Therefore, the neutral term - SSD is currently used as a compromise (Shriberg et al., 2010).

According to ASHA (2019), SSDs with a known cause or Organic SSDs result from an underlying motor/neurological (e.g., childhood apraxia of speech (CAS) and dysarthria), structural/articulatory (e.g., cleft lip and palate) or sensory/perceptual cause (e.g., hearing impairment). Functional SSDs are called idiopathic (i.e., they have no known cause). This categorisation is visualised in Figure 2.

Figure 2

Categorisation of SSDs based on causes (ASHA, 2019).



According to McCormack et al. (2009) speech impairment of unknown origin is one of the most common communication impairments in childhood, among children who are without any genetic, structural or perceptual disorders. It has been reported that the prevalence of SSDs of unknown origin is estimated to be approximately 3.6% among children aged 8 to 9 years (Wren et al., 2016b). It is difficult to clearly distinguish between articulation disorders and phonological disorders of unknown origin, among such children (ASHA, 2019; Farquharson, 2015).

In the past, perceptual or auditory skill-based measures (e.g., traditional transcription methods) have mainly been used to identify and treat SSDs. Typical perception-based intervention approaches include minimal pairs, auditory discrimination, phonological awareness and core vocabulary therapy. Such intervention strategies mainly consist of clients listening to their own productions and modifying them using auditory cues (Broomfield and Dodd, 2011; Joffe and Pring, 2008; Law et al., 2003). It has been reported that some clients are unresponsive to these conventional intervention methods. Therefore, they were defined as persistent SSDs or intractable SSD (Wood and Scobbie, 2003).

This has important implications for the types of intervention that is chosen for a child. Therefore, SLTs use phonological and speech production measures to assist in the evaluation of SSDs. These measures contribute to the diagnosis, clinical intervention decisions and the monitoring of the development of the client's SSD treatment.

2.3 Persistent velar fronting

Cleland and Scobbie (2021) stated that "velar fronting is a rule-based pattern and hence is most often conceptualized in the literature as a phonological process" (p. 1). In fact, the absence of velars in the phonetic inventory at 3 years of age has been claimed to be predictive of phonological disorder Stoel-Gammon et al. (1996). Therefore, it has been stated that, velar fronting is normally successfully treated with phonological interventions using auditory or perceptual based measures (McLeod and Baker, 2016).

Wood and Scobbie (2003), define persistent velar fronting of unknown origin as: velar fronters who have passed the expected age for typically developing children's phonological process (e.g., 3 years), but have not shown progress using traditional auditory based intervention methods, and who do not have any genetic, structural or perceptual disorders.

Identifying covertness is considered as a very important factor in regard to remediating persistent velar fronting. Gibbon (1999) argued that velar fronting could be a result of a failure to dissociate different parts of the tongue. For example, raising the tongue dorsum while keeping the tip/blade low. Gibbon (1990) states that, the presence or absence of covertness can offer evidence regarding the level of processing at which a child's speech errors apply. Therefore, covertness can be interpreted as evidence of errors at a more peripheral or articulatory level. Identifying covertness is important diagnostically because it may provide evidence of motoric difficulties which, perhaps, may require different therapeutic approaches.

There have been few studies reported in the literature which attempted to identify the covertness, including covert contrast, covert errors and undifferentiated lingual gestures among persistent velar fronters as well as other persistent SSDs.

2.4 Covertness

Covertness refers to covert contrast, covert errors or undifferentiated lingual gestures. In the past, there have been few instrumental studies and acoustic studies undertaken to gather precise phonetic details to identify covertness. Methods such as Visual Analogue Scales (VAS) and articulatory analysis studies (e.g., Electropalatography (EPG) studies, ultrasound tongue imaging (UTI) studies and acoustic studies) were developed to aid visualizing such details not observed during perceptual studies. It has been said that, identifying covertness is diagnostically important because it may provide evidence of motoric difficulties which, may require different therapeutic approaches (Cleland et al., 2017).

Munson et al. (2010) has described covert contrast as the phenomenon when a child appears to neutralize a contrast in the adult language during speech, but on closer examination is found to produce a contrast. Covert contrast is a sub-phonemic difference that is usually not large enough to be noticed by a listener and be transcribed by a different phonetic symbol, but can be measured acoustically instead (Munson et al., 2010). According to Cleland and Scobbie (2021), "covert contrast is a term used to describe the phenomenon where, despite a perceptual neutralization, there is any measurable acoustic or articulatory difference between attempts at different phonemes" (p. 2). Covert contrast occurs when there exists a significant articulatory difference between two phonemes in a child's speech, but both the phonemes are transcribed with the same symbol by a listener (Smit et al., 1990).

Covert errors (i.e., a number of specific errors which defy broad phonetic transcriptions) are persisting speech errors such as, double articulations (n → n͡ɲ,n͡d),retroflexion (s → ʂ), distortions (r → ɹ̞). Therefore, covert errors are visible in articulatory data but sound 'correct' in perception analysis. For example, in Cleland et al. (2017), several 'correct' sounding instances of /k/, when investigated using articulatory analysis were found to be actually a retroflex or an undifferentiated lingual gesture.

According to Gibbon (1999), traditional transcription studies do not always give reliable results compared to acoustic or instrument based studies. Gibbon (1999) goes on to claim that, when standard transcriptions cannot detect undifferentiated lingual gestures (ULG), in some transcribed contexts these errors are categorized as speech errors (e.g., phonological substitutions, phonetic distortions) while in some other contexts these lingual errors are transcribed as correct productions. ULG typically occur during productions of lingual consonant targets and are characterized by contact that lacks clear differentiation between the tongue apex, tongue body, and lateral margins of the tongue (Gibbon, 1999). ULG are said to reflect a speech motor constraint resulting from either a delayed or lack of correct control over the function of independent regions of the tongue (Gibbon, 1999). According to Cleland and Scobbie (2021) ULG occur in children when they are unable to learn how to control independently the coronal and dorsal regions of their tongue. Furthermore, Gibbon (1999) and Mefferd (2017) have also suggested that ULG occur due to constraints in independent movements of the tongue and jaw, suggesting that ULG may arise from decreased oro-motor control abilities.

Covert contrast and covert errors are considered as a very important factor in determining whether to use an articulatory/motor approach or a phonological approach for the remediation of SSDs (Cleland et al., 2017). Gibbon (1999) states that the presence or absence of covert contrast can offer evidence regarding the level of processing at which a child's speech errors apply. For example, children who make a covert distinction between two targets often progress to an overt contrast even without any intervention (Forrest et al., 1990). This means that covert contrast is a positive prognostic indicator in child speech development (Forrest et al., 1994).

Covert contrast is one type of measure that is used to identify covertness. According to Cleland et al. (2017), identifying covertness is considered as a very important factor for remediating persistent velar fronting. Covert contrast among velar fronters can occur when there is a perceptual neutralization of a contrast, despite the existence of a clear statistical difference in tongue spline curves when measured. For example, some children who substitute [t] for /k/ show a neutralization between the two phonemes in phonetic transcriptions. However, when measured instrumentally (e.g., using UTI), there exists a significant statistical difference between productions of target /t/ and target /k/. Therefore, the present study aims to identify whether children with persistent velar fronting have a phonetic or a phonemic cause, by investigating any covertness that exists.

2.4.1 Methods used to identify covertness among persistent velar fronters

A smaller number of studies have been conducted to identify covertness among persistent velar fronters to gather phonetic details. They have been developed to aid in visualizing such details not observed in perceptual studies. These studies are: VAS based study by Munson et al. (2012) (see section 2.4.1.3), an instrumental study involving EPG by Gibbon (1999) (see section 2.4.1.3.1) and UTI based studies by Byun (2012), Cleland et al. (2015b), Cleland et al. (2017) and Cleland and Scobbie (2021) (see section 2.5).

2.4.1.1 Traditional perception-based transcription

Most SLTs in speech therapy settings use traditional, perception-based transcription methods to diagnose SSDs. However, it is well known that transcription-based studies do not always give reliable results compared to acoustic and instrumental based studies (Pandeli et al., 1997). For instance, Gibbon (1999) claims standard transcriptions cannot detect ULGs. In some transcribed contexts, these are categorized as speech errors (e.g., phonological substitutions, phonetic distortions). In some other contexts, these lingual errors are transcribed as correct productions (Gibbon, 1999). In another example, Cleland et al. (2020) carried out a study to investigate whether transcription studies, supplemented with an additional modality (i.e. using UTI), can improve the detection of covert errors with children with cleft lip and palate. The study results revealed that both UTI and traditional transcription produced similar error detection rates. However, these were significantly higher than the observations recorded live in the clinic when using UTI. Therefore, the authors concluded that by using UTI aided transcription, the clinicians were more likely to identify covert errors such as, double articulations and retroflexion, than the traditional phonetic transcription method (Cleland et al., 2020).

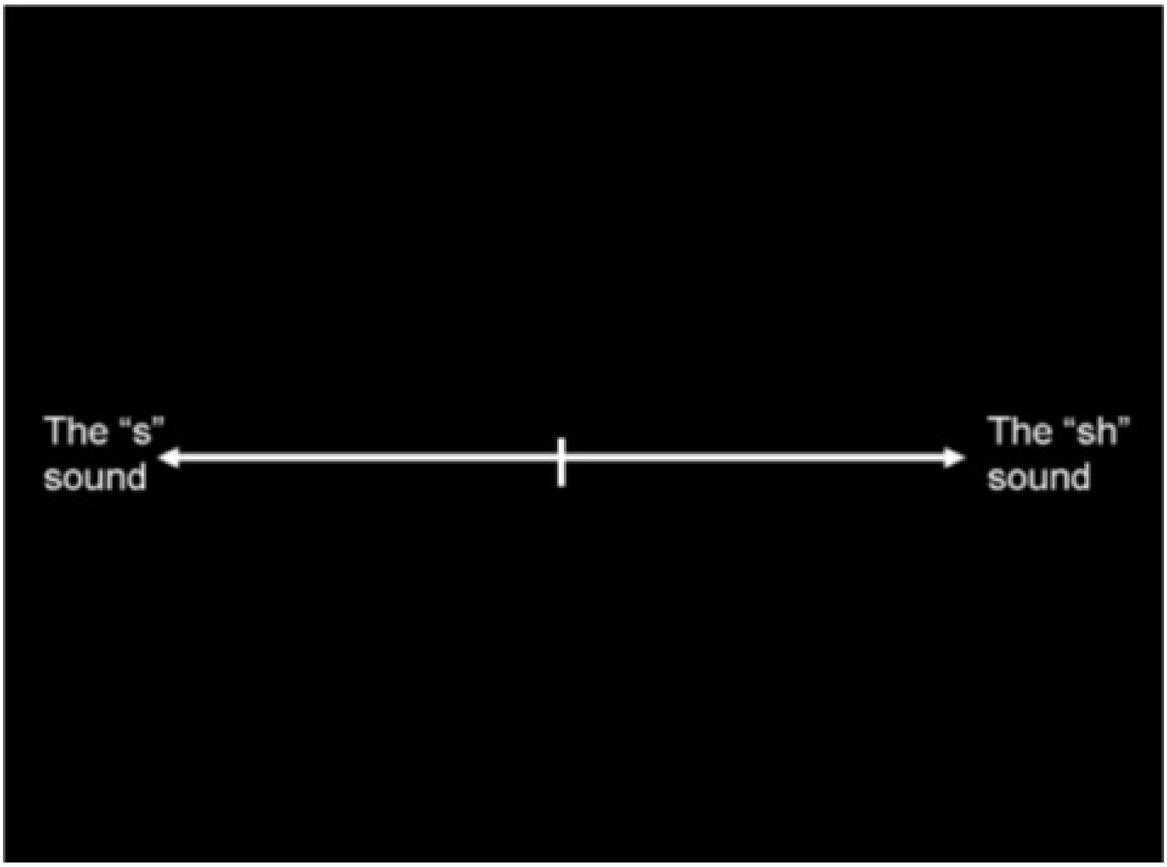
2.4.1.2 Visual Analogue Scale (VAS)

In a study conducted by Munson et al. (2012) the Visual Analogue Scale (VAS) method was used to detect small changes which indicate covertness, during therapy which would otherwise not be detected using phonetic transcription alone. By using this VAS method, Munson et al. (2012) examined the listeners' perception of productions of minimal contrasts such as /s/–/ʃ/, /s/–/θ/, /t/–/k/, and /d/–/g/. The authors chose these contrasts for the study because they are typically acquired later by English learning children. The participants were children aged 2 to 5 years old (Munson et al., 2012).

The VAS used in the study by Munson et al. (2012) consisted of a double-headed arrow which is graphically designed on paper. An example of VAS used in the study is shown in Figure 3.

Figure 3

The Visual Analogue Scale (VAS) by Munson et al. (2012).



One end of the arrow is marked as 's', 's', 't', 'd' sound while the other end is marked as 'ʃ', 'θ', 'd', 'g' sound (Munson et al., 2012). Listeners who participated in the study were not specially trained in speech and language therapy. The listeners were required to listen to the children's productions of consonant-vowel sequences, which were exercised from productions of real words and non-words (Munson et al., 2012).

The listeners were then asked to rate the initial consonant using the VAS. They were told to click on the line on the VAS the point at which they believed the child's production would fall in the scale. The listeners were encouraged to use the entire line in the VAS (Munson et al., 2012). The listeners were not given any specific instructions on their assessment but were allowed to apply their own criteria on how close a production was to an ideal /s/, /ʃ/, /θ/, /k/, /t/, /d/, or /g/. The authors of this study concluded that SLTs who use VAS to measure children's progress in therapy

”might also be more likely to detect small changes during therapy, than would those who use phonetic transcription alone” (Munson et al., 2012, p. 10).

2.4.1.3 Acoustic studies

The most widely used method for identifying covertness has been acoustic analysis. Both temporal (i.e., time-based) and spectral measures (frequency-based) derived from acoustic recordings have revealed the presence of different types of covert contrasts. Voice onset time (VOT) is the most frequently reported temporal metric used in investigate velar fronting (Tyler and Saxman, 1991; Young and Gilbert, 1988). VOT measures the time interval between consonant onset and the onset of low-frequency periodicity which is generated by rhythmic vocal cord vibrations (Steinschneider et al., 1999). Another type of acoustic measure commonly used to identify covert contrast in cases of velar-to-alveolar fronting is, spectral moments (Forrest et al., 1990; Li et al., 2009).

Overall, acoustic studies identify articulatory features that relate to vocal tract shape and movement, signalling place and manner distinctions. Priestly (1980) however, pointed out that acoustic analysis has limited use in cases where an articulatory distinction exists, but an acoustic difference is lacking. Lawson et al. (2008) also stated that although useful, auditory and acoustic analysis may suggest what happens in the vocal tract. A direct articulatory study is thus preferable. Furthermore, acoustic analyses do not provide direct information about the tongue’s behaviour.

A study by Byun et al. (2016a) however combined both acoustic and UTI measures to detect covertness among children with velar fronting (see section 2.5.4). The acoustic measures in this study were VOT and spectral moments of the burst. By combining results from both acoustic and ultrasound methods, the study discovered that, one participant out of two persistent velar fronters demonstrated a covert contrast (Byun et al., 2016a).

Due to the various limitations in studies involving acoustic analysis, other physiological techniques such as EPG (see section 2.4.1.3.1) have been used in the past (Gibbon, 1990; Gibbon and Crampin, 2001; Gibbon et al., 1995; Gibbon and Lee, 2017). In more recent times, UTI based studies have been conducted as well (Byun, 2012; Cleland and Scobbie, 2021; Cleland et al., 2017, 2015b). Both methods have proven to be valuable in uncovering covert contrasts and covert errors, since these two physiological techniques provide direct information about tongue behaviour and can reveal details of articulatory activity such as tongue's position and its movement during events such as the closure period of stops during speech. These advantages over acoustic analysis have been a strong motivation for researchers to use them when studying covert contrasts.

2.4.1.3.1 Electropalatography (EPG)

EPG is an instrumental articulatory technique which visualises patterns of tongue-palate contact. During this method, the contact between the tongue and the palate is measured during speech production in real-time, using pseudo-palates equipped with special sensors (Mat Zin et al., 2021). Such tongue-palate contact information provides rich detail and information about articulatory parameters such as the place of articulation, lateral bracing, groove formation, timing of tongue movements and co-articulation (Gibbon and Lee, 2017).

When using EPG, speakers wear a custom-made artificial palate which is approximately 1 to 3 mm thick and contains 62 sensing electrodes. The first row of the palates contains six electrodes which are spaced closely together, next to the upper incisors. The next seven rows (i.e., rows two to eight) contain eight sensing electrodes which are more widely spread at the junction of the hard and soft palate. When the tongue contacts an electrode on the palate, a signal is sent to the processor and the pattern of contacts is schematically displayed on a computer screen.

Gibbon (1999) conducted an EPG-based study involving 12 children who were persistent velar fronters of unknown origin. ULGs were detected using EPG instrumental data when a large amount of tongue-palate contact occurred during attempts at velar and/or alveolar stops. In some cases, contact extended all the way from the alveolar region to the velar region (Gibbon, 1999). Therefore, Gibbon (1999) argued that velar fronting could be a result of a failure to dissociate different parts of the tongue such as, when raising the tongue dorsum while keeping the tip/blade low.

According to Gibbon and Lee (2017) the EPG technique can be used to identify covert contrasts in lingual sounds among participants with SSDs, by measuring the amount of contact for most lingual sound targets. For example, when studying velar fronters with SSDs who frequently produce errors, the use of EPG is particularly useful in investigating cases of potential covert contrast. When using EPG to identify the possible presence of covert contrasts, the first step is to carry out a transcription-based assessment of the child's speech to highlight contrasts that are lingual and therefore likely to have measurable amounts of tongue-palate contact. Gibbon and Lee (2017) also claimed that EPG studies can be used to detect covert contrasts, because EPG allows the subjective visual inspection of the raw data and then objective quantification using numerical indices. According to Gibbon and Lee (2017), a number of metrics/indices derived from EPG tongue-palate contact raw data allow investigating aspects such as: place of articulation, amount of contact, symmetry of contact and variability.

Gibbon and Lee (2017) have also stated that EPG can detect covert contrast and covert errors manifesting in a variety of phonetic and articulatory problems among velar fronters by revealing visual and quantifiable features. Zharkova (2013) however stated that, "most articulatory information to date has been collected using EPG, which records the location and size of tongue-palate contact but not the tongue shape. The latter type of data can be provided by ultrasound studies" (p. 1). For example, ultrasound tongue imaging can show the tongue shape in real time.

It has been said that when studying velar fronters, EPG is not suitable for all speakers because the EPG plate does not capture the velar closure for /k/, since plates typically end at the juncture of the hard and soft palates (Cleland and Preston, 2020). Therefore, UTI could be a more useful approach.

2.4.1.3.2 Ultrasound Tongue Imaging (UTI)

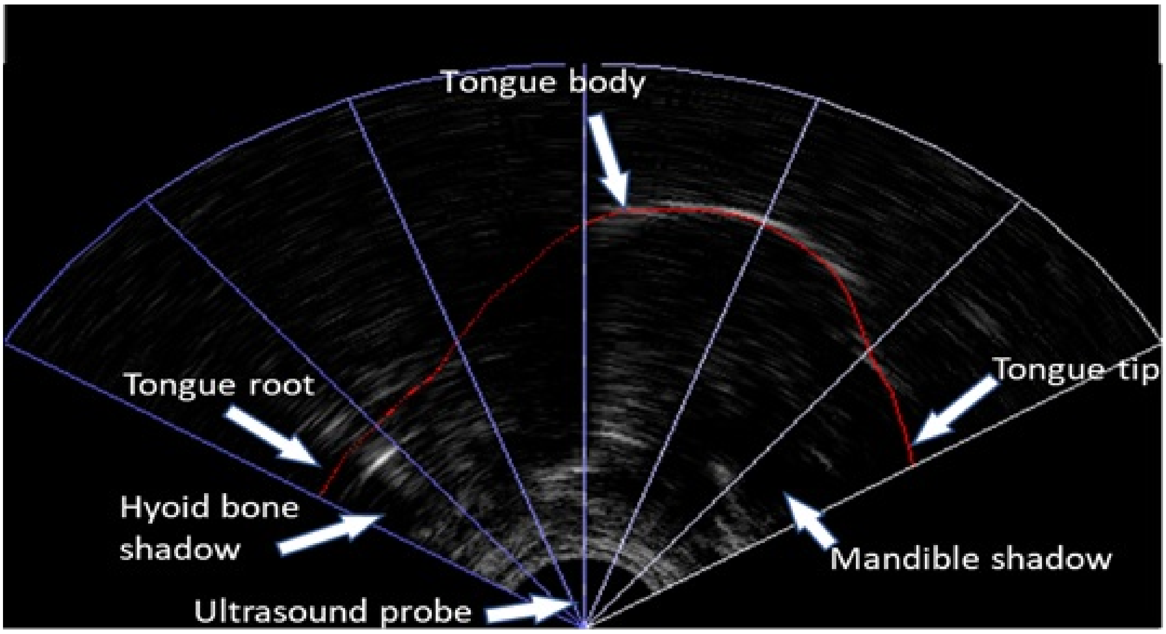
UTI is an articulatory technique which shows movement of the speaker's tongue in real-time allowing direct measurement of tongue-movements using a medical ultrasound probe placed under the chin. Therefore, it allows real time visualization of the surface of the tongue in the mid-sagittal view from near the tip to the root. This is achieved using the reflective properties of ultrasound waves when they travel through soft tissue. UTI first came into regular clinical use in the 1960s and 1970s and has been used for research on speech production and linguistics research since then Gick (2002). It is a non-invasive technique allowing researchers to view the shape, position and movements of the tongue (from root to apex) in real time during speech (Preston et al., 2017).

As described by Kabakoff et al. (2023), when these waves encounter a medium having a different density to the one it was travelling through (e.g., air or bone) the waves reflect at this barrier between the mediums. The reflected waves are sensed by a probe which converts it into an electrical signal. The probe is synced with a processor in the imaging system which uses this signal to create an image on screen. The tongue's surface appears as a white line on screen. As sound waves do not pass-through bone, they appear as a black shadow on screen (Kabakoff et al., 2023).

Typically, UTI is used as a 2D image. However, in some cases 3D images were used to analyse tongue movements (Burton et al., 2013; Lulich et al., 2018). Ultrasound can be used to image either a midsagittal or coronal image of the tongue. As shown in Figure 4, using a midsagittal view, we can observe the tongue's surface and movement patterns from the tongue tip to the root.

Figure 4

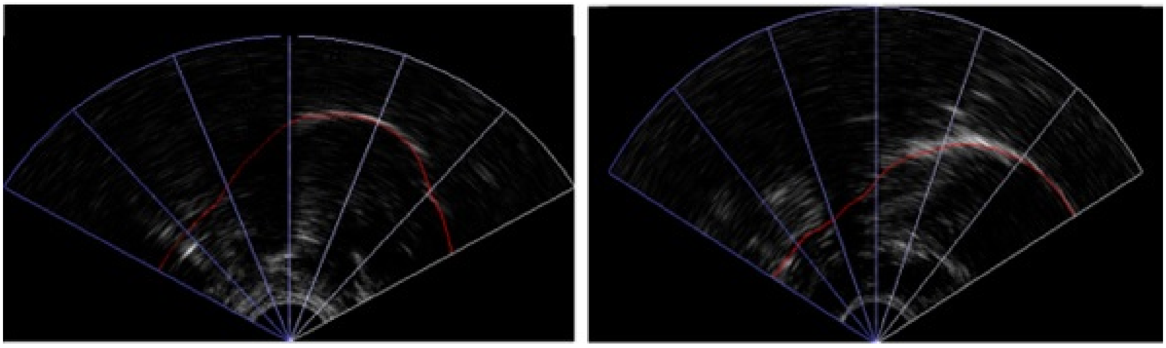
Mid sagittal view of Ultrasound image taken from at the maximum point of constriction of /k/. Tongue tip is to the Right.



An example of a midsagittal view of the UTI from the present study, which involved mainly looking at tongue movements of alveolar and velar stops, is shown in Figure 5. The [k] tongue spline is shaped like a dome, characterised by its raised tongue dorsum. The [t] tongue shape is characterised by a raised tongue tip and a usually flat tongue body. The maximum constriction for velar [k] in Figure 5 occurs in middle third posterior part of the tongue or the dorsum of the tongue.

Figure 5

Tongue spline fitting for [k] (Left) and [t] (Right) at the maximum point of constriction.



For [t], the maximum constriction occurs at the anterior part of the tongue. In the UTI image, only the tongue and chin fat are visible (Seeingspeech, nd). It has been noted that, near 100% reflection at the tongue-air-boundary often provides a bright tongue surface image. Therefore, the tongue's contour region can be seen in UTI.

Since no wave energy is transmitted into the air above the tongue or to its sides, the hard palate and sides of the tongue are usually not visible in UTI (Seeingspeech, nd). As there are no hard structure references on the image, it is challenging to locate the tongue position in the sides in UTI (Hamed Mozaffari and Lee, 2020). Therefore, when producing /t/, although the tongue's tip touches the sides, it is not visible. In addition, UTI does not normally capture the palate since the air at the tongue surface reflects the ultrasound beam back to the transducer (Epstein and Stone, 2005).

Using UTI is less costly than EPG. This is because UTI does not require custom made palates unlike for EPG (Cleland and Scobbie, 2021). Another advantage of using UTI is that vowels, and the coarticulatory effects of vowels on consonants, can be easily viewed.

According to Sugden et al. (2019) UTI is becoming an increasingly popular technique in phonetics assessment in SLT clinics for measuring tongue shape and movement because the hardware setup for UTI is becoming increasingly more affordable. UTI coupled with visual biofeedback has been used in clinical settings for both diagnosis and therapy (Cleland and Scobbie, 2021; ClinicalTrials.gov, 2018; Roxburgh et al., 2015; Sugden et al., 2019).

Since UTI can be applied to larger groups of children than in EPG in clinical settings providing UTI, it would be useful to look at the various norms used in UTI based studies. They are discussed next.

2.4.1.4 Measures used in UTI to identify covert contrast

UTI studies reported in the literature have used quantitative measures such as: t-test (Cleland et al., 2017), KTmax (Cleland and Scobbie, 2021), DEI (Byun and Tessier, 2016) as well as qualitative measures such as: direct visual inspection of tongue splines (Cleland et al., 2017), in order to identify covertness among the persistent velar fronters.

2.4.1.4.1 t-test

A t-test is a statistical measure calculated to determine whether there is a significant difference between the means of two groups. The t-test had been used in some UTI studies to ascertain whether there is a statistical difference between alveolar /t/ and velar /k/ tongue spline shapes among persistent velar fronters Cleland et al. (2017) and Cleland et al. (2015b) as well as among children with submucous cleft palate (Roxburgh et al., 2022). The study by Cleland et al. (2015b) and Cleland et al. (2017) reported using t-tests along with the qualitative measure of direct visual interpretation of tongue splines in their UTI-based study.

In UTI studies a t-test can be computed using the in-built function in AAA software (AAA, 2012). According to Cleland et al. (2015b) this type of t-test is computed radially along 42 fan-lines emanating from the probe. The threshold used in this study for determining a significant difference between the alveolar and velar tongue splines was a Mean difference of $p \leq 0.05$ along a minimum of seven adjacent radii. This was indicative of a covert contrast (Cleland et al., 2015b). Other studies such as Cleland et al. (2017) and Cleland and Scobbie (2021) also have reported using t-test to identify covert contrast in persistent velar fronters of unknown origin. With a greater difference between /k/ and /t/ tongue splines corresponding to a lower p-value. According to Roxburgh et al. (2022), "non-significant t-tests do not disprove the null hypothesis of complete neutralisation, whereas significant t-test results indicate a covert contrast" (p. 4).

The present study uses the same null hypothesis as the one used in the study by

Cleland et al. (2017). That is, a covert contrast is indicated by lack of statistically significant difference between tongue shapes for /k/ and /t/ attempts at the point of maximum constriction.

A p-value in statistics is used to determine the significance of data or to ascertain if the observed relationship happened as a result of chance. If statistical results show a lower p-value, it gives strong evidence that the tested outcome has a very low probability of having a result by chance (SimpliLearn, 2022). A lower p-value in a t-test shows there is a statistically significant difference between the Means of two groups. The level of statistical significance is usually represented as a p-value between 0 and 1. A greater difference between two groups corresponds to a lower p-value. Therefore, the smaller the p-value the more likely the test will reject results that have happened due to chance (e.g., null hypothesis) and suggest there is a significant difference (e.g., alternative hypothesis). A p-value less than or equal 0.05 ($p \leq 0.05$) means that there is a 5% probability that results are due to random chance. Although p-values of 0.05 and 0.001 both are considered statistically significant, the present study used $p \leq 0.05$ criterion to test the hypothesis. Using $p \leq 0.05$ criterion means there is less or equal to one in thousand probability of the result occurring due to chance. Therefore, the credibility of the test's outcome is high. The present study uses $p \leq 0.001$ to test the null hypothesis (i.e., strong evidence /k/ and /t/ tongue splines have no statistically significant difference). Using the outcome of such a t-test we can reject the null hypothesis and accept the alternative hypothesis that /k/ and /t/ tongue splines shows a significant difference.

2.4.1.4.2 Dorsum Excursion Index (DEI)

DEI is another quantitative measure used in UTI. Zharkova (2013) presented the DEI for use in clinical analysis of abnormal lingual stops produced by children with a cleft palate. DEI measures the magnitude of excursion of the tongue's dorsum, relative to the front-to-back extension of the tongue (Cleland and Scobbie, 2021). A higher DEI value indicates more dorsum excursion. Therefore, a typical articulation of /k/ is expected to have a higher DEI than a typical articulation of /t/ (Cleland and Scobbie, 2021).

Byun et al. (2016a) used DEI to identify covert contrasts among persistent velar fronters of unknown origin. According to the study's results, two participants who were typically developing children, showed a significant higher DEI in velar target contexts than in alveolar target contexts (Byun et al., 2016a). It was stated that their DEI values were roughly consistent with the DEI values that Zharkova (2013) described as typical for velar (DEI = 0.5) and alveolar stops (DEI = 0.26). In the study by Byun et al. (2016a), one participant who presented positional persistent velar fronting (syllable initial position) showed a covert contrast in DEI. In this participant, the velar targets were produced with a substantially higher DEI than that for alveolar targets. This suggests the existence of a covert contrast. However, it was reported that, in the back vowel context, there was no apparent difference in DEI (Byun et al., 2016a).

2.4.1.4.3 KTmax

KTmax is another quantitative metric that has been used to measure a covert contrast in UTI. Cleland and Scobbie (2021) recently introduced two new measures: KTmax and 'KT crescent area' in their UTI-based study. The authors defined KTmax as the single maximum radial difference (measured in mm), within the crescent shape formed by the overlap between a typical /k/ tongue spline and a /t/ tongue spline (i.e., KT crescent).

KTmax can be used to measure the degree of differentiation between stops in the dorsal region (Cleland and Scobbie, 2021). In this study, the authors used KTmax and KT crescent area to distinguish /k/ and /t/ by quantifying the magnitude of this distinction in spatial terms (i.e., linear dorsal difference in mm). To determine this, they calculated the area of the crescent formed when one tongue spline crosses above the other (i.e., KT crescent area) and the maximum radial difference (i.e., KTmax) between the /k/ and /t/ tongue splines in this region. The ultrasound probe's location was used as the origin in this measurement (Cleland and Scobbie, 2021). The authors reported that, the width and the area of the contrast between /t/ and /k/ are highly correlated. This means that, either KTmax or KT crescent area can be used as metrics to compare the typical velar vs. alveolar difference (see section 2.5.3).

2.4.1.5 Direct inspection of tongue splines

Although quantitative measures are crucial in the analysis of ultrasound data, especially in a clinical population, directly inspecting and identifying an abnormal tongue shape is helpful as well in identifying any covert errors or undifferentiated lingual gestures (Roxburgh et al., 2015). Cleland et al. (2017) discovered covert errors when one participant's ultrasound data was examined visually, in more detail. This was in addition to using the quantitative measures described previously. The method of direct visual inspection of tongue splines was also used in other recent studies, to identify covert contrasts among persistent velar fronters (Cleland and Scobbie, 2021; Cleland et al., 2015b). These studies also combined the statistical t-test method ($p \leq 0.001$) along with visual inspection of tongue splines when investigating the existence of a covert contrast.

Cleland et al. (2017) reported that velar fronters who substituted [t] for /k/ showed overlapping tongue splines, which is indicative of having no covert contrast, because there is no difference between the visual data and the perceptual data. On the other hand, if tongue spline shapes of /t/ and /k/ were different (i.e., they did not overlap), it suggested there can be a covert contrast. This is because, despite the perceptual neutralisation, there was a visual difference in tongue splines between those two phonemes. Overall, this was further confirmed by t-test showing statistically significant difference (p-value ≤ 0.001) and having measurable KTmax measure.

2.5 UTI-based studies on persistent velar fronting

2.5.1 The study by Cleland et al. (2015b)

The study by Cleland et al. (2015b) involved six participants who were monolingual English speakers who displayed persistent primary SSDs. They were children aged 6;0 to 10;1 and were unresponsive to traditional auditory-based/linguistic-based speech therapy intervention. Every participant underwent two baseline probe sessions during weeks 1 and weeks 6. There was no contact between these two sessions. Then, intervention using UVBF occurred over 12 weekly sessions. There was one more mid-therapy probe at week 12. Two post-therapy maintenance were then conducted 6 weeks apart, at week 19 and week 25 respectively, with no contact in between (Cleland et al., 2015b).

Before intervention, among the participants all attempts of /k/ were transcribed as [t] in their broad phonetic transcriptions. According to Cleland et al. (2015b), in their UTI analysis no statistically significant difference was shown between /k/ and /t/ using the criterion $p \leq 0.001$. This showed that both /k/ and /t/ tongue splines overlapped.

After the intervention using UTI, all participants were shown to have made significant progress as evidenced by the perceptual measures and UTI measures (Cleland et al., 2015b). Change in their tongue-shapes were evident as they showed non-overlapping tongue splines and significant difference in their statistical t-tests. The participants' KTmax measure between /t/ and /k/ average tongue splines showed maximum differences between 6.5 mm to 11.0 mm (Cleland et al., 2015b).

Therefore, Cleland et al. (2015b) concluded that these children presented articulatory difficulties because they showed improvements slowly and progressively with UVBF intervention. This observation is in line with by Weaver (2015) who claimed that, if children had a motoric difficulty in producing velar gestures, they were expected to show a gradient or variable acquisition of the new gesture, which is consistent with the view that learning a new motor gesture is slow and inconsistent process.

2.5.2 The study by Cleland et al. (2017)

Cleland et al. (2017) conducted a UTI-based study to compare tongue contours for /k/ and /t/ in seven children aged 6;0 to 10;11 years, who presented persistent velar fronting. The study aimed to identify covertness at the articulatory level using a method of overlaying tongue contours. All the participants presented persistent primary SSDs and highly consistent velar fronting in all word positions according to the transcriptions. Participants were selected for study if the percentage of correct velars at baseline probes were less than 20% and according to clinical referral. All participants had to complete a word list which was designed to sample velar consonants and /t/ and /k/ minimal pairs in variety of word positions and vowel environments. Speech measures were recorded simultaneously using ultrasound audio and video (Cleland et al., 2017). Therefore, the following two types of analysis were conducted by the authors:

- Broad phonetic transcriptions.
- Ultrasound Analysis.

Broad transcriptions results showed that six out of seven (6/7) children produced no correct velars in all attempts of velar /k/ being transcribed as alveolar [t]. The other participant produced 5% of all attempts of /k/ correctly. The rest (95% of all attempts) were transcribed as [t], which had two variants: centrally released or laterally released (Byun et al., 2016b).

Cleland et al. (2017) reported that in the ultrasound analysis, six children produced near identical shapes for /t/ and /k/ tongue splines, which overlapped indicating no covert contrast. The results of the t-tests with $p \leq 0.05$ criterion also showed no statistically significant differences were found between tongue contours. Therefore, contrary to the expectations, none of the children showed obvious evidence of covert contrast and displayed near identical tongue contours for /t/ and /k/ (Cleland et al., 2017).

Cleland et al. (2017) also conducted a second study, which was a case study on one child identified during the broad phonetic transcription analysis. The clinician also suspected unusual tongue shapes in this participant based on her own visual judgement of the participant's ultrasound tongue images. This child was noted to have retroflexed [ ] during production of /k/. Although this child showed no evidence of covert contrast, she clearly showed a 'covert error' (i.e., retroflexion) when her ultrasound tongue images were visually inspected. Therefore, Cleland et al. (2017) concluded that although they were unable to find covert contrasts in the seven participants in mid-sagittal ultrasound images, one child in particular showed evidence of unusual tongue shapes during productions of /t/ and /k/. This may suggest that some children with persistent velar fronting could present with articulatory (motor-based) difficulties.

2.5.3 The study by Cleland and Scobbie (2021)

Cleland and Scobbie (2021) conducted a study comprising of two groups of participants using two types of ultrasound measures: KTmax and direct visual inspection of tongue splines. One group consisted of typically developing children ($n = 30$) who showed appropriate alveolar-velar distinction. The other group consisted of participants who were persistent velar fronters ($n = 5$) with SSDs who were undergoing U-VBF intervention (Cleland and Scobbie, 2021). The aims of the study were twofold: The first aim was to "provide developmental articulatory norms for the alveolar-velar distinction in 30 English-speaking typically developing (TD) children" (Cleland and Scobbie, 2021, p. 1). The second aim was to compare TD children's KTmax norms among the persistent velar fronters before, during and after intervention, thereby demonstrating the usefulness of the reported measures (Cleland and Scobbie, 2021).

Cleland and Scobbie (2021) reported that TD children showed KTmax measures in their tongue splines which showed that they reliably distinguished /k/ and /t/. However, ”the children with persistent velar fronting showed KTmax values near zero before intervention, showing a complete merger between /k/ and /t/” (Cleland and Scobbie, 2021, p. 1). During intervention stage, these children showed KTmax values which varied. But during the post intervention period, they showed KTmax values which were within the range of TD children (Cleland and Scobbie, 2021).

This study suggests that similar to Cleland et al. (2015b), children who were persistent velar fronters acquired velar-alveolar contrast in an articulatorily gradient manner. Furthermore, this study found that KTmax measure can be a useful articulatory norm for the dorsal differentiation in alveolar and velar stops in TD children (Cleland and Scobbie, 2021).

2.5.4 The study by Byun et al. (2016a)

Byun et al. (2016b) conducted a study with four preschool children, two of whom aged 4;2 years and 4;5 years respectively, displayed velar fronting while the other two displayed an accurate velar-alveolar contrast. These initial assessments were based on the first author’s transcriptions. This study used two types of instrument-based studies. One was an acoustic study using a spectrogram, which utilised two different acoustic measures: VOT (conducted with t-tests) and spectral moments of the burst (see section 2.4.1.3). The other study used UTI which was an instrumental articulatory study. This method used DEI (Byun et al., 2016a).

Out of the two participants displaying velar fronting, one showed ”across-the-board velar fronting” while the other participant Max, showed ”positional fronting affecting only syllable-initial velars” (Byun et al., 2016a, p. 9). Byun (2012) previously explained that this could be due to phonological process brought about by specific constraints in children’s speech-motor control. According to Namasivayam et al. (2020) previous studies by Byun (2012) and Byun and Tessier (2016) have shown that ”In the case of positional velar fronting, the phonetically grounded ’MOVE-AS-UNIT’ constraint is eliminated from the grammar as the tongue-jaw complex matures” (p. 2).

Although in Byun et al. (2016b) the participant Max does not have across-the-board velar fronting but rather displays positional velar fronting, the child already is able to perceptually differentiate alveolar-velar positions linguistically/cognitively. Therefore, Cleland and Scobbie (2021) suggest that the child ”clearly does have the ability to differentiate the coronal and dorsal parts of the tongue some of the time: It is therefore tempting to conclude that this must be a truly phonological issue rather than motoric or articulatory” (p. 2). According to James et al. (2000), the existence of a covert contrast shows that, the child has phonological knowledge but is having difficulties in the articulatory movements.

2.6 Rationale for present study

In the literature surveyed above there have been only two prior UTI studies which attempted to investigate covert contrast among children with persistent velar fronting with an unknown cause and who substituted [t] for /k/. They are: the studies by Cleland et al. (2017) and Byun et al. (2016b) (see section 2.5.2 and 2.5.4). Both these studies used ultrasound measures such as direct visual inspection of tongue splines (see 2.4.1.5) and t-tests (see section 2.4.1.4.2). The aims of these studies were to investigate whether school-aged children with SSDs of unknown origin and persistent velar fronting, showed a covert contrast at the articulatory level (i.e., in tongue movements), suggesting an articulatory/motoric cause rather than phonological cause to their SSD.

Cleland et al. (2017) reported no evidence of covert contrast among the participants who substitute [t] for /k/. Instead, the authors found evidence of covert errors. It is to be noted that Cleland et al. (2017) used word level speech material data (i.e., singleton minimal pairs beginning with /t/ and /k/) only. These were a longer word list of single productions of many words containing velars and alveolars including minimal pairs. The participants in the present study used non-words (e.g., /aka/ & /ata/) and words (e.g., singleton minimal pairs begin with /t/ and /k) have aided with visual instructions on a computer screen.

The study by Byun et al. (2016b) used both acoustic measures (VOT and spectral moments) (see section 2.4.1.3) as well as ultrasound measures (DEI) (see 2.4.1.4.2), to investigate covert contrast. The study found evidence of covert contrast in one out of two participants having velar fronting. However, it was reported that the acoustic measure (spectral burst data) and the ultrasound measure (DEI) differed significantly between target alveolars and fronted velars (Byun et al., 2016a). These two measures also did not agree during the recording session in which covert contrast was detected. Byun et al. (2016a) stated that, ”spectral burst data appear to be more informative than the DEI measure” (p. 15). Since DEI quantifies the extent of excursion of the tongue’s dorsum (see 2.4.1.4.2), calculation of DEI may be adversely affected by shadows in the ultrasound image formed by the hyoid bone and the mandible on either end of the tongue spline (Byun et al., 2016a).

What is also noteworthy is that, both these previous studies did not include a Control Group (CG) of children with SSDs of unknown origins, but without velar fronting, for comparison.

Therefore, a knowledge gap was identified where the present study could replicate one of these studies, with new speech material data, to corroborate or compare with the findings of the two previous studies. Due to the availability of new speech material data for secondary data analysis from the ULTRAX2020 project (Ultrax, 2017, nd), whose participants’ demographic data (e.g., age, diagnosis) were similar to that of participants in Cleland et al. (2017) study, it was decided to replicate and extend the study by Cleland et al. (2017). The present study did not use DEI but rather, used the UTI based methods reported by Cleland et al. (2017). In addition, the present study extended the previous study by Cleland et al. (2017) using extra methods and measures found in the literature. They are described next.

It was decided to use two types of speech material data, where one uses t/k singleton minimal pairs, while the other proceeds along the articulatory route as using repeated imitations of non-lexical /aka/ & /ata/. It was thought that this would be useful for comparison when trying to identify covert contrast. To our knowledge, this is the first study that has used two different types of speech material data among persistent velar fronters of unknown origin, using UTI.

It was also decided to add a Control Group (CG) to the present study. The CG consists of participants who did not substitute [t] for /k/ but also have SSD of an unknown origin. The inclusion of CG in the present study would be useful to compare between participants who did not substitute [t] for /k/ with those who did. Finally an extra ultrasound measure - KTmax, which was not used in the Cleland et al. (2017) study, was added to the present study as a way of quantifying the degree of covert contrast. KTmax was previously introduced by Cleland and Scobbie (2021) (see section 2.4.1.4.3). The present study uses KTmax measure for two purposes:

- To compare coronal and dorsal separation between /k/ and /t/ tongue splines for each participant and,
- To compare SG participants KTmax measures with normative KTmax (e.g., CG participants).

2.7 Overall Aims, Research Questions and Hypotheses

The overall aim of the present study is to replicate and extend the study by Cleland et al. (2017). That is, to investigate whether school-aged children with SSDs of unknown origin and persistent velar fronting, show covert contrasts at the articulatory level. This would suggest an articulatory, rather than phonological cause to their SSD.

The present study has four main aims, each are presented next with the respective research questions hypotheses.

Aim 1:
To replicate the ultrasound t-test method described in Cleland et al. (2017) with a different group of children and different speech materials.

Research Question 1:
Does using a different group of children and different speech materials lead to the same finding that there is no evidence of covert contrast between /t/ and /k/ in ultrasound data?

Hypothesis:
We hypothesised that most children with persistent velar fronting would not show a covert contrast, indicated by no statistically significant difference between tongue shapes for /k/ and /t/ attempts at the point of maximum constriction, in line with Cleland et al. (2017).

Aim 2:
To extend the study by Cleland et al. (2017) by including a Control Group of children

with an overt contrast between /k/ and /t/, also with persistent SSDs.

Research Question 2:
Do children with an overt contrast (i.e., CG) show a difference at the articulatory level between /t/ and /k/?

Hypothesis:
We hypothesised that CG would show a consistent, significant, difference between tongue shapes for /k/ and /t/ attempts at maximum constriction.

Aim 3:
To compare the magnitude of KTmax to the norms presented in Cleland and Scobbie (2021).

Research Question 3:
Does KTmax for children with persistent SSDs fall within the norms for typically developing children?

Hypothesis:
We hypothesise that children with an overt contrast (i.e., CG) would have a KTmax value within the normal range, whereas children with a covert contrast would have values below the normal range.

Chapter 3

Method

3.1 Study Design

The present study is based on secondary data from the Ultrax2020 project (Ultrax, 2017). It involved analysing secondary data from 16 participants, divided into two groups called the Study Group (SG), who substitute [t] for /k/ and the Control Group (CG), who did not substitute [t] for /k/, instead producing an overt, appropriate contrast between /t/ and /k/.

To achieve the Aim 1 described in section 2.7 and to find evidence of any covert contrast in the SG, first a transcription-based phonological assessment, or perceptual study, was carried out. Next, we carried out an articulatory-based study using ultrasound and audio data, drawing ultrasound tongue splines at the maximum point of constriction, using AAA software (Cleland et al., 2017). Then articulatory data was analysed using three measures, namely the articulatory t-test (see section 2.4.1.4.1), KTmax (see section 2.4.1.4.3) and visual inspection (see section 2.4.1.5).

We then analysed the data to determine whether there was any evidence of a quantifiable difference between UTI splines between /k/ and /t/ attempts. Finally, we compared the perceptual analysis study results with the UTI-based articulatory instrumental study results, to find out which participants, if any, showed a covert contrast.

If covert contrast was present, we would expect a perceptual neutralisation where /k/ would sound the same as /t/ (i.e., both [t]). In the UTI based articulatory study, we would expect to see a quantifiable difference.

3.2 Ethical consent for the present study

Ethical approval had already been obtained for the Ultrax2020 project by the primary researchers and approval to share its data was obtained from South East Scotland Research Ethics Committee 22001 (reference 18/SS/0072). Therefore, we were granted access to the Ultrax2020 project's ultrasound data to perform secondary data analysis.

3.3 ULTRAX2020

The overall aim of Ultrax2020 was to develop an automatic ultrasound based diagnostic assessment for identifying and categorising speech errors in children with SSDs (Cleland et al., 2018). To achieve this, clinicians collected ultrasound data from children in partnership with three NHS boards across Scotland. Data was collected by the children's Speech and Language Therapists. Data from 44 children (31 males and 13 females) aged 5-12 years was collected between February 2019 and January 2020. The most common type of SSD observed was an articulation disorder among 15 of these children. Out of these, fourteen children had a cleft lip and/or palate. The remaining children had a range of SSDs including phonological disorders, childhood apraxia of speech, residual speech sound errors, and inconsistent phonological disorders (Cleland et al., 2018).

The main eligibility criteria to be included as a participant for the Ultrax2020 project were:

- be an English-speaking child,
- be aged between 5-16 years,
- have a current diagnosis of Speech Sound Disorder (SSD).

3.4 Selection of participants in the present study

In the Ultrax2020 project there was a range of children with SSDs, aged between 5-12 years. However, for the present study a sub-set of participants with a current diagnosis of SSDs with an unknown cause were selected. Therefore, among the original participants of Ultrax2020 study, participants who had a current or repaired syndromic or non-syndromic cleft lip and palate of any type, those with evidence of severe/profound hearing loss or those with evidence of severe/profound learning disability, were excluded. Overall, there were data from a total of 21 children who qualified under the inclusion and exclusion criteria above, were selected from the Ultrax2020 project.

Out of these 21 participants, a further five participants were removed due to their data being unsuitable. Amongst them, three participants were excluded due to poor quality of the UTI data, while two participants were excluded due to insufficient ultrasound and audio recordings. Therefore, UTI and audio data from 16 participants were analysed. Their demographic data are shown in Table 1.

Table 1

Demographic data of the participants included in the present study

Participant	Age	Sex	Speech disorder recorded by SLT	Additional information by SLT
SG_1 (UX05)	6;3	M	Motor speech disorder	Poor production
SG_2 (UX12)	5;11	M	Inconsistent phonological disorder or CAS	
SG_3 (UX15)	7;5	M	Articulation and phonological disorder, with possible coordination difficulties	Difficulties with grammar
SG_4 (UX02)	9;2	M	Inconsistent speech sound disorder, query CAS	Swallowing difficulties
SG_5	7;4	F	Articulation disorder	
SG_6	8;5	M	Articulation disorder	
CG_1 (UX03)	9;2	M	Articulation disorder	
CG_2 (UX21)	11;2	F	Articulation disorder	
CG_3 (UX19)	8;4	M	Articulation disorder	Bilateral conductive hearing loss (wears hearing aids/corrected)
CG_4 (UX18)	12;11	M	Phonological disorder	Global Developmental delay
CG_5 (UX01)	8;2	M	Phonological disorder	
CG_6 (UX13)	6;4	M	Inconsistent phonological disorder	
CG_7 (UX14)	5;2	M	Diagnostic process ongoing	
CG_8 (UX20)	11;6	M	Motor speech disorder	ADHD learning difficulties
CG_9 (UX23)	8	M	Unknown	Unknown
CG_10 (UX02)	6;3	M	Articulation disorder	

3.5 The ultrasound recording set-up

The present study used speech material data from ULTRAX2020 project (Ultrax, 2017). In Ultrax2020, an equipment set-up comprising of a laptop running SonoSpeech® Software (version 2.17) and a Micromachine supplied by Articulate Instruments Ltd (AAA, 2012) along with an ultrasound microconvex probe, was used to record high-speed ultrasound data (Cleland et al., 2018). The echo return data from the ultrasound probe was recorded at roughly 100 fps over a field of view of 162^o. This field of view of 162^o allowed the best view of the tongue, including both the hyoid and mandible shadows. Using this, the clinician was able to view a midsagittal view of the tongue. The ultrasound probe was stabilised using a custom-made lightweight flexible headset (Sugden and Cleland, 2022). This stabilised headset is shown in Figure 6.

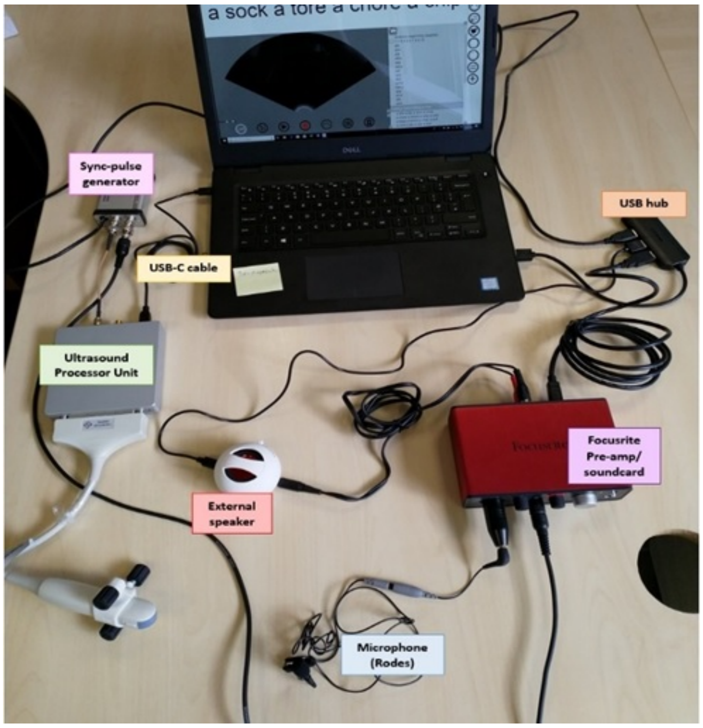
Figure 6

The stabilisation headset used for ultrasound recordings in Cleland et al. (2018).

A microphone attached to this headset recorded live audio recordings simultaneously. The software running in the laptop internally synchronised the ultrasound and recorded audio (Sugden and Cleland, 2022). This recording set-up is shown in Figure 7.

Figure 7

Synchronised Micro Speech Research Ultrasound System coupled to the 5-8 MHz micro convex probe.



The recording process was conducted in a quiet clinical ultrasound recording setting by community-based SLTs, who were trained in data collection and assessment by university speech scientists conducting the research (Cleland et al., 2018).

3.5.1 UTI data recording process

3.5.1.1 Speech recording materials

According to Cleland et al. (2020), the ULTRAX 2020 project used two types of speech material data. They were non-lexical speech material data and real word sentences (from two to six words in length) (Cleland et al., 2020). Therefore, the present study also used the following two types of speech material data, adapted from the ULTRAX 2020 project.

- Non-lexical speech material data: ten repetitions of all voiceless English obstruents and sonorants in /aCa/ context. There were 14 /aCa/ context consonants, for /l, j, w, ŋ/ (i.e., k, t, n, ɹ, θ, l, j, t͡ʃ, ŋ, p, m, f, w, s, ʃ), and one /bang/(ŋ) context consonant. From this /t/ and /k/ were selected mainly.

- Singleton minimal pairs that begin with /t/ and /k/. Six pairs of singleton minimal pairs, or near minimal pairs, that begin with /t/ and /k/: tore/core, tool/cool, tip/kip, top/cop, tan/can, team/keep

The speech material data used in the ULTRAX2020 project contained a longer list which included minimal pairs (e.g., a core a sip a cop a tool). In the ULTRAX 2020 project, the SLT in clinic requested the participant to produce both types of speech material data using visual prompts on screen. If the participant was unable to do so, then the SLT verbally prompted, saying the production when and where necessary (Cleland et al., 2018).

3.6 Perceptual study method

The aims of the perceptual study were:

- To divide the participants into a SG and CG, by conducting inter-rater broad phonetic transcriptions of /aka/ & /ata/ and t/k singleton minimal pairs,
- To carry out baseline measures including (a) percentage of consonant correct-adjusted (PCC-A) and (b) to analyse speech sound error types,
- To conduct inter-rater broad phonetic transcriptions to investigate transcription reliability, by involving an experienced SLT as the second transcriber.

3.6.1 Dividing participants into SG and CG

In total there were 16 participants. we transcribed all participants, out of these participants who substituted [t] for /k/ in /aka/ & /ata/ context or in t/k minimal pairs were named as the SG. The rest of the participants were named as the CG.

3.6.1.1 Study Group (SG)

The SG comprised of five males and one female, aged between six to eight years. All participants in SG substituted [t] for /k/ in the /aCa/ context (e.g., ten repetitions of

/aka/ - non lexical speech material data) or in word level (e.g., singleton minimal pairs begin with /t/ and /k/). SG_3 was the only participant who did not substitute [t] for /k/ in the /aCa/ context but substituted [t] for /k/ in t/k singleton minimal pairs. All other participants (n = 5) substituted [t] for /k/ in both levels. Furthermore, 5/6 participants substituted ɳ → n and 3/6 participants substituted ɳ → ɲ, which are alveolar and palatal substitutions for the velar nasal /ŋ/. The present study also included broad phonetic transcriptions for t/k singleton minimal pairs for SG participants to clarify how they substitute [t] for /k/ in word level.

Table 2 presents the Velar Fronting found among SG and CG in /aCa/ context.

Table 2

Other Velar Fronting difficulties found among participants in their broad phonetic transcriptions of /aCa/

Participant	**Velar fronting errors in aCa context**			
	k → t	ɲ → ŋ	ɲ → n	g → d
SG_1	✓	✓	✓	Did not assess
SG_2	✓	×	✓	Did not assess
SG_3	×	×	✓	Did not assess
SG_4	✓	✓	✓	Did not assess
SG_5	✓	✓	✓	Did not assess
SG_6	✓	×	×	Did not assess
CG_1	×	×	×	Did not assess
CG_2	×	×	×	Did not assess
CG_3	×	×	×	Did not assess
CG_4	×	×	×	Did not assess
CG_5	×	×	×	Did not assess
CG_6	×	×	×	Did not assess
CG_7	×	×	×	Did not assess
CG_8	×	×	×	Did not assess
CG_9	×	×	×	Did not assess
CG_10	×	×	✓	Did not assess

3.6.1.2 Control Group (CG)

The CG comprised ten children, six males and four females, who were aged between six to twelve years. None of the participants in the CG substituted [t] for /k/ in broad phonetic transcriptions in /aka/ context. In other words, they presented with no velar fronting difficulties in non-lexical speech material data. However, one participant, CG_10, presented other velar fronting difficulties, substituting [ɲ] for /ŋ/. This is evident in the summary presented in Table 2.

The present study also conduced broad phonetic transcriptions for t/k singleton minimal pairs for the CG participants to verify whether they substitute [t] for /k/ in word level.

3.6.2 Baseline Measures

Three types of baseline measures were conducted in the present study. They were: Percentage of Correct Consonants-Adjusted (PCC-A), speech error types and transcription reliability.

3.6.2.1 Percentage correct consonants-adjusted (PCC-A)

The main purpose of calculating PCC-A is to allow clinicians to assess intelligibility of speech. In the present study PCC-A was carried out to obtain further information about each participant's presentation of correct number of consonants or speech intelligibility, and to categorise participants into the SG and CG. The PCC metric is a quantitative severity measure developed by Shriberg and Kwiatkowski (1982). It calculates the percentage of intended consonant sounds that were articulated correctly, usually in a conversational sample of typically 5-10 minute duration (Shriberg and Kwiatkowski, 1982). Dale et al. (2020) mentions that "the PCC metric considers only the production of consonants, which are judged as correct or incorrect based on the PCC sampling rules" (p. 23).

Shriberg and Kwiatkowski (1982) suggest the following four severity classifications levels based on calculated PCC values:

- PCC $>$ 90%: mild,

- 65% < PCC < 85%: mild-moderate,
- 50% < PCC < 65%: moderate-severe and
- PCC < 50%: severe.

To calculate PCC-A in the present study, we first identified uncommon clinical distortions, deletions or substitutions as incorrect, in all participants (n = 16). As all the participants were older than five years and were past what is commonly considered as the typically developing stage for speech, they were not expected to have common speech distortions/errors (Smit et al., 1990). Therefore, these distortions/errors identified were marked as incorrect. We then counted the total number of consonants and the number of correct consonants for each /aCa/ consonant resulted in 15 consonants for each participant (n = 16) in the present study. Next, I divided the number of correct consonants by the total number of consonants and multiplied by 100% to determine the PCC-A. Finally, we classified SG and CG according to the severity rating above.

3.7 UTI-based articulatory study

The aim of carrying out an UTI-based articulatory study was to identify quantifiable, subtle differences (e.g., covert contrast, covert errors, or undifferentiated lingual gestures) among SG participants, which we were unable to observe during the perceptual study.

Therefore, using AAA software (AAA, 2012), ultrasound data was transformed into quantitative/qualitative articulatory data. For this articulatory study, we used ultrasound corresponding to the same speech material used to conduct the perceptual study. They were recordings of ten repetitions of all voiceless English obstruents and sonorants in /aCa/ context and singleton minimal pairs that began with /t/ and /k.

These UTI data were transformed into articulatory information by fitting tongue splines at the maximum point of constriction (i.e., place of articulation) mid-sagittally. The maximum point of constriction was marked using semi-automatic detection function offline by visually looking at the highest position that the tongue could reach.

To discover subtle differences in tongue splines indicating a covert contrast, covert errors, or undifferentiated lingual gestures, three types of analysis were performed offline using the AAA software. They were articulatory t-test, KTmax and direct visual inspection of tongue splines. This was necessary because, such subtle differences are not easily discerned from live, real-time, or recorded raw UTI.

3.7.1 Articulatory analysis

The steps that were taken to annotate the ultrasound data were:

- Step A: Annotation
- Step B: Alignment of splines at key-frames
- Step C: Identifying keyframes
- Step D: Copying splines into the spline workspace

They are described next.

Step A: Annotation

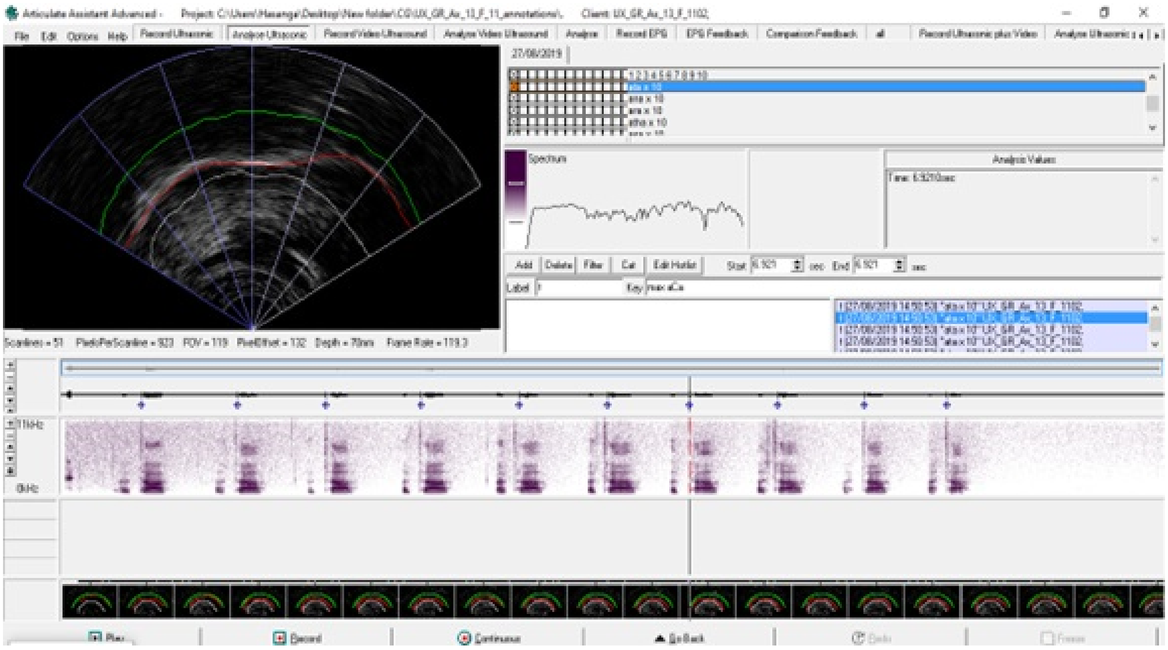
Using the AAA software (AAA, 2012), we annotated the maximum point of constriction for /k/ and /t/ attempts for 10 repetitions of /aCa/ context. Therefore, for each segment one annotation was marked and labelled by looking at the maximum point of constriction with the hard or soft palate. This resulted in ten labelled annotations each for /k/ and /t/ per participant. However, one participant in SG, namely SG_4, had only three repetitions of /k/ and three repetitions of /t/.

Step B: Alignment of splines at key-frames

The tongue splines were drawn using AAA software after labelling the annotations for /k/ and /t/ for each participant as described in step A. A spline indicating the tongue surface was fitted to the image semi-automatically for /k/ and /t/ using the 'automatic edge tracking function' Found in the AAA software. This is shown in Figure 8.

Figure 8

Screenshot of AAA software (AAA, 2012) showing ultrasound image and spectrogram



Step C: Identifying keyframes

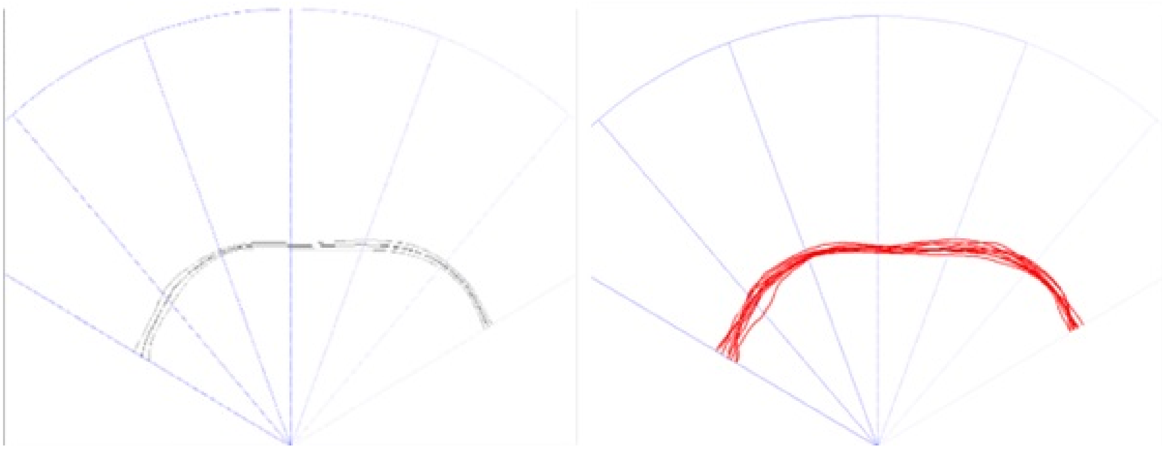
Keyframes are a central concept in the creation of tongue splines in UTI. They are points in time for which a spline can be defined or drawn (AAA, 2012). There are three options for creating keyframes: 'single keyframe', 'every Nth video frame' and 'every N milli-seconds'. For the present study we used the 'every Nth video frame' method, where $N = 1$ was specified by me in the "frames" contour in AAA software.

Step D: Copying splines into spline workspace

Although splines were fitted every $N = 1$ frame, we exported only splines that had maximum point of constriction. Therefore, generally there were ten /aka/ and ten /ata/ tongue splines copied into the spline workspace. Then the 'average' fan spline for /t/ and /k/ was calculated and distances between these splines was measured in the spline workspace. After this step, the Mean and standard deviation (S.D.) for /k/ and /t/ was calculated using the AAA software. Finally, Mean tongue contours for /k/ and /t/ were drawn in spline workspace is shown in Figure 9.

Figure 9

Calculating the Mean and S.D. (Left] by averaging all tongue splines for /t/ [Right]. Tongue root to the Left and tongue tip to the Right.



3.7.2 Transcriptions

The main author listened to each ultrasound audio recording offline a multiple number of times, then carried out broad phonetic transcriptions for each participant's productions. In doing so, the following guidelines were followed, as described in Sugden and Cleland (2022):

- listen to each audio recording containing a participant's ten repetitions of the target consonant in /aCa/ context, and t/k singleton minimal pairs. There was no restriction on the number of times the audio file was played back.
- Note down their transcription using both IPA and extIPA symbols (including diacritics where necessary). Then move on to the next recording.
- Both transcribers to use broad phonetic transcription for most of the speech samples, but use narrow transcription for errors,
- Transcribe only the target consonant in the /aCa/ recording (i.e., ignore the vowels), include all different productions made by the participant or include only one transcription if all the repetitions sounded the same.

It took multiple sessions of approximately six hours to transcribe all the participants data. Using these recorded ultrasound and audio data, we conducted two main studies: broad phonetic transcription based perceptual study and UTI-based articulatory study.

3.7.3 Conducting articulatory measures

Next, three articulatory measures: t-test, KTmax and direct visual inspection of tongue splines, were conducted to analyse /k/ and /t/ ultrasound tongue image splines to measure any subtle differences.

3.7.3.1 t-test

A t-test was carried out using the calculated average tongue splines and S.D. for /k/ and /t/ for each participant. These averaged splines were compared statistically using the built-in statistical difference function by pressing 'Diff' in the AAA software. Using this function, the significance, is calculated along 42 fan-lines emanating radially from the probe to the tongue surface (Cleland et al., 2017). The Mean difference for a given fan-line is regarded as significant if $p \leq 0.001$. When this is the case, the AAA software draws a thicker spoke as shown in Figure 10. It has been stated that, AAA software uses a 2 tailed test using Welsh-Satterthwaite equation (Cleland et al., 2017).

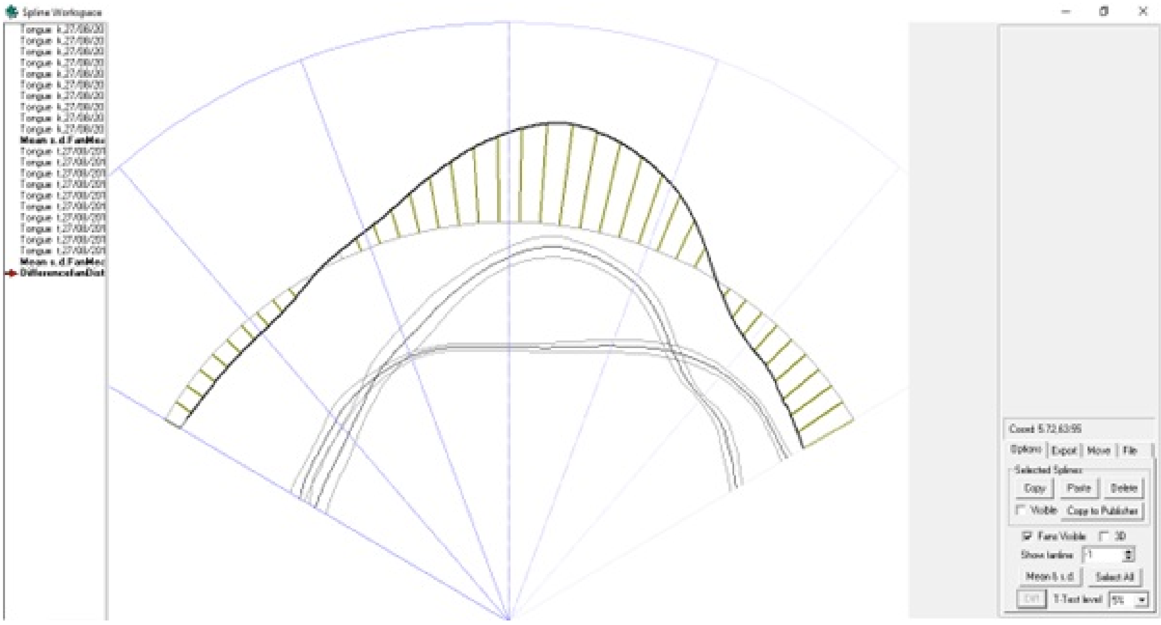
This t-test function in AAA uses a fan-shaped measurement space that aligns with the fan shaped imaging area. A set number of fan lines is drawn radially starting from the probe's location (Cleland and Scobbie, 2021). In the present study, the fan shaped area comprised of 42 fan lines as shown in Figure 10. We used the criteria of a minimum of seven adjacent fan-lines, three fan lines to the Left of the maximum point of constriction and three fan lines to its Right, to test for a significant difference between /t/ and /k/ attempts. That is, if the average of the said seven adjacent fan lines had a computed $p \leq 0.001$, it was considered as a significant difference.

3.7.3.2 KTmax

Next, we calculated KTmax measure. Roxburgh (2018) stated that, when there is a global significant difference between two tongue splines, it is important to quantify this difference. Even in cases where there is no area of significant difference, computing KTmax measure still allows us to quantify the relative similarity of two tongue splines

Figure 10

Fan shaped measurement space in AAA (AAA, 2012) consisting of 42 fan-lines for computing t-test (fanlines 1-42 from Right to Left)



that may have insufficient number of significantly different adjacent fan lines, as per the criteria in section 3.7.3.1 for the t-test. Thus, KTmax potentially allows tracking of non-significant changes in /k/ and /t/ tongue splines.

It has been said that it is important to quantify raw similarity/difference and to quantify the size of significant differences (Roxburgh, 2018). According to Cleland and Scobbie (2021), KTmax compares the coronal and dorsal separation between /k/ and /t/ in an articulatory context. The KTmax, which is also known as the maximum radial difference, indicates the highest degree of dorsal separation in terms of linear depth of /k/ and /t/ in the crescent.

Calculating KTmax

To calculate KTmax, the polar co-ordinates of the tongue spline knot locations (i.e., distance from the origin at a given fan angle) for [k] and [t] were exported to MS Excel as per the method reported in Cleland and Scobbie (2021). Then, we calculated the differences between the two knots on each fan-line in which, [k] is further from the probe than [t] (i.e., Mean [k] for each fan-line and Mean [t] for each fan-line). This maximum radial difference between [k] and [t] generate a crescent shape which called the KT crescent depth. This is calculated as:

Maximum radial difference (KTmax) = longest fan line in the KT crescent depth(mm) or which fan line has the biggest distance between Mean [k] for each fan-line and Mean [t] fan-line.

Then the highest maximum radial difference (KTmax) among the 10 repetitions is selected (Cleland and Scobbie, 2021).

3.7.3.3 Direct visual inspection of tongue splines (qualitative measure)

In this step, we visually compared similarities and differences in tongue spline shapes offline, to identify subtle differences between corresponding /k/ and /t/ tongue spline shapes. I visually inspected small sections in detail, whether the small section of curves are different (i.e., the tongue root is the same in both /k/ and /t/ splines, but the tip of one production is raised compared to the other). Furthermore, we visually inspected where the maximum point of constriction occurred, and whether it was palatalised or not.

3.7.3.4 Speech error types

Speech error classification will help to spot any patterns among participants in SG and CG in /aCa/ speech material data. Speech error classification was not conducted for t/k singleton minimal pairs because of limited time availability and because assessing CG was not a priority in the present study.

In the present study, we looked for both developmental speech error types (e.g., inter-dentalisation, stopping, fronting gliding, affrication, deaffrication, alveolarization, depalatalization, labialization and disordered speech error types (e.g., double articulation, retroflexion, Palatalisation, Palatal fronting, Nasalisation, Lateralisation, r-distortion, Retracted, backing and lateralisation).

Then the total number of speech errors each participant displayed was counted. Note that, in this method some participants could present the same type of speech error more than once. For example, SG_2 presented lateralisation twice for two different /aCa/ productions.

3.7.3.5 Transcription reliability

An inter-rater reliability check was carried out to determine the validity and reliability of the phonetic transcriptions completed by me (the author). It has been said that, the 'reliability' of a phonetic transcription estimates the repeatability of judgements generated by assessors within a given transcription system (Shriberg and Lof, 1991). Therefore, "both intra-judge and inter-judge agreements assess the degree of similarity in descriptions of speech by persons, trained to use the same transcription system" (Shriberg and Lof, 1991, p. 3).

For the present study, inter-rater reliability for phonetic transcriptions of /aCa/ productions were calculated using the concept of near functional equivalence introduced by Seifert et al. (2020). The process involved two transcribers identifying differences in the two transcriptions in vowels, consonants and diacritics. One transcriber was me, while the second transcriber was an experienced SLT with a PhD and over 8 years researching disordered speech.

We transcribed all ten repetitions of /aCa/ productions for the 15 consonants for the 16 participants. The second transcriber then transcribed data from six (i.e., 40%) randomly selected participants. They comprised data from four participants in SG and two participants in CG. The selected participants were SG_1, SG_2, SG_3, SG_4, CG_3 and CG_8. The overall number of /aCa/ consonants the author transcribed was 2284, while the second transcriber transcribed a total number of 794 (i.e., 35%) /aCa/ consonants.

Then finally, we calculated near functional equivalence Seifert et al. (2020) by calculating consonant difference and diacritic difference between the two transcribed data by the two transcribers.

Chapter 4

Results and analysis

This chapter presents the results of both the transcription based perceptual analysis and UTI based articulatory analyses. The results of perceptual and articulatory analysis are based on two types of speech material data: ten repetitions /aka/ & /ata/ context (non-lexical) and t/k singleton minimal pairs (single word level), which may reflect different levels of psycholinguistic processing.

4.1 Perceptual study results

The perceptual study's results contained broad phonetic transcriptions of ten repetitions of all (15) voiceless English obstruents and sonorants in an /aCa/ context and one /bang/(ŋ) for both SG and CG participants and broad phonetic transcriptions of t/k singleton minimal pairs for the SG participants only.

There were 14 /aCa/ context consonants, for /l, j, w, ŋ/ (i.e., k, t, n, ɹ, θ, l, j, t͡ʃ, ŋ, p, m, f, w, s, ʃ), and one /bang/(textipaŋ) context consonant. From the ten repetitions of these, four different types of results were generated and analysed. They were:

- results and analysis of broad phonetic transcriptions of /aka/ & /ata/ context
- results of PCC-A,
- results of speech error types,
- transcription reliability test.

4.1.1 Results and analysis of broad phonetic transcriptions

Table 3 presents the summary of all participants' broad phonetic transcriptions of /ata/ and /aka/ and indicates how they were subsequently categorised into either the SG or CG.

Table 3

Summary of participants' broad phonetic transcriptions of ten repetitions of [k] in /aka/ context and [t] in /ata/ context.

Participant	/k/ realisations	/t/ realisations
SG_1	6;3	M
SG_2	5;11	M
SG_3	7;5	M
*SG_4	9;2	M
SG_5	7;4	F
SG_6	8;5	M
CG_1	9;2	M
CG_2	11;2	F
CG_3	8;4	M
CG_4	12;11	M
CG_5	8;2	M
CG_6	6;4	M
CG_7	5;2	M
CG_8	11;6	M
CG_9	8	M
CG_10	6;3	M

***Note:** SG_4 only had three repetitions in of /aka/ and three repetitions of /ata/.

When looking at SG participants' /k/ realisations in the /aka/ context, 5/6 participants substituted [t] for /k/ in at least one attempt out of ten repetitions. SG_1 did this substitution three times, SG_2 and SG_4 did this twice, while SG_5 and SG_6 did this in all ten attempts. The only participant that did not substitute [t] for /k/ in /aka/ context in any attempt was SG_3.

However, SG_3 was included in the present study due to substituting [t] for /k/ in

Singleton minimal pairs that begin with /t/ and /k/ following the present study's inclusion and exclusion criteria. In other words, real words have influenced the allocation of SG_3 into SG, even though SG_3 did not substitute [t] for /k/ in nonsense-words. First, nonsense words were analysed in the /aka/ & /ata/ context. See section 4.2.5 for further details of SG_3.

In the CG, all participants were able to produce [k] in the /aka/ context in ten out of ten attempts. With regards to [t] productions in the /ata/ context, CG_7 substituted [k] for /t/ in 3 /10 attempts.

4.1.2 Results and analysis of PCC-A

The four severity classifications levels based on calculated PCC values suggested by Shriberg and Kwiatkowski (1982) were mentioned previously in section 3.6.2.1.

Table 4 summarises the results of the PCC-A calculation and each participant's severity classification alongside the Mean and S.D. of the PCC-A for SG and CG as a whole.

Table 4

Summary of participants' broad phonetic transcriptions of ten repetitions of [k] in /aka/ context and [t] in /ata/ context.

Participant	Number of correct consonants	Total number of consonants	Percentage of correct consonant (severity rating)
SG_1	76	150	50.7 (moderate to severe)
SG_2	52	150	34.7 (severe)
SG_3	55	149	36.9 (severe)
SG_4	24	45	53.3 (moderate to severe)
SG_5	86	150	57.3 (moderate to severe)
SG_6	89	140	63.6 (moderate to severe)
Mean for SG			49.4
S.D. for SG			11.4
CG_1	103	150	68.7 (mild to moderate)
CG_2	119	150	79.3 (mild to moderate)
CG_3	90	150	60 (moderate to severe)
CG_4	97	150	64.5 (moderate to severe)
CG_5	111	150	74 (moderate to severe)
CG_6	111	150	74 (moderate to severe)
CG_7	74	150	49.3 (severe)
CG_8	88	150	56.7 (moderate to severe)
CG_9	94	150	62.7 (moderate to severe)
CG_10	100	150	66.7 (moderate to severe)
Mean for CG			65.8
S.D. for CG			8.8

When looking at SG, participants SG_1, SG_4, SG_5 and SG_6 fall into the "moderate to severe" category, while SG_2 and SG_3 fell into the "severe" category. In the CG, participants CG_1, CG_2, CG_5 and CG_6 fell into the "mild to moderate" category, while CG_3, CG_4, CG_8, CG_9 and CG_10 participants fell into "moderate to severe". However, CG_7 fell into the "severe" category.

Next, we conducted a t-test using the Mean of SG and CG participants presented above in Table 4. This was conducted using the "Data Analysis" Tool in MS Excel.

The two-tailed t-test method was used to calculate the p-value. The result of the t-test is shown in Table 5.

Table 5

Calculated t-test for SG and CG PCC-A.

Variable	**CG**	**SG**
Mean	65.8	49.4
Variance	77.0	130.8
Observations	10	6
Hypothesized Mean Difference	0	
p-value ($T \leq t$) two-tail	0.01	

4.1.3 Results of Speech error types

Table 6 and Table 7 show the speech error types that the SG and CG participants displayed and the total number of different speech errors for each participant. The participants with the highest number of speech errors were SG_1 and SG_2 showing a total of 15 errors. The highest number of disordered speech errors were shown by SG_5 (6 errors) followed by SG_1 (5 errors).

Table 6

Summary of participants' broad phonetic transcriptions of ten repetitions of [k] in /aka/ context and [t] in /ata/ context

Speech error type	**SG_1**	**SG_2**	**SG_3**	**SG_4**	**SG_5**	**SG_6**.
th-fronting	1	1	1			1
Labialisation	1	1				
Velar fronting	3	2	1	3	3	1
Stopping				1		
Total developmental speech errors	5	4	2	4	4	1
Double articulation	1	1	2	2		
Retroflexion	2		1		1	
Palatalisation	1					
Palatal fronting				1	2	
Nasalisation	1					
Dentalisation		1				
Lateralisation	2			1	1	
r-distortion				1		
Interdentalisation					1	
Retraction					1	1
Backing						
Total disordered speech errors	5	4	3	4	6	1
Other speech errors	5	7	8	3	1	2
Total speech errors	15	15	13	11	11	4

Table 7

Speech error type classification for CG participants.

Speech error type	**CG_1**	**CG_2**	**CG_3**	**CG_4**	**CG_5**	**CG_6**	**CG_7**	**CG_8**	**CG_9**	**CG_10**
th-fronting					1			1	1	1
Labialisation				1	2			1	1	
Velar fronting										1
Stopping							2		1	
Gliding	1							1	1	
Lisp	2		4			1				
Total developmental speech errors	3	0	4	1	3	1	2	3	4	2
Double articulation								3		
Retroflexion				1						
Palatalisation						1	1			
Palatal fronting				1				2		
Nasalisation										
Dentalisation		1								
Lateralisation	1	2	1	1						
r-distortion	1					1				
Interdentalisation	1							1		
Retraction					1					
Backing										
Voicing error							1			
Inconsistency and deaffrication								1		
Total disordered speech errors	3	3	1	3	1	2	2	7	0	0
Other speech errors	2	1	2	0	1	1	2	1	4	3
Total speech errors	8	4	7	4	5	4	6	11	8	5

In both tables, where a speech error type could not be clearly identified, these instances were categorised under ”Other” category. Most participants presented ”Other” speech errors which were comparable in number to their total developmental speech errors and total disordered speech errors. Due to this, classification of participants of SG or CG as primarily having disordered or developmental speech errors was not possible. For example, SG_4 presented many speech error types such as double articulations (n →n͡ɲ, n͡d), distortions (r→ɹ), voicing errors (s→s̥), stopping (θ→d) and some unclassified (i.e., ”Other”) speech error types such as (r→z) and (j→l).

4.1.4 Transcription reliability

As explained in section 3.7.3.5, the diacritic difference calculated for six randomly selected participants in SG and CG are presented in Table 8.

Table 8

Diacritic differences calculated for each selected participant and the total diacritic difference.

Participant	Diacritic difference for 15 consonants of 10 repetitions of /aCa/ consonants	Total /aCa/ context consonants
SG_1	20	150
SG_2	31	150
SG_3	36	149
SG_4	10	45
CG_3	44	150
CG_8	9	140
Total	150	794

Then the diacritic difference for all the participants ($n = 6$) were calculated as:
Total number of diacritic differences = 150
Total number of consonants = 794
Diacritic difference = $\frac{150}{794} \times 100\% = 18.9\%$

Similarly, results of consonant difference for the same participants are tabulated in Table 9.

Table 9

Consonant differences calculated for each participant and total consonant differences.

Participant	Consonant difference for 15 consonants of 10 repetitions of /aCa/ consonants	Total /aCa/ context consonants
SG_1	27	150
SG_2	14	150
SG_3	41	149
SG_4	17	45
CG_3	36	150
CG_8	36	150
Total	171	794

Then the consonant difference for all the participants ($n = 6$) were calculated as:
Total number of consonant differences = 171
Total number of consonants = 794
Consonant difference = $\frac{171}{794} \times 100\% = 21.5\%$

Therefore, the present study's comparison of the two transcriptions indicated 81% agreement between the two assessors, having a diacritic difference of 18.9% and a consonant difference of 21.5% resulting in a 78.5% consonant agreement. Both the diacritic difference and consonant difference were less than 25%. For comparison, Ramsdell et al. (2007) reported a 58% to 62% agreement among coders while Zanichelli and Gil (2011) reported reliability of 89% between two examiners, for PCC reliability test. This shows that the present study's transcription reliability falls within the commonly acceptable levels for diacritic difference agreement and consonant difference agreement.

4.1.5 Broad phonetic transcriptions of singleton minimal pairs

Table 10 summarises the broad phonetic transcriptions of t/k singleton minimal pairs for SG participants. SG_1 and SG_2 substituted [t] for /k/ in 5/6 pairs, and once substituted [tʃ] for /k/. Participant SG_3 only substituted [t] for /k/ once and in other attempts /k/ was produced correctly. Participant SG_4 substituted [t] for /k/ twice, substituted [tʃ] for /k/ three times, while one attempt produced correctly. Participant SG_5 substituted [t] for /k/ in all six attempts. SG_6 substituted [t] for /k/ 4 times, substituted [tʃ] for /k/ once and one attempt was correct.

Table 10

Broad phonetic transcriptions of minimal pairs beginning with /k/ and /t/ for SG.

Singleton Minimal Pairs	SG_1	SG_2	SG_3	SG_4	SG_5	SG_6
core	tɔ:	tɔ:	kɔ:	tɔ:	tɔ:	tɔ:
tore	tɔ:	tɔ:	tɔ:	tʃɔ:	tɔ:	tɔ:
cool	tu:l	kɪu:l	ku:l	tʃu:l	tu:l	tu:l
tool	tu:l	tu:l	tu:l	tu:l	tu:l	tu:l
kip	tʃɪp	tɪp	kɪp	tʃɪp	tɪp	tʃɪp
tip	tʃɪp	tɪp	tɪp	tɪp	tɪp	tɪp
cop	tɒp	tɒp	tɒp	tɒp	tɒp	tɒp
top	tɒp	tɒp	tɒp	tɒp	tɒp	tɒp
can	tan	tan	kan	tʃan	tan	tan
tan	tan	tan	tan	tan	tan	tan
keep	ti:p	ti:p	ki:p	ki:p	ti:p	ki:p
team	ti:m	tʃi:m	ti:m	ti:m	ti:m	ti:m
Substituting [t] for /k/	5/6	5/6	1/6	2/6	6/6	4/6

Note: Underlined phonetic transcriptions indicated instances of [t] substituted for /k/, [t] substituted for /k/ and [t] substituted for /t/.

Table 11 summarises the broad phonetic transcriptions of t/k singleton minimal pairs for CG. CG_7 indicated a speech error when replacing /cool/ by [plɒs]. In addition, CG_7 substituted /k/ with [p] in /kip/ while substituting [k] for /t/ in /tip/. Therefore, CG_7 substituted [t] for /k/ once while CG_9 substituted [tʃ] for /k/ once. CG_9 ($n = 1$) also made other substitutions.

Table 11

Broad phonetic transcriptions of minimal pairs beginning with /k/ and /t/ for CG.

Singleton Minimal Pairs	**CG_1**	**CG_2**	**CG_3**	**CG_4**	**CG_5**	**CG_6**	**CG_7**	**CG_8**	**CG_9**	**CG_10**
core	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
tore	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
cool	✓	✓	✓	✓	✓	✓	plɒs	✓	✓	✓
tool	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
kip	✓	✓	✓	✓	✓	✓	pɪp	✓	✓	✓
tip	✓	✓	✓	✓	✓	✓	kɪp	twɪp	✓	✓
cop	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
top	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
can	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
tan	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
keep	✓	✓	✓	✓	ti:p	✓	✓	✓	tʃi:p	✓
team	✓	✓	✓	✓	✓	✓	✓	✓	di:m	✓
Substituting [t] for /k/	0	0	0	0	0	0	1/6	0	0	0

4.2 Articulatory study results of /aka/ & /ata/ context speech material

The present study conducted t-test results, KTmax and, direct visual inspection of UTI splines of /aCa/ context speech material data for both SG and CG participants.

4.2.1 t-test results of /aka/ & /ata/ context speech material data

In this section, we examine /t/ and /k/ tongue splines of /aka/ & /ata/ context speech material data results. Table 12 presents the t-test results of /aka/ & /ata/ speech material data for SG and CG.

Table 12

Consonant differences calculated for each participant and total consonant differences.

Participant	$p \leq 0.001$ along seven fan-lines?	Maximum adjacent fan-lines with $p \leq 0.01$	Maximum adjacent fan-lines with $p \leq 0.05$
SG_1	Yes	9	15
SG_2	No	0	3
SG_3	Yes	17	21
SG_4	Yes	10	16
SG_5	No	5	8
SG_6	No	6	13
CG_1	Yes	17	20
CG_2	Yes	16	18
CG_3	Yes	25	25
CG_4	Yes	18	20
CG_5	Yes	17	17
CG_6	Yes	15	16
CG_7	Yes	10	12
CG_8	Yes	15	17
CG_9	Yes	32	33
CG_10	Yes	12	13

Note: $p \leq 0.001$ for positive difference in radial distances along a minimum seven adjacent fan-lines was the criteria for significant difference.

Overall, SG participants SG_1, SG_3 and SG_4 showed a clear significant difference between the Mean /k/ and Mean /t/ splines with $p \leq 0.001$.

SG_3 showed a significant difference in most of the fan lines except 13^{th}, 14^{th}, 36^{th} and 37^{th} fan-lines which included both positive and negative radial differences. Therefore, SG_3 did not show a covert contrast but rather showed an overt contrast in 10/10 reps of /aka/ & /ata/ or non-words (see Figure 17 and section 4.2.5). Because SG_3 had one token of VF in the minimal pairs or real words (see Table 2 and section 3.6.1.1), it has resulted in this participant being included in SG at the beginning of the present study. This single VF token could have been due to a normal speech error such as an error of target choice instead of an error of speech production.

Therefore, although grouped into SG, participant SG_3 in fact showed characteristics typical of the CG participants. This shows that participant grouping criteria should reconsidered to include only those persistent Velar Fronters in CG by excluding those who indicate a single VF token from CG. Furthermore, SG_3 is the only participant who displayed similar a significancy to the CG participants. The main reason could be that SG_3 did not substitute [t] for /k/ in any of the ten repetitions perceptually (see Table 4). **As a result of these findings, SG_3 will be grouped into CG and discussed as a CG participant from here on**.

Of the remaining three SG participants, SG_2 and SG_5 did not show a significant difference in their p-values between the Mean /k/ and Mean /t/ tongue curves to meet the minimum requirement of seven adjacent fan lines. Although SG_6 showed a significant difference ($p \leq 0.001$) along a minimum of six adjacent fan-lines, if the p-value criterion was relaxed to $p \leq 0.05$, then SG_6 would show a significant difference along 14 fan-lines. SG_5 would also show a significant difference along 8 fan-lines then. However, for SG_6, the KTmax value was 1.99 mm which is very low. In addition, visual inspection of UTI also showed that this participant's /k/ and /t/ tongue splines look similar. Therefore, the overall conclusion was that SG_6 did not show a covert contrast.

All CG participants' ($N = 11$ with SG_3) p-values and their Mean /k/ and /t/ show a significant difference with a $p < 0.001$ along a minimum of 35 fan lines out of 42. However, which fan lines show this significancy varies from one participant to another. For example, CG_3 showed a significant difference in all the fan lines except 14^{th}, 15^{th} and 41^{st} fan lines while CG_4 showed a significant difference in all fan lines except the 7^{th}, 8^{th}, 31^{st}, 32^{nd} and 33^{rd} fan lines.

4.2.2 KTmax results of /aka/ & /ata/ context speech material data

KTmax gives information about maximum radial difference between the dorsal and coronal separation between /k/ and /t/ tongue splines (Cleland and Scobbie, 2021). The calculated KTmax measures for SG and CG participants are tabulated in Table 13.

Table 13

KTmax results for SG and CG participants for /aka/ & /ata/ context speech material data.

Participant	**KTMax (MM)**
SG_1	5.27
SG_2	0.24
SG_3	9.27
SG_4	0.32
SG_5	1.99
Mean KTmax of SG	1.50
S.D. of SG	3.85
CG_1	9.81
CG_2	17.56
CG_3	16.42
CG_4	15.77
CG_5	9.73
CG_6	9.05
CG_7	8.25
CG_8	11.19
CG_9	18.48
CG_10	8.53
SG_3	12.14
Mean KTmax of CG	11.90
S.D. of CG	4.08

Note: SG_3 who showed the showed the highest KTmax in SG (KTmax = 12.14 mm) is now grouped into CG. SG_3 did not substitute [t] for /k/ in /aka/ & /ata/ speech material data.

Table 14 shows the summary of Fan-line, t-test result and the KTmax data of SG participants.

Table 14

Summary of the KTmax values of SG participants and their representative fan lines and p-value stated at the maximum point of constriction

	SG_1	SG_2	SG_4	SG_5	SG_6
Fan-line	4	7	14	19	4
p-value	0.001	0.8	0.001	0.5	0.001
KTmax	5.27	0.24	9.27	0.32	1.99

Amongst SG, the highest KTmax was with SG_4 (KTmax = 9.27 mm) which was well above the Mean KTmax of 1.75 mm for the SG. The lowest KTmax was from SG_5, having a KTmax of 0.32 mm.

Among the CG participants, the highest KTmax was recorded by CG_2 (KTmax = 17.56 mm) which is well above the CG's Mean KTmax of 12.48 mm. In CG, the lowest KTmax was presented by CG_7, having a KTmax of 8.25 mm.

4.2.3 Broad results for /aka/ & /ata/ context speech material data for SG

This section summarises the average UTI splines of [k] and [t] for SG along with t-test results, KTMax results and phonetic transcriptions for /aka/ and /ata/ in Figure 11.

4.2.4 Summary of SG participants' /aka/ & /ata/ context speech material data and evidence of covert contrast

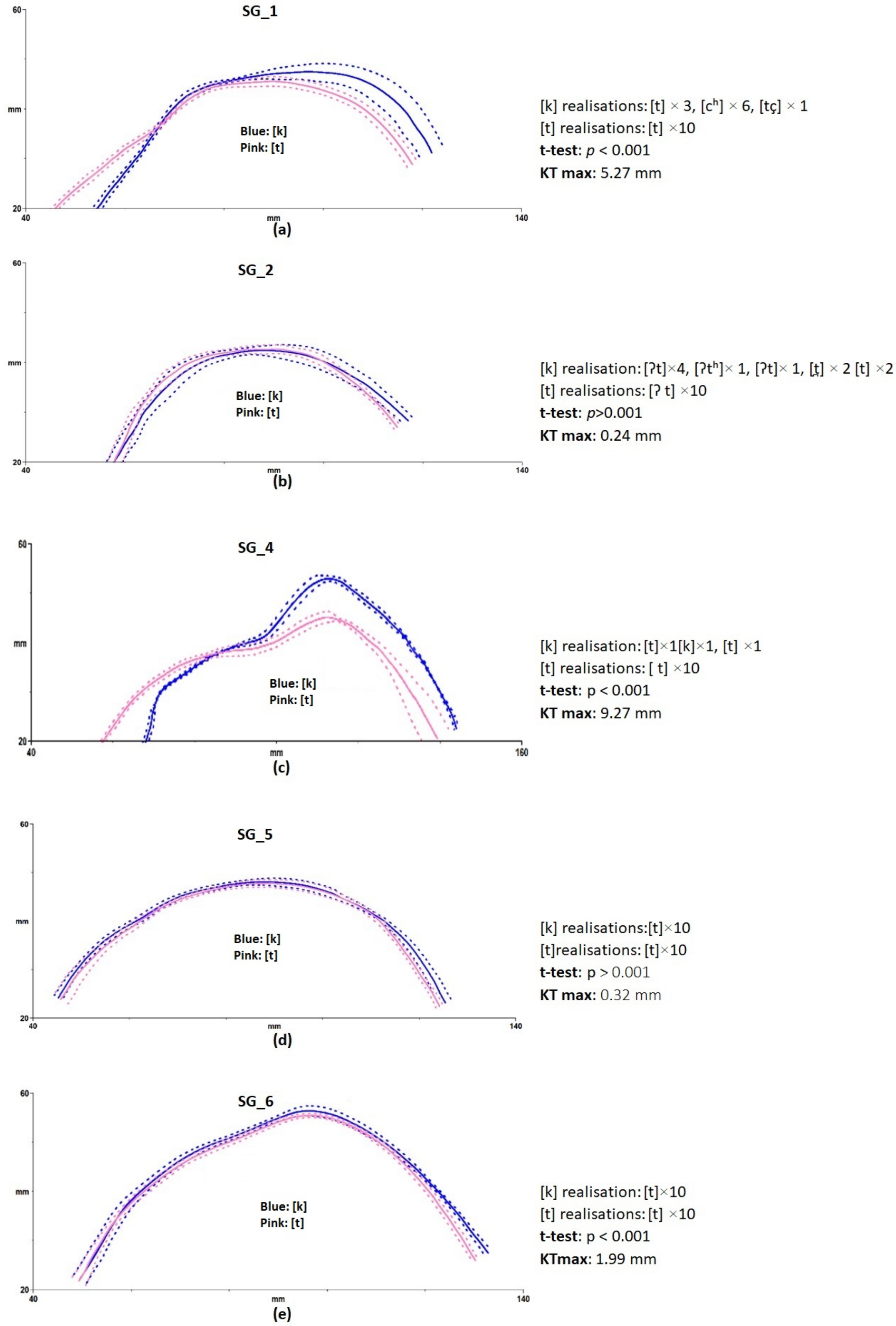
This section discusses the tongue splines for /aka/ & /ata/ context speech material data for SG participants reported in Figure 11 with their KTmax results (see section 4.2.2) and their broad phonetic transcriptions, to conclude whether they showed a covert contrast or not. The SG participants who indicated a significant difference in t-test are discussed first, followed by those who did not show a significant difference.

4.2.4.1 SG_1

The average tongue splines for /k/ and /t/ for SG_1 is shown in Figure 11(a). SG_1's t-test (i.e., $p \leq 0.001$) indicates a significant different in most of the fan lines except

Figure 11

Summary of /aka/ & /ata/ context speech data for (a)SG_1 (b)SG_2 (c)SG_4 (d)SG_5 and (e)SG_6 showing average tongue spline of [k](blue) and [t](pink) (S.D. in dotted lines). Tongue root to the Left. Tongue tip to the Right.



from the 17^{th} to 22^{nd} fan-line.

Although the back of the /k/ and /t/ tongue splines overlaps, we can see that Mean /t/ tongue spline is different to Mean /k/. Therefore, SG_1 looks like backing not fronting. The /t/ tongue spline does not show a typical tongue shape and it is not normal. It looks like /t/ is backed to [k]. But when looking at broad phonetic transcriptions, SG_1's [t] realisations is 10/10, indicating no signs of backing. However, the blade of the tongue is missing to provide more clarity, for a conclusion on SG_1.

SG_1 substituted alveolar /t/ with velar [k] in 3/10 attempts, while in the remaining attempts (7/10) this participant produced them as palato-alveolar [tʃ]. This is shown in Figure 12 and Figure 13.

Figure 12

SG_1's average tongue spline when substituting [t] for /k/ (in 3/10 attempts).

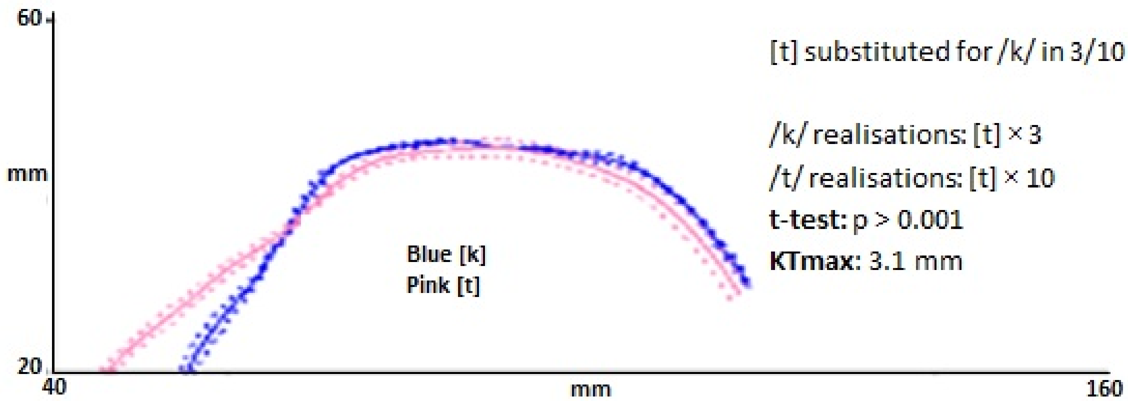
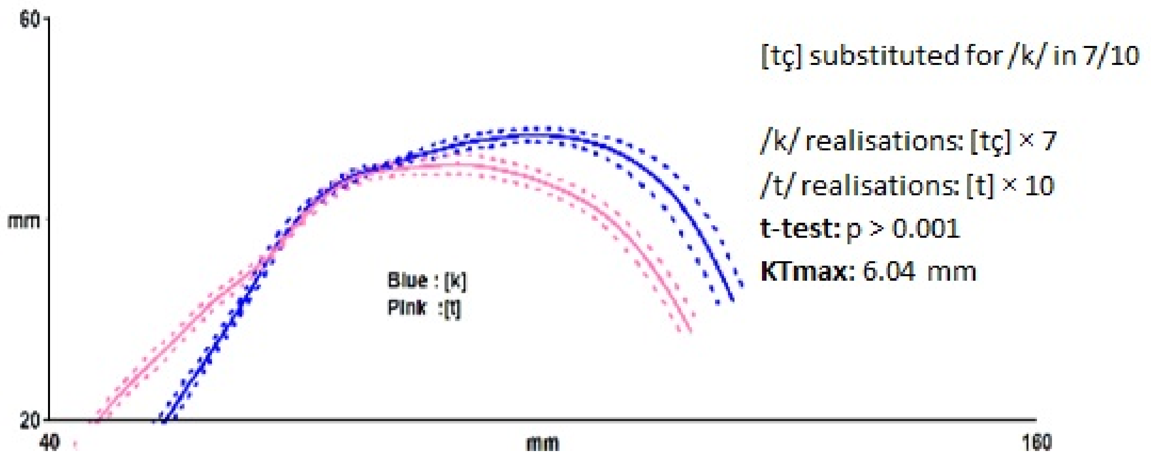


Figure 13

SG_1's average tongue splines when substituting [tʃ] for /k/ (in 7/10 attempts).



Although SG_1, and SG_4 showed a significant difference in their t-tests, existence of a significant difference alone is insufficient to claim a covert contrast exists in them. Significant difference in a t-test is only one factor among others that may indicate a covert contrast. The KTmax measure (see section 4.2.2) can be used in conjunction to corroborate the existence of a covert contrast. In some cases such as SG_1, direct visual inspection of tongue splines was crucial to conclude whether a participant presented a covert contrast as shown in Figure 12 and Figure 13.

The KTmax measure for SG_1 for [t] substitutions for /k/ (3/10 attempts) was 3.1 mm (+/-0.86 mm), while the KTmax when [tʃ] was substituted for /k/ (7/10 attempts) was 6.04 mm (+/-1.2 mm). The Mean KTmax for all ten repetitions of /aka/ & /ata/ (10/10) was 5.27 mm. The three productions where [t] was substituted for /k/ had a lower KTmax compared to when [tʃ] was substituted for /k/. Therefore, seven /k/ attempts (7/10) by SG_1 were palatalised, while for the other three /k/ attempts, this was not the case. We can conclude that, SG_1 did not show a covert contrast. Rather, this participant displayed an unusual production of /k/ that is fronted or distorted to a palatal position.

During direct visual inspection of UTI splines it was observed that, for the three [t] substitutions for /k/ (3/10 attempts), the /k/ tongue spline did not entirely overlap the /t/ tongue spline. This was also evidenced by the lower KTmax measure of 3.1 mm (+/-0.86 mm). It was also observed that, when [t] was substituted for /k/ (3/10 attempts), the tongue spline did not distort to a palatal position, whereas when [tʃ] was substituted for /k/ (7/10) it did distort to a palatal position.

Overall, SG_1's perceptual studies/broad phonetic transcriptions match articulatory UTI data which indicate that no covert error or covert contrast is present, although his [t] tongue spline when he substituted for /k/ (3/10 attempts) showed no complete merger.

4.2.4.2 SG_2

The average tongue splines for /k/ and /t/ for SG_2 is presented in Figure 11(b). Although the t-test revealed $p \leq 0.001$ for a few fan lines (2^{nd} to 3^{rd} fan-lines), it did not satisfy the minimum required consecutive fan-lines to show a significant difference according to the t-test.

From this figure, we can see that both /k/and /t/ have tongue tip elevation with the tongue tip going towards the anterior region of the oral cavity. Generally, the tongue tip and the body of the /k/ and /t/ tongue splines overlap with each other (within +/- 1 S.D.). The tongue shape and the place of articulation for SG_2 in Figure 11(b) show that the Mean /k/ and Mean /t/ tongue spline curves are both similar to alveolar [t].

SG_2 showed a complete merger by entirely overlapping /k/ and /t/ tongue splines which matches the articulatory results with perceptual study results. This is further corroborated by SG_2 showing a near zero KTmax of 0.24 mm at the 7^{th} fan line. Therefore, no covert contrast was evident.

4.2.4.3 SG_4

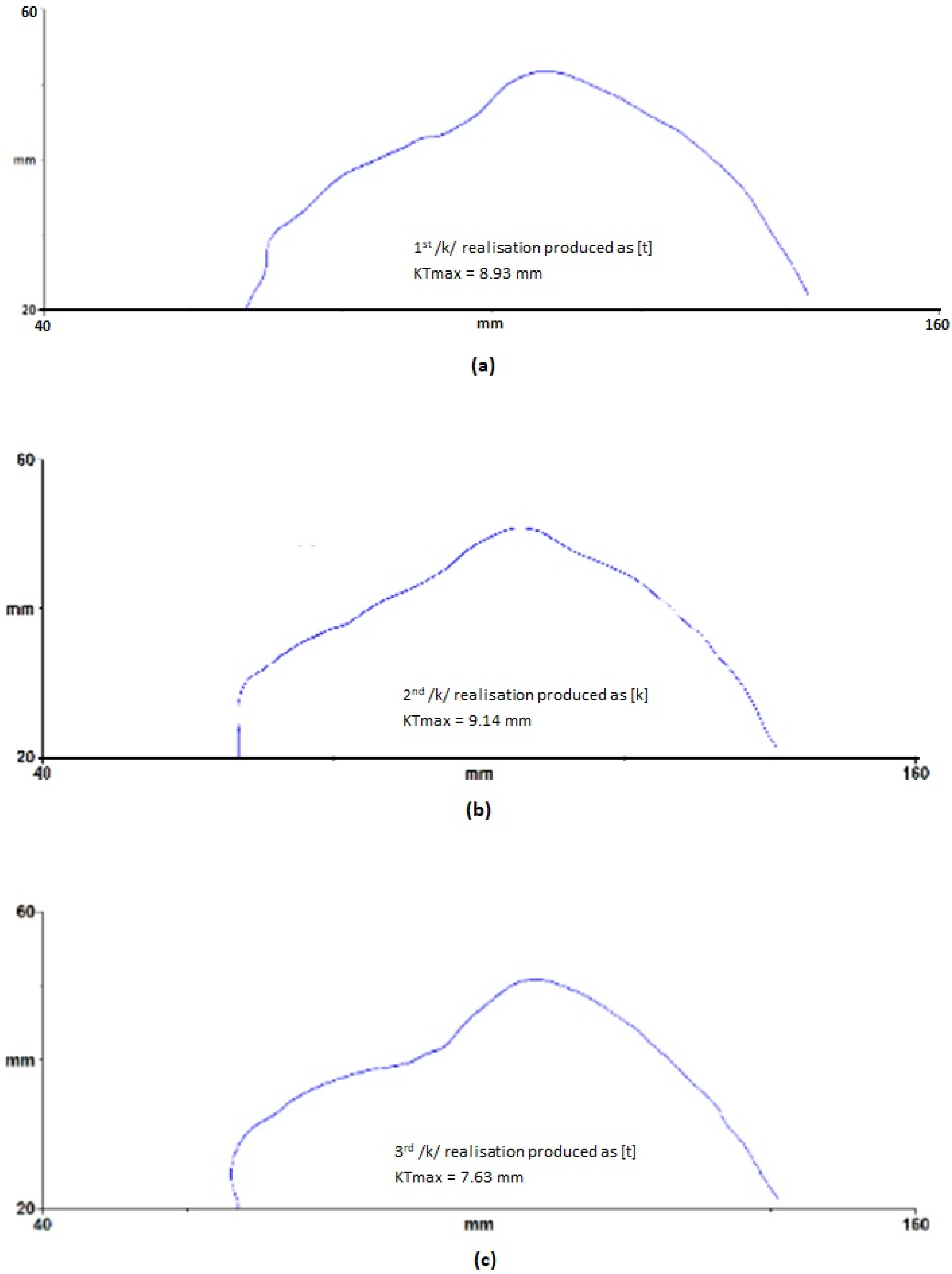
The average tongue splines for /k/ and /t/ for SG_4 is presented in Figure 11(c). Significant difference (i.e., $p \leq 0.001$) was displayed from fan line 11th to 18th. SG_4's broad phonetic transcription sequence of /aka/ context was [t]×1, [k]×1 and [t]×1. SG_4's broad phonetic transcriptions or perceptual study results do not agree with the articulatory results which suggest a covert contrast. SG_4's tongue splines show a significant difference statistically in the t-test. The KTmax measure was 9.27 mm, which is indicative of a measurable difference. Overall, SG_4 also shared similar features such as shape of spline and maximum constriction at the anterior region, which was also slightly different than UTI of the [t] realisations. Therefore, in conclusion, we believe this is indicative of a covert contrast.

SG_4's Mean /k/ and /t/ tongue splines show similarities as well as differences. One similarity is that each spline is directed to the anterior region of the oral cavity, giving a similar tongue shape. The maximum point of contact for Mean /k/ and Mean /t/ also seem to be at a similar place along the horizontal axis, but with different heights along the vertical axis. However, the +/- S.D. curves do not overlap except at the back of the tongue. This is because the Mean /k/ tongue spline is positioned higher than the Mean /t/ spline, and the root part of /t/ is placed further back than that of /k/. However, the anterior /k/ is located further forward than the /t/. Although the shape of Mean /t/ spline is similar to a typical [t] tongue spline, the Mean /k/ spline seems problematic (see 4.2.4.2 for further details).

It can also be argued that SG_4 does not show covert contrast since it is variable. This is because SG_4's [k] realisations only have 3 repetitions of /aka/ context and one of them was produced correctly as [k] as shown in Figure 14.

Figure 14

(a)1st (b)2nd and (c)3rd /aka/ realisations of by SG_4.



It has been said that children can show less differentiation between tip and dorsum gestures, and greater variability between multiple repetitions (Cleland and Scobbie, 2021). Therefore, when calculating KTmax for only those repetitions where SG_4 substituted [t] for /k/ (2/3 attempts), the KTmax was 9.27 mm with a significant difference also shown in the respective t-test (see Figure 14).

By looking at both quantitative and qualitative articulatory results of SG_4, it can be assumed that all /k/ realisations from SG_4 have similar characteristics. Firstly, the Mean S.D. (for KTmax) for each /k/ and /t/ realisations by SG_4 is less than 1 mm. A low S.D. suggest that all three tongue splines were clustered around the Mean spline. Secondly, the KTmax measure difference between each /k/ realisation of SG_4 was less than 2mm. The first attempt of SG_4 was recorded as a [t] in the broad phonetic transcription. The corresponding KTmax was 8.93 mm, while the second attempt was perceived as [k] and the KTmax was 9.14 mm. The third attempt was perceived as [t] and the KTmax was 7.63 mm. Finally, each tongue spline of SG_4 also shared similar features such as shape of spline and maximum constriction at the anterior region, which was also slightly different than UTI of the [t] realisations. However, the third [k] realisation used as [t] tongue shape is slightly different to other two attempts. Therefore, we can conclude that this is not indicative of a covert contrast in SG_1 but infact, is indicative of a variable.

4.2.4.4 SG_5

The average tongue splines for /k/ and /t/ for SG_5 is presented in Figure 11(d). Tongue tip elevation for both Mean /k/ and Mean /t/ tongue splines is similar, with the tongue tip pointing towards the alveolar ridge. It is evident that the tongue tip and the body of the tongue +/- S.D. curves overlap for both /k/ and /t/. There is no significant difference shown by the t-test with $p > 0.001$.

Overall, both the perceptual study results and the articulatory study give similar results for this speaker. The t-test results did not show a significant difference between the tongue splines. The KTmax measured was 0.32 mm. UTI inspection showed that the degree of constriction, the tongue shape and the place of articulation for /k/ and /t/ were similar. Therefore, SG_5's velar /k/ spline has a similar tongue shape to alveolar /t/, thereby showing a complete merger.

4.2.4.5 SG_6

The average tongue splines for /k/ and /t/ for SG_6 is shown in Figure 11(e). Both Mean /t/ and /k/ tongue splines have tongue tip elevation towards the alveolar ridge. The tongue tip and the tongue body of the tongue's +/- S.D. curves overlap. Therefore, considering the degree of constriction we can see that their place of articulations is similar and both have similar tongue shapes to alveolar [t] rather than to velar [k]. SG_6 showed a p-value < 0.001 in six adjacent fan lines from 3rd fan line to 9^{th} fan line.

Overall, both the perceptual study and articulatory study give similar results. The t-test results did not show statistically significant difference. The KTmax of SG_6 was 1.99 mm. UTI inspection shows that the place of articulation for /k/ and /t/ are similar and that SG_6 show a complete merger.

4.2.5 Summary of CG participants' results in /aka/ & /ata/ speech material data

Figure 15 and Figure 16 show the average tongue shapes for /k/ (in blue) and /t/ (in pink) for CG participants including SG_3.

Figure 15

Average tongue splines of [k] (blue) and [t] (pink), with S.D. in dotted lines for (a)CG_1, (b)CG_2, (c)CG_3 and (d)CG_4.

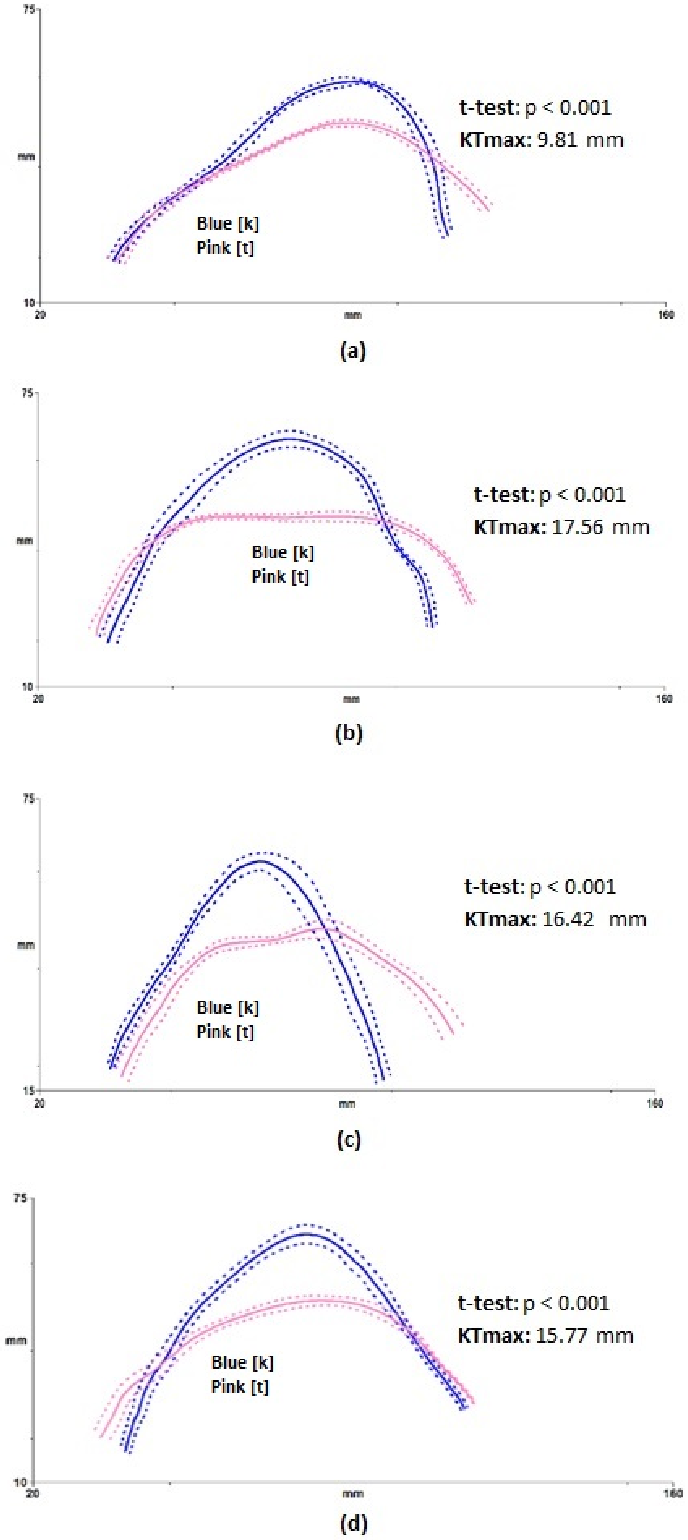
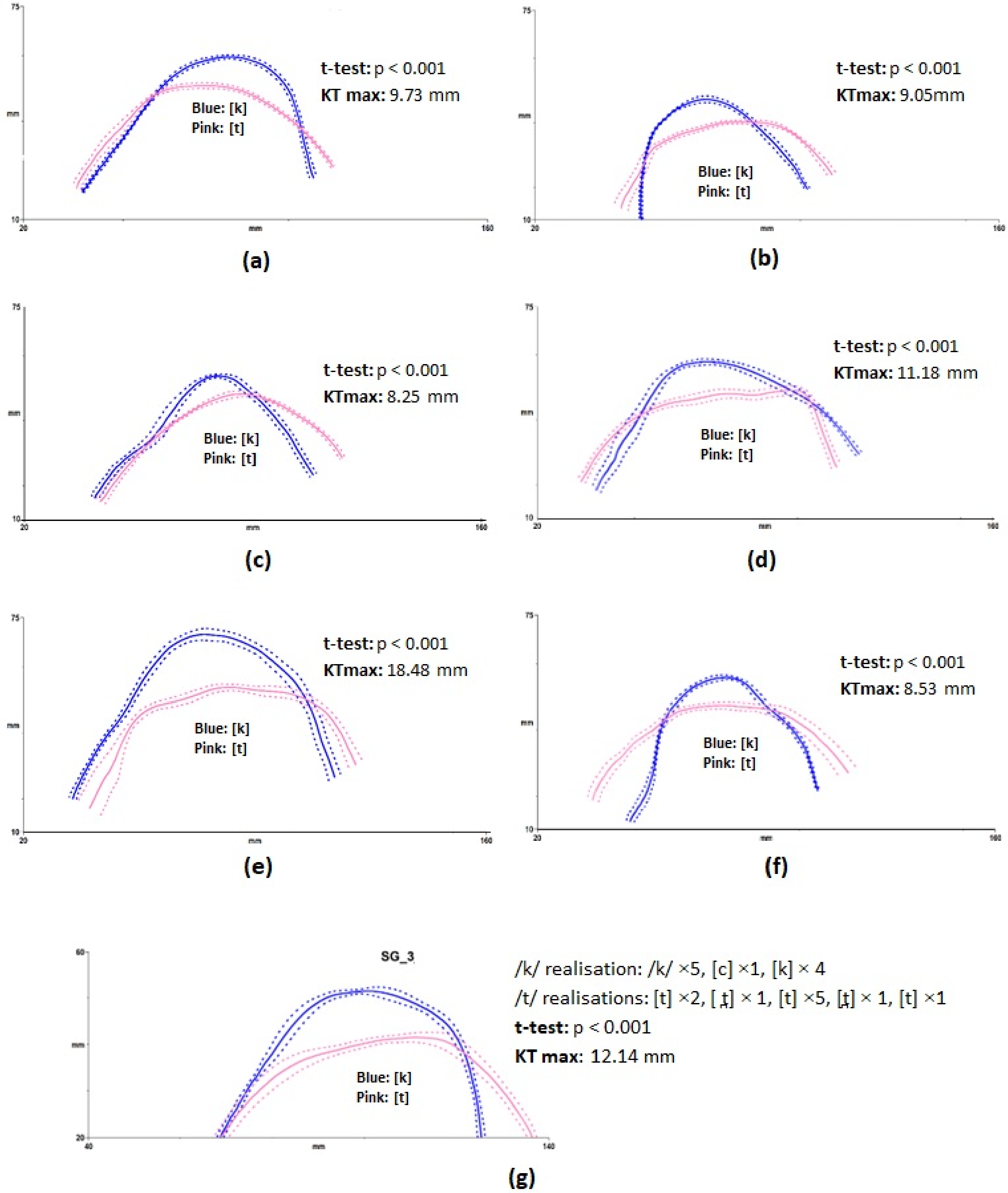


Figure 16

Average tongue splines of [k] (blue) and [t] (pink), with S.D. in dotted lines for (a)CG_5, (b)CG_6, (c)CG_7, (d)CG_8, (e)CG_9, (f)CG_10 and (g)SG_3.

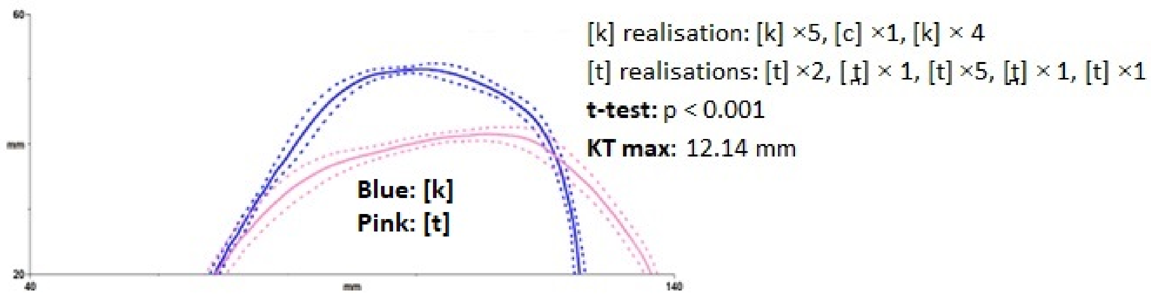


When observing CG participants' Mean /k/ and Mean /t/ tongue splines, we can see that in CG_3 and CG_9 the root of /t/ spline does not cross over the /k/ spline at the back of the tongue. We can also see that in all CG participants, their /t/ tongue tip is directed to the anterior region of the oral cavity. The tongue body of the Mean /k/ tongue spline (in blue) of all CG participants is raised. Furthermore, their /k/ and /t/ splines have different shapes. This is evidenced by the fact that their +/- S.D. curves do not overlap.

The average tongue splines for /k/ and /t/ for SG_3 is presented in Figure 17.

Figure 17

SG_3's average tongue splines of [k] (blue) and [t] (pink) with S.D. in dotted lines.



The p-value is significant in most of the fan lines except for the 27^{th} to 26^{th} and from 13^{th} to 14^{th}. We can see that SG_3's Mean /t/ tongue spline is directed to the anterior region of the oral cavity. During the production of the velar /k/, the Mean tongue spline shows posterior tongue movement associated with tongue body elevation. Furthermore, the +/-S.D. curves of /k/ and /t/ do not overlap with each other. This suggests that SG_3 may have different /k/ and /t/ Mean tongue splines. When looking at their degree of constriction we can see that the place of articulation is different. The maximum point of constriction for /k/ is more towards the posterior region, and for /t/ this is towards the anterior region. Therefore, the Mean tongue spline of /t/ is a typical alveolar [t], while the Mean tongue spline of /k/ is a typical velar [k], indicating an overt contrast.

SG_3 produced [k] in 9/10 attempts while producing /t/ correctly in 6/10 attempts. Therefore, SG_3's /t/ and /k/ tongue splines have unique alveolar and velar features in their respective tongue splines, thereby showing an overt contrast. Furthermore, SG_3's broad phonetic transcriptions and perceptual study results matched the UTI based articulatory study results. Therefore, SG_3 did not demonstrate evidence of a covert contrast but showed an overt contrast instead. Hence, SG_3 had been correctly grouped into CG.

When looking at the point of constriction, tongue shape and place of articulation, we can see that /k/ and /t/ are different to each other in all CG participants. The tongue spline of /k/ has a raised tongue dorsum and the tongue root and is similar in shape to a typical velar [k] than a typical alveolar [t]. The average tongue spline of /t/ in the CG participants have elevation in the tongue tip region that represent a typical alveolar [t].

All CG participants' ($n = 11$ with SG_3) p-values and their Mean /k/ and /t/ showed a significant difference along a minimum of 35 fan lines out of the reasonable maximum fan-lines available for each participant's dataset. It is unreasonable to expect a participant to display $p > 0.001$ in 42/42 fan-lines, since there will be no tongue data for some anterior or posterior fan-lines at all. For some others there will be no confident tracking. These minimum 35 fan lines include those with positive (i.e., /k/-/t/ positive) and negative (i.e., /t/-/k/ positive) radial differences resulting in $p \leq 0.001$. Therefore, the CG participants showed a significancy in their t-test.

All CG participants also had KTmax values near zero, suggesting a significant difference between their /k/ and /t/ tongue shapes. The CG's Mean KTmax was 11.90 mm with a maximum KTmax of 18.48 mm and a minimum KTmax was 8.25 mm for /aka/ & /ata/ context (see section 4.2.2). Furthermore, direct visual inspection of tongue splines also revealed that their /k/ and /t/ tongue splines were distinct. Their broad phonetic transcriptions and the perceptual study results are also consistent with these articulatory study results, suggesting an overt contrast.

4.3 Articulatory study results of t/k singleton minimal pairs

The articulatory study of t/k singleton minimal pairs consisted of conducting the t-test, calculating KTmax and visually inspecting UTI splines of the SG participants only. Table 15 presents the t-test results of t/k singleton minimal pairs for SG. Table 16 presents the KTmax results for SG.

4.3.1 t-test results of t/k singleton minimal pairs for SG

The SG participants' t-test results are summarised in Table 15.

Table 15

t-test results of t/k singleton minimal pairs for SG.

Participant	p-value	Showed significant difference (i.e., $p < 0.001$)?	Showed significant difference (i.e., $p < 0.05$)?
SG_1	0.5	No	No
SG_2	0.8	No	No
SG_4	0.02	No	Yes
SG_5	0.05	No	Yes
SG_6	0.5	No	No

None of the SG participants showed a significant difference in their t-test ($p \leq 0.001$) for t/k singleton minimal pairs along seven adjacent fan lines, which was the minimum requirement used for the present study. However, when the t-test criteria was relaxed to $p \leq 0.05$, then both SG_4 and SG_5 showed a significant difference. SG_4 in particular, showed $p = 0.02$ which is a clear indication of a significant difference in this case.

4.3.2 KTmax results of t/k singleton minimal pairs for SG

The KTmax results of t/k singleton minimal pairs for SG are summarised in Table 16.

Table 16

KTmax results of t/k singleton minimal pairs for SG.

Participant	**KTmax (mm)**
SG_1	0.86
SG_2	0.33
SG_4	4.33
SG_5	0.99
SG_6	0.84
Mean	2.02
S.D.	1.99

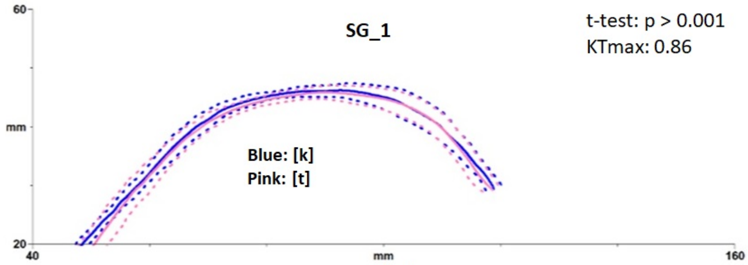
The highest KTmax was shown by SG_4 (4.33 mm). The remaining participants in SG showed a KTmax of less than a 1 mm.

4.3.3 Average tongue splines for t/k singleton minimal pairs and word level transcriptions

Figure 18 presents the average tongue shapes of SG participants for t/k singleton minimal pairs along with their respective t-test result, KTMax and the word level transcriptions of t/k singleton minimal pairs.

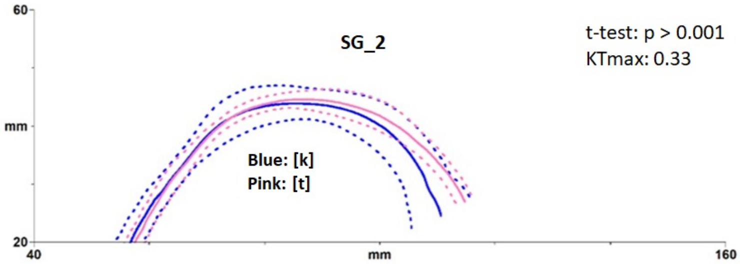
Figure 18

Summary of t/k singleton minimal pairs for (a)SG_1 (b)SG_2 (c)SG_4 (d)SG_5 and (e)SG_6.



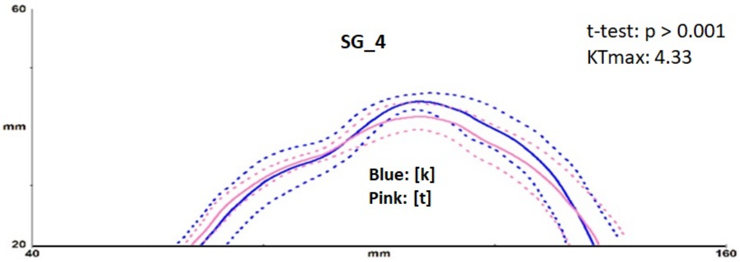
/k/	/t/
core - tɔ:	tore - tɔ:
cool - tu:l	tool - tu:l
kip - tʃɪp	tip - tʃɪp
cop - tɒp	top - tɒp
can - tan	tan - tan
keep - ti:p	team - ti:m

(a)



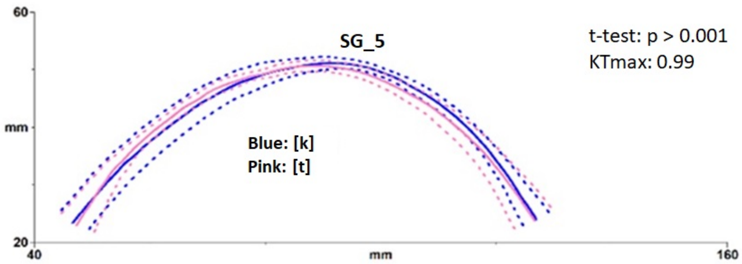
/k/	/t/
core - tɔ:	tore - tɔ:
cool - kɪu:l	tool - tu:l
kip - tɪp	tip - tʃɪp
cop - tɒp	top - tɒp
can - tan	tan - tan
keep - ti:p	team - tʃi:m

(b)



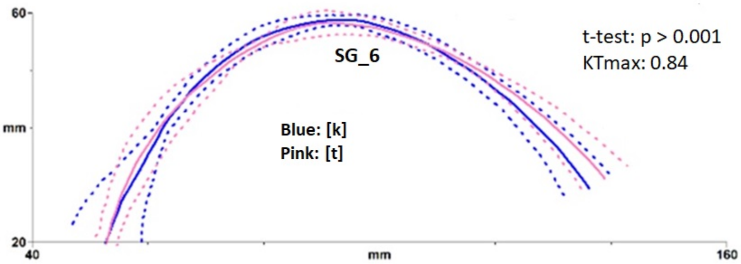
/k/	/t/
core - tɔ:	tore - tʃɔ:
cool - tʃu:l	tool - tu:l
kip - tʃɪp	tip - tɪp
cop - tɒp	top - tɒp
can - tʃan	tan - tan
keep - ki:p	team - ti:m

(c)



/k/	/t/
core - tɔ:	tore - tɔ:
cool - tu:l	tool - tu:l
kip - tɪp	tip - tɪp
cop - tɒp	top - tɒp
can - tan	tan - tan
keep - ti:p	team - ti:m

(d)



/k/	/t/
core - tɔ:	tore - tɔ:
cool - tu:l	tool - tu:l
kip - tʃɪp	tip - tɪp
cop - tɒp	top - tɒp
can - tan	tan - tan
keep - ki:p	team - ti:m

(e)

4.3.4 Summary of t/k singleton minimal pair analysis for SG

SG_1, SG_2, SG_5 and SG_6 showed complete mergers in t/k singleton minimal pairs, compared to the remaining participant, SG_4 (see Figure 18). Furthermore, direct visual inspection of their tongue splines also revealed that their /k/ and /t/ tongue splines overlapped with each other.

In t-tests with the $p \leq 0.001$ criterion, none of the SG participants showed a significant difference in their p-values. However, when $p \leq 0.05$ was used as the criteria, SG_4 and SG_5 both indicated a significant difference between their /k/ and /t/ tongue splines.

When computing KTmax, four participants (i.e., SG_1, SG_2, SG_5 and S_6) presented a KTmax measure of less than 1 mm. The highest KTmax measure was shown by SG_4 (4.33 mm, S.D. 1.99 mm). SG_4 showed a significant difference between /k/ and /t/ tongue splines by way of KTmax. However, the broad phonetic transcriptions of SG_1 and SG_4 showed no covert contrast.

Therefore, by combining the above observations from the articulatory study, we can conclude that SG participants with the exception of SG_4, presented complete mergers in t/k singleton minimal pairs. If a different criterion of $p \leq 0.05$ was used for the t-test, combined with KTmax, it might suggest that SG_4 showed a covert contrast in t/k singleton minimal pairs during articulatory analysis.

When looking at SG_4's t/k singleton minimal pairs, according to broad phonetic transcriptions of t/k singleton minimal pairs, [t] was substituted with /k/ only twice in core → [tɔ:], cop → [tɒp]. The KTmax measure when substituting a [t] for /k/ was 4.18 mm. During the other four /k/ productions, /k/ was substituted by a post alveolar affricate [tʃ] three times (e.g., cool → [tʃu:l], can → [tʃan], kip → [tʃip]). During these SG_4's KTmax measure was 5.04 mm. The other /k/ production was correct (e.g., keep → [ki:p]) and its KTmax measure was 8.35 mm (see section 4.2.5). Furthermore, the t-test indicated no significance during those two substitution patterns. Since the KTmax measure [tʃ] for/k/ was higher (5.04 mm) than [t] for /k/ (4.18 mm), it suggests that [tʃ] was palatalised. This was further evidenced when directly inspecting SG_4's tongue splines. Overall, SG_4's neither showed a complete merger or a covert contrast during t/k singleton minimal pairs.

Chapter 5

Discussion

The present study aimed to replicate and extend the Cleland et al. (2017) study (see section 2.5.2). The aim of both studies was the same: to investigate whether school-aged children with SSDs of unknown origin and persistent velar fronting, show covert contrasts at the articulatory level, suggesting an articulatory rather than phonological cause to their SSD. Both studies were UTI based articulatory studies, which investigated tongue movements of /t/ and /k/ articulations and aimed to determine whether covert contrast can be found in children who perceptually neutralise this distinction, using qualitative and quantitative measures.

This discussion will be centred around answering each Research Question 1 to 3 introduced in section 2.7, followed by an overall discussion, limitations, clinical implications, future work and finally a conclusion.

5.1 Research Question 1

Does using a different group of children with persistent velar fronting and different speech materials lead to the same finding that there is no evidence of covert contrast between /t/ and /k/ in ultrasound data in line with Cleland et al. (2017) study?

Although the participants were different in the present study, their demographic data (e.g., age, diagnosis) were similar to participants in Cleland et al. (2017) study. We hypothesised that when using the same speech material data (i.e., t/k singleton minimal pairs), most children with persistent VF would not show a covert contrast, indicated by no statistically significant difference in their t-tests between tongue shapes for /k/ and /t/ attempts at the point of maximum constriction, in line with Cleland et al. (2017). In addition, the present study also used non-lexical words that comprise of open vowel contexts: VCV (e.g., /ata/ & /aka/) that were not used by Cleland et al. (2017). In doing so, we attempted to find out whether we would be able to replicate the same results as in Cleland et al. (2017). This newly introduced speech material data resulted in one participant in SG presenting with a covert contrast, which was not in line with the hypothesis.

This suggests that they were persistent velar fronters (see section 2.5.2) with persistent SSDs and highly consistent VF, who had no reported structural or underlying difficulties. In the present study, the participants were aged between 6;3 to 11;2 years and presented with functional SSDs along with velar fronting difficulties, substituting alveolar [t] for velar /k/ (see section 3.6.1.1). Therefore, in both studies the participants were past 5;0 years, which is the age reported by McLeod and Crowe (2018) as the expected age for typically developing children to acquire velars during their phonological development process (see section 2.1.3.2). This suggests that they were persistent velar fronters (see 2.5.2). However, one participant (SG_4) in the present study was identified as having swallowing difficulties by the clinicians. SG participants who were perceptually persistent velar fronters.

Cleland et al. (2017) used t-test and direct visual inspection of /k/ and /t/ tongue splines to identify covert contrast. In addition, the present study also used an additional measure (i.e., KTmax introduced by Cleland and Scobbie (2021)) to identify covert contrast.

5.1.1 t/k singleton minimal pairs speech material data

In both studies, the sample word list included t/k singleton minimal pairs (or near minimal pairs) in a range of vowel contexts as shown in Table 17. In Cleland et al. (2017), the word lists were gathered from two types of intervention studies: Ultrax 1 (Ultrax, nd) and Ultraphonix (Cleland et al., 2015a; Ultrax, 2017).

Table 17

t/k singleton minimal pairs used by clinicians in Ultrax 1 (Ultrax, nd), Ultraphonix (Cleland et al., 2015a; Ultrax, 2017) and the present study.

Ultrax	Ultraphonix	Present Study
tore/core	keep/team	core/tore
table/cable	cape/tape	cool/tool
tie/Kai*	Ken/ten	kip/tip
toot/coot	cab/tab	cop/top
	cop/top	can/tan
	core/tore	keep/team
	cool/tool	
	cub/tub	
	kip/tip	
	coy/toy	

***Note:** Kai is a first name in the UK.

The present study used participants' speech material data from ULTRAX2020 project (Ultrax, 2017). The original word lists used by the researchers involved: a general word list and a target specific word list. The general word list was used in Ultrax (2017) for all children. It covered all lingual consonants to identify specific areas of difficulty in a participant. Next the target-specific wordlist was designed to probe specific errors in children who required U-VBF (Ultrax, 2017). Since not all children would benefit from U-VBF, such an assessment was deemed useful in case a different intervention other than U-VBF was then required. Therefore, the t/k singleton minimal pairs word list used in the present study is a subset of this target-specific word list used in (Ultrax, 2017).

As can be seen from Table 17, the t/k singleton minimal pairs used were different from the targeted words used by Cleland et al. (2017). Therefore, while both studies shared similar type of speech material data (i.e., single word level /t/ and /k/ minimal pairs), each study used different targeted words.

In both previous studies the speech material data used were a longer word list of single productions of many words containing velars and alveolars, including minimal pairs. In the present study, the t/k singleton minimal pairs were aided with written sentences on the computer screen enabling participants to read themselves (see Appendix A). However, the SLT prompted reading word by word once to the SG participants who attempted to imitate them, similar to the study in Cleland et al. (2017).

5.1.1.1 Comparison of t-test results

In the present study, none of SG participants (n = 5, without SG_3) who were persistent velar fronters, showed a significant difference in their t-test for the t/k singleton minimal pairs. This was measured along a minimum of seven adjacent fan lines (i.e., 16.7% or 7/42 fan-lines), which was the minimum requirement criteria with respect to adjacent fan-lines, used in the present study.

It can be argued that some SG participants did not show a significant difference in their t-test due to two very different but typical tongue splines for [t] and [k] crossing over at one fan line. In this case, the radial difference measured from the origin of the fan-line would be near zero, resulting in no statistical difference during the t-test. The use of a t-test to measure a significant difference is therefore not suitable where the tongue splines cross over, as discussed by Cleland and Scobbie (2021).

The participants who did not show a significant difference in seven adjacent fan lines in /aka/ & /ata/ context were SG_2, SG_5 and SG_6. Among them SG_2 and SG_5 had both anterior and posterior crossover points in their tongue splines. However, SG_6 showed only an anterior crossover point. However, this did not affect the determination of a significant difference in SG_6 because this participant's KTmax (1.99 mm) was on the 4th fan line which was in the anterior side of the tongue. When looking at the posterior side of the tongue, none of the (k-t) or (t-k) differences exceeded a KTmax of 1.99 mm.

5.1.1.2 Comparison of tongue splines using direct visual inspection

In the Cleland et al. (2017), all the persistent velar fronters produced both /t/ and /k/ with near identical tongue shapes or complete merges in t/k singleton minimal pairs, with no evidence of covert contrast found along a minimum of six adjacent fan lines, which was the criteria adopted for that study. However, it was noted that one participant, namely Rachel, showed unusual tongue shapes such as retroflexes, using direct visual inspection. In the present study, tongue splines during production of t/k minimal pairs among 5/6 participants were similar to the results reported in Cleland et al. (2017), indicating complete mergers. The tongue splines of the remaining participant, namely SG_4, neither overlapped completely, nor show a covert contrast in the t-test.

5.1.2 /aka/ & /ata/ context (non-lexical) speech material data

Cleland et al. (2017) did not conduct a UTI study for non-lexical speech material data. Therefore, the present study's additional speech material data (i.e., repetitions of /aka/ & /ata/), will be compared with the results of the t/k singleton minimal pairs in Cleland et al. (2017) study.

SG_1, showed neither a covert contrast nor an overt contrast (see section 4.2.4.1). Rather, SG_1 showed an unusual production of /k/ that is fronted or distorted to a palatal position. In the perceptual studies carried out, SG_1 did not show perceptual neutralisation in 7/10 attempts of /t/ (e.g., [t] ×3, [c^h]×6, [tç]×1) (see Table 3). In SG_1's broad phonetic transcription, [t] for /k/ was substituted in 3/10 attempts. When inspecting the UTI for SG_1, it was observed that three /t/ productions were substituted with [k]. As shown in Figure 12 they were perceptual mergers.

Cleland et al. (2017) in their study noted that one participant, namely Rachel, showed unusual tongue shapes such as retroflexes, when UTI of /k/ was inspected with direct visual gestalt. They called it a covert error which is also an indication of articulatory difficulty. Covert errors are those errors visible in articulatory data, but sound 'correct' or perceptually neutralised. Therefore, SG_1 did not show a covert error unlike Rachel in Cleland et al. (2017) study. It was instead a distortion error with an overt contrast since /t/ and /k/ sounded different in the perceptual study.

On the other hand, SG_4 showed a covert contrast in /aka/ & /ata/ speech material data (see section 5.5). According to Gibbon (1999) ULG typically occur during productions of lingual consonant targets and are characterized by contact that lacks clear differentiation between the tongue apex, tongue body and lateral margins of the tongue. SG_4 showed less tongue-palate contact in UTI during direct offline visual inspection. However, the lateral margins of the tongue could not be viewed in the present study because the UTI were recorded in midsagittal plane. Therefore, without such evidence we cannot deduce that SG_4 did not present ULG.

Gibbon and Lee (2017) stated that EPG can be used to identify covert contrasts with

SSDs, using visual inspection of the raw data (e.g., tongue palate contact). Then objective quantification using numerical indices such as variability, place of articulation and tongue palate contact can be used to define increased variability, as different tongue shapes are formed for different attempts of repetition. In a similar manner, SG_4 showed slightly different tongue shapes for three repetitions of /aka/ (see Figure 14), but the place of articulation remained the same. However, SG_4 did not demonstrate covert contrast in t/k singleton minimal pairs or real words. Therefore, SG_4 could likely be in the process of resolving his velar fronting difficulties as suggested by Cleland and Scobbie (2021) who termed this as gradual gradient acquisition of velars (see section 2.5.3).

SG_4 showing a covert contrast in non-lexical speech material data but not in t/k singleton minimal pairs or real words could suggest that repetitions of non-lexical speech material data is easier to produce than imitations of t/k singleton minimal pairs. We can assume that due to the simple structure of the non-lexical repetitions, they are easier to produce than real single words. Showing a covert contrast in non-lexical speech material data may be step in the gradual gradient acquisition suggested by Cleland and Scobbie (2021). Munson et al. (2010) have also stated that, before children produce a contrast between two sounds, they may produce a covert contrast.

This evidence of a covert contrast reveals more about the articulatory pathway than the phonological pathway because tongue movements were viewed from a phonetic perspective. According to Levelt (1999) and Levelt et al. (1999) imitations do not require activation of lexical concepts or lemma selection. Instead, it starts at phonological encoding through the process of auditory word perception. Then syllabification is performed on the segment string that has been activated. In the present study t/k singleton minimal pairs were proceeded as imitations, because all the participants imitated the targeted word once, after the SLT demonstrated each word verbally. It has been said that model repetitions use a different path compared to production of words more related to cognitive systems (Levelt, 1999; Levelt et al., 1999).

Finally, compared to Cleland et al. (2017), the present study also used other types of non-lexical speech material data (e.g., repetitions of /aka/ & /ata/) which have a syllabification pattern of vCv, comprising of one consonant with two vowels. These have a very short token length and contain a high frequency context. It has been stated that vowels have a higher acoustic energy and that they are easier to perceive than consonants as they do not reach to semantic systems to retrieve the words (Ladefoged and Disner, 2012). Therefore, these speech material data are more related to passive perception.

5.1.3 /aka/ & /ata speech material data vs t/k singleton minimal pairs

SG_4 substituted [t] for /k/ in 2/6 attempts (33.3%) in t/k singleton minimal pairs, while in /aka/ he substituted [t] for /k/ in 2/3 attempts (66%) in the perceptual study results. This could indicate that percentage of perceptual neuralisation (e.g., substituting [t] for /k/) could be lower in t/k singleton minimal pairs compared to /aka/ speech material data. Due to the small sample size used in the present study, we cannot conclude that non-lexical speech material data is more likely to show covert contrast than word level speech material data of t/k singleton minimal pairs.

To investigate this claim further, we calculated KTmax and t-test only for [t] substitutions for /k/ by SG_4. When they were separately calculated, for t/k singleton minimal pairs, the KTmax for 2/6 attempts (e.g., core→[tɔː], cop→[tɒp]) was 4.18 mm and the t-test was non-significant. With regards to /aka/ speech material data, in 2/3 attempts where SG_4 substituted [t] for /k, the calculated KTmax was 8.23 mm and 7.63 mm with the t-test also showing a significant difference. This may suggest that for the occasions where SG_4 substituted [t] for /k/ do not show a covert contrast.

Study by Byun et al. (2016a) is the only ultrasound study that discovered evidence of a covert contrast in repetitions of t/k singleton minimal pairs, in a preschool child having positional velar fronting. This is in contrast to the findings of Cleland et al. (2017) and the present study which also used t/k singleton minimal pairs or in real words. The targeted t/k minimal pair words used in the present study were: cup-tub and key-T (letter), which were used to elicit the targets in initial position and back-bat and bug-mud, which were used to elicit the contrast in the final position. Although the present study and Byun et al. (2016a) both used minimal pairs, there were differences in the process of collecting the targeted words.

In the present study, the target words were recorded after the SLT demonstrated them verbally to the participants and then the participants repeated them. But Byun et al. (2016a) conducted the following protocol: first, to familiarise the targeted eight words practice sessions were conducted. During these practice sessions, stimulus images, prompts, feedback and producing each word multiple times by repeating it until the children succeeded, were carried out. In the exact recording environment, during the experimental task the child was seated in a chair facing a laptop screen that displayed the stimulus images, then the child produced each word in a single block of five consecutive trials. Verbal models were not used when participants produced the t/k minimal pair words, because the child had practiced naming the stimulus images in the practice sessions (Byun et al., 2016a).

The Schema Theory of motor learning introduced by Schmidt and Lee (2005) states that motor learning as a production of rapid discrete movements, involves units of actions (motor programs) that are retrieved from memory and then adapted to a particular situation. Therefore, the method by adopted by Byun et al. (2016a) prior to data collection using t/k minimal pair words, followed principles of the motor learning theory. This contrasts with the data collection method used for t/k minimal pair words in the present study, which found no evidence of a covert contrast among SG participants.

As shown in Figure 1, the psycholinguistic framework has two routes/pathways: The nonword repetition route or the word repetition route. In the non-word repetition route, there is a need to create a new motor program due to the lack of stored representations for non-words. The word repetition route, however, starts with 'semantic representations' and does not require any input skills. In this pathway, a stored motor program is accessed (Nathan, 2001). Therefore, we can conclude that existence of a covert contrast among the perceptually persistent velar fronters could depend on the speech material data type used, since the two speech material data types used in the present study represent different pathways in the psycholinguistic framework by (Stackhouse and Wells, 1997).

Word repetition, like non-word repetition, requires input processing. In addition, it will generally require access to a stored phonological representation and a stored motor program. However, it is possible for a word to be processed through a non-lexical route.

5.2 Research Question 2

Do children with an overt contrast (CG) show a difference at the articulatory level between /t/ and /k/ in /aka/ & /ata/ speech material data compared to the SG?

One of the aims of the present study was to extend it by adding a CG, to quantify the size of any difference between Mean tongue splines of /k/ and /t/ (see section 2.7). We hypothesised that CG participants would show a consistent, significant, difference in t-test in most of the fan lines, higher KTmax for /k/ and /t/ attempts, and /t/ and /k/ tongue spline shapes which are different upon visual inspection. Overall, we expected that CG would show an overt contrast between /k/ and /t/ tongue splines.

All CG participants' ($n = 11$ with SG_3) p-values and their Mean /k/ and /t/ show a significant difference with $p \leq 0.001$ along a minimum of 35 fan lines out of 42. All CG participants showed a significancy in their t-test, while in SG only two participants (SG_1 and SG_4) showed a significant difference by way of t-test. Furthermore, CG have shown a significant difference in t-test in more adjacent fan-lines compared to these two SG participants.

Participants in CG presented a mix of disorders (see Table 1). Among CG, 4/10 participants were diagnosed with articulation disorders by the clinicians. Another 3/10 participants were diagnosed with phonological disorders, while one participant (1/10) was diagnosed with a motor speech disorder. The remaining participants in CG (2/10) were not given specific diagnoses. Although nearly half of the CG participants were diagnosed with SSDs with articulation difficulties, they did not show a covert contrast in /aka/ & /ata/ speech material data. This is mainly because they were not persistent velar fronters, although they were shown to produce other speech error types. Therefore, as hypothesized all CG participants demonstrated distinct Mean /k/ and /t/ tongue spline shapes (i.e., overt contrast), while none of the SG participants showed a covert contrast.

5.3 Research Question 3

Do children with persistent SSDs without VF (i.e., CG) and those with SSDs and VF (i.e., SG) fall within the norms for typically developing (TD) children's KTmax range?

5.3.1 KTmax of CG vs KTmax norms of TD children

Previously, Cleland and Scobbie (2021) reported KTmax for 30 TD children aged between 5;8 to 12;10 years. The reported KTmax for /aka/ & /ata/ context speech material was 11.9 mm with S.D. = 3.0 mm (Cleland and Scobbie, 2021). Similarly, in the present study, CG (n = 11 with SG_3) showed a similar Mean KTmax of 11.90 mm ($8.25mm <$ KTmax < 18.48 mm, S.D. = 4.08 mm) for /aka/ & /ata/ context (see section 4.2.2). All CG participants with overt contrast had a KTmax measure which fell within the typically developing children's KTmax range reported by Cleland and Scobbie (2021). Although CG participants were diagnosed with SSDs, they displayed similar results to TD children's KTmax range in /aka/ & /ata/ context.

The reason for displaying similar KTmax norms to TD children's despite having SSDs, is that the CG participants did not have any difficulties producing target alveolar /t/ and velar /k/ in ten repetitions of /aka/ & /ata/, according to both their broad phonetic transcriptions and the articulatory analysis. Furthermore, visual inspection of their UTI splines also revealed that all the CG participants were able to produce correct velar /k/ and alveolar /t/ similar to TD children's /k/ and /t/ productions.

5.3.2 KTmax of SG vs KTmax norms of TD children

In the present study, SG participants who showed complete mergers (i.e., SG_2, SG_5 and SG_6) had Mean KTmax values ranging from 0.24 to 1.99 (see section 4.2.2), which was outside the reported KTmax range for TD children in Cleland and Scobbie (2021). Furthermore, SG_1's KTmax measure (5.27 mm) did neither belong to the range of KTmax of SG participants who showed complete mergers, nor did it fall within the TD children's KTmax range. However, the one participant who showed a covert contrast in the SG, namely SG_4, had a KTmax value of 8.45 mm which fell within TD children's range reported in Cleland and Scobbie (2021).

The Mean KTmax of SG (n = 5) was 4.47 mm (S.D. = 3.85 mm). Therefore, compared to Mean KTmax norms for TD children (11.9 mm), the present study's participants who substituted [t] for /k/ (i.e., SG) had a Mean KTmax well below this level.

Overall, perceptually persistent velar fronters showed covert contrasts or perceptual mergers, and their KTmax range was in the intermediate stage in participants with a covert contrast, and near zero in participants with a complete merger. In contrast, CG participants who did not substitute [t] for /k/, showed an overt contrast with higher Mean KTmax in /aka/ & /ata/ context speech material data. As hypothesised, children with an overt contrast (i.e., CG) had KTmax values within the normal (TD) range, whereas SG participants' KTmax values were below the normal range.[1]

The next section compares these KTmax values with those reported in other studies in literature.

5.4 Comparison with covert contrast reported in related work

According to Cleland and Scobbie (2021), all five participants with persistent velar fronting in Cleland et al. (2017) who substituted [t] for /k/ had Mean KTmax values of near zero (1.29 mm, S.D. = 0.94 mm).This suggests that there was a complete merger of [k] and [t] for t/k singleton minimal pairs. In comparison, the Mean KTmax for t/k singleton minimal pairs in the present study was 2.02 mm (S.D. = 1.99 mm) (see Table 15). This Mean KTmax is within one S.D. of the Mean KTmax measure reported for TD participants in Cleland et al. (2017) study, by Cleland and Scobbie (2021). Therefore, the KTmax of /aka/ & /ata/ speech material data (Mean = 2.02 mm (w/o SG_3)) for the SG participants in the present study is comparatively higher than that for t/k singleton pairs of the present study as well as the KTmax of participants in the Cleland et al. (2017) study (see Table 13).

Overall, in both studies Velar Fronters' Mean KTmax values were well below that of TD children who acquired velar and alveolar stops with no other SSDs (Mean = 11.9 mm, S.D. = 3.03 mm) (Cleland and Scobbie, 2021). Therefore, we can conclude that regardless of the type of speech material used, perceptually persistent velar fronters with SSDs have Mean KTmax which is lower than that of TD children's /aka/ & /ata/ speech material data.

The reasons for the present study's SG showing a higher Mean KTmax than participants in Cleland et al. (2017), can be explained by the fact that SG in the present study included SG_4 who showed a covert contrast with a higher KTmax (9.27 mm).

The use of t/k singleton minimal pairs in the present study as well as the study by Cleland et al. (2017) did not reveal a covert contrast by any participant. The speech material data used in both studies were related to automatic cognitive systems. This is similar to the results in Byun et al. (2016a). However, Byun et al. (2016a) found

[1]KTmax for t/k singleton minimal pairs was measured for only SG. There are no corresponding KTmax for CG except for SG_3. All SG participants showed complete mergers except SG_4.

a covert contrast in repetitions of singleton minimal pairs (see section 2.5.4). They reported using minimal pairs which were recorded as repetitions, with the SLT making verbal prompts.

A possible reason for Byun et al. (2016a) showing a covert contrast at word level could be due to this different protocol used in collecting speech material data. In Byun et al. (2016a), the participants were given stimulus and practice sessions that may have made the task easier for the participants on the day of recoding. Furthermore, Byun et al. (2016a) stated that participants Max (3;9 years) and Rory (4;1 years) were receiving treatment for speech sound errors including VF. Max, who was the participant who demonstrated a covert contrast, presented positional VF in initial position, which has a significant effect on vowel context as claimed by Cleland et al. (2017). In both the present study and in Cleland et al. (2017), the participants have received treatment targeting production of velars prior to the study, but were unresponsive to treatment. However, none of the participants received therapy via VBF method.

In Byun et al. (2016a) the participant who showed a covert contrast was 4;1 year old with SSD. In both the present study and in Cleland et al. (2017) the participants were perceptually persistent velar fronters. SG_4 who showed a covert contrast was aged 9;2 years while in Cleland et al. (2017) the participant Rachel who showed a covert error was aged 8;8 years. Therefore, in Byun et al. (2016a) the participant who showed a covert contrast was younger and was at the borderline of the TD children's phonological development process. Byun et al. (2016a) found evidence of covert contrast in both Ultrasound measures and in acoustic studies. The measure used in the Ultrasound study was DEI, which directly compared the tongue shapes (see section 2.4.1.4.2).

Although Cleland et al. (2017) did not find evidence of a covert contrast, one participant showed a covert error at word level speech material data. Therefore, the present study is the only study that did not show covertness in single word level speech material data compared to Cleland et al. (2017) and Byun et al. (2016a).

In related work, Baum and McNutt (1990) investigated /s/ fronting among participants divided into two SG (5-6 years and 7-8 years) and two CG (5-6 years and 7-8 years). This articulatory analysis assessed the duration, intensity (i.e., amplitude) and spectral (i.e., centroid frequency) parameters of /s/ and /θ/ in several contexts. The two SG comprised of misarticulating children and the two CG comprised of normally articulating children. Their reported results showed that all four groups distinguished /s/ and /θ/ to some extent on all parameters, revealing a covert contrast among the misarticulating group (i.e., SG). The study found out that misarticulated /s/ productions were lower in centroid-frequency than normally articulated /s/ in the spectral analysis. Therefore, acoustic analysis involving temporal and spectral parameters may be useful in identifying a covert contrast. It also revealed that the older group of children in SG showed similar results to the younger group in SG. This may indicate a delay in the older group's development of fine (adult-like) phonetic control resulting in them functioning at a less mature stage in their speech sound development (Baum and McNutt, 1990).

Maxwell and Weismer (1982) investigated voicing contrast for word-initial obstruents in a child who showed a functional misarticulation characterized by a severe lack of contrasts. While the perceptual analysis suggested that the child represented a large number of voiced and voiceless obstruents with [d], the acoustic analysis revealed that the participant produced a covert contrast. After undergoing speech therapy, the participant produced acceptable voicing distinction, at least for stops. Furthermore, this revealed that the child had delayed articulation skills attainment, which was not deviant (Maxwell and Weismer, 1982).

Forrest et al. (1990) used spectral moments and VOT to investigate the acoustic characteristics of voiceless velar and alveolar stop consonants among normally articulating (CG) and phonologically disordered children (SG). The perceptual analysis indicated that SG substituted /t/ for /k/. They developed a discriminant function based on Mean (first moment), skewness (third moment) and kurtosis (fourth moment) derived from the first 40 ms of the VOT interval to investigate covert contrast between velar and alveolar stops (Forrest et al., 1990). The study revealed a velar-alveolar covert contrast in one SG participant who then underwent therapy. Phonological analysis of the child's speech after therapy, then revealed target appropriate productions of both /t/ and /k/. By contrast, the other three SG participants (i.e., phonologically disordered children), who did not show a covert contrast in the acoustic analysis, did not evidence any a contrast after treatment.

5.5 Discussion on SG_4

5.5.1 Differential diagnosis of SG_4

SG_4 is a 9;2 years old male who has been diagnosed by the clinicians as having inconsistent SSD and swallowing difficulties (see Table 1). According to the overall results of perceptual study and articulatory studies, SG_4 showed a covert contrast in /ata/ & /aka/, but not in t/k singleton minimal pairs.

Speech error types that SG_4 presented in non-lexical speech material data (/aCa/) were: velar fronting [k→t] , [ŋ→ɲ], [ŋ→n], palatal fronting [ʃ→s], double articulations [n→n͡ɲ,n͡d], distortions [r→ɹ̝̊], voicing errors [s→s̬], stopping [θ→d] and some unclassified error types such as [r→z] and [j→l], [t͡ʃ→l t͡ʃ]. Therefore, SG_4 presented both developmental speech errors and disordered speech error types (see Table 6).

According to Bishop et al. (2014) a phonological impairment is an impaired ability to learn the speech-sound contrasts that discriminate words and the constraints that govern how those sounds can be combined. Because SG_4 produced the same sound production both correctly and incorrectly during attempts of /aCa/ (e.g. [k→[k]/[t], [r→[ɹ̝̊]/[z̥], [n]→[n͡ɲ]/[n͡d], [j]→[j]/[l], [t͡ʃ]→[t͡ʃ]/[lt͡ʃ]), it shows that SG_4 perceives and contrasts speech sounds in non-lexical speech material data albeit in an inconsistent manner. Similarly, in t/k singleton minimal pairs too SG_4 showed similar inconsistency in productions (e.g., core→tɔ:, keep→ki:p) (see Table 10). This, along with the fact that SG_4 showed a covert contrast in /aka/ context in UTI data analysis, may suggest that SG_4 has a motor/articulatory related SSD rather than a SSD with a phonological cause, as also mentioned in Cleland et al. (2017).

This claim is further supported by the results of the perceptual study. SG_4 is the only participant who displayed r-distortions [r→ɹ̝̊] amongst the Velar Fronters (i.e., SG) in 1/3 attempts in the perceptual studies. It has been said that, distortions appear to result from leftovers from an earlier speech delay, which were originally substitutions or omissions, which migrated closer to normal before becoming distortions (Flipsen Jr., 2015). Preston et al. (2019) claimed that production of /r/ results in differentiated movement of the anterior and posterior parts of the tongue. This results in multiple curves (i.e., "bends") of the tongue. In contrast, distorted /r/ production results in a near lack of differentiation among distinct parts of the tongue, thereby resulting in a "rainbow shape" of the tongue from back to front in the sagittal section (Gick et al., 2007; Klein et al., 2013; Preston et al., 2019).

Furthermore, SG_4's Mean /k/ tongue spline in /aka/, where SG_4 showed a covert contrast, showed a steep tongue-palate contact. This suggested that SG_4 has an inflexible tongue spline shape, which does not curve properly in a smooth manner (see Figure 14). This also suggests that SG_4 may be having difficulties in manipulating tongue movements properly, possibly because of motoric difficulty. Moreover, to support with the fact that SG_4 has an articulatory related SSD is that he showed disordered speech error types (Preston et al., 2019).

It has been said that, an articulation difficulty with an unknown origin exists when a child has difficulty in producing sounds due to a difficulty in the proper placement of articulators to create the target sound (Flint et al., 2020). This may be due to incoordination of movements of the lips, tongue, teeth, palate or lungs. Cleland et al. (2017) also found retroflexion among a child called Rachel and referred to it as an articulatory difficulty of an unknown origin. The clinicians of ULTRAX 2020 however, had diagnosed SG_4 as having inconsistent speech sound disorder as well as swallowing difficulties. it has been said that articulation difficulties of unknown origin do not have known neurological difficulties, unlike swallowing difficulties van den Engel-Hoek et al. (2015). Furthermore, it has also been said that children with articulation difficulties of unknown origin can't make that sound in any context/word position (e.g., beginning, middle or end of the word) or in a single position such as at the end of a word Feldman and Messick (2016). SG_4 however, produced /k/ correctly in 1/3 attempts in /aka/ and in 2/6 attempts in t/k singleton minimal pairs. Although SG_4 showed disordered speech error types (distortions, retroflexion, double error types as previously mentioned), which are features of an articulatory SSD of unknown origin, SG_4 also present symptoms that does not belongs to articulatory SSD of unknown origin as above mentioned.

It is also useful to compare SG_4 with SG_1 in the perceptual and articulatory results to investigate whether SG_4 has articulation difficulties related to unknown origin or known (e.g., motor speech disorders). SG_1 was diagnosed with motor speech disorder by the clinicians in the ULTRAX 2020 (see Table 1). Both SG_1 and SG_4 showed higher KTmax in /aka/ & /ata/ speech material compared to t/k singleton minimal pairs. Furthermore, both participants have also shown ”moderate to severe” PCC rating (see section 4.1.2). With regards to the speech error types, both participants share similar kind of speech errors such as velar fronting (see Table 2), double articulation and retroflexion (see Table 6). Therefore, both participants showed disordered SSDs. Both participants also showed common types of substitutions in /aka/ & /ata/ and t/k singleton minimal pairs.

It can be stated that, overall SG_4 presented a motor speech disorder than an articulation difficulty related to the unknown origin. Dysarthria and childhood apraxia of speech (CAS) are motor speech disorders derived mainly due to neurological difficulties. Children with dysarthria have weak muscle tone and those with CAS have difficulties with motor planning (the plan and act of moving muscle (unknown, 2020). CAS may occur as a result of known/unknown neurological impairment, in association with complex neurobehavioral disorders speech sound disorder (ASHA, 2019).

As both childhood Dysarthria and CAS are related to motor speech disorders, other related symptoms (e.g., swallowing difficult) can be used to find the best matched diagnosis for the SSD in SG_4.

The clinicians have recorded SG_4 as having inconsistent speech sound difficulties with swallowing difficulties. Generally, children with CAS do not have swallowing difficulties unless there is an oral Apraxia. Speech inconsistency however, is a core feature of CAS (Iuzzini-Seigel et al., 2017). Children with Dysarthria can be accompanied by swallowing difficulties which is related to the involuntary system (van den Engel-Hoek et al., 2015).

It has been said that, swallowing difficulties can occur among children who have Dysarthria due to muscle weakness. According to Potter et al. (2019), children with Dysarthria do not show inconsistent speech errors, despite imprecise articulation due to muscle weakness and in coordination with a limited range of movements.

In conclusion, although it was difficult to determine which type of motor speech disorder (e.g., CAS or childhood Dysarthria) SG_4 presented, because SG_4 showed symptoms of both CAS and Dysarthria, we can however suggest that SG_4 presented a SSD with a motoric cause.

5.6 Clinical implications

The present study suggests that covert contrast can vary in the same speaker, depending on the speech material type. Therefore, it is important when clinicians are assessing children with persistent SSDs that they consider using different types of stimuli as corresponds to different paths of the psycholinguistic framework (see Figure 1).

The present study also revealed that children with persistent velar fronting with motor/articulatory difficulties, can show neither covert contrast nor a complete merger. For these children distortions were evident in the phonetic transcription and KTMax was lower than expected for an overt contrast. This may be due to gradient improvement in the acquisition of velars as suggested by Cleland and Scobbie (2021). Therefore, the existence of such a considerable KTmax measure, even when they do not show a covert contrast in UTI, may be a prognostic indicator of a resolving difficulty. Alternatively, it may suggest that articulation therapy is needed to achieve more perceptually correct productions.

Furthermore, in both the perceptual and articulatory analysis, some children showed palatalisation when producing /k/. For example, when looking at broad phonetic transcriptions of t/k singleton minimal pairs of real words, SG_4 substituted [t] for /k/ only twice (e.g., core→[tɔ:], cop→[tɒp]), while /k/ was substituted by a post alveolar affricate [tʃ] three times (e.g., cool→[tʃu;l], can→[tʃan], kip→[tʃip]). This shows that in SG_4 the number of palatalised [tʃ] was higher than [t].

According to Keating and Lahiri (2005), palatalization of velars is equivalent to fronting phonetically. However, neither fronted nor palatalized velars are like palatal consonants. Therefore, looking at the number of palatalisations can be useful as a symptom related to articulatory difficulties, when diagnosing children with persistent VF with articulatory difficulties. Clinicians in such cases can assess whether palatalisation is part of the gradient development for [k] productions.

Another finding of the present study was that children can show small KTmax differences in their tongue shapes during different attempts of the same /k/ production. Despite this, SG_4 had the same place of articulation in each attempt. This variability speaks to a motor cause warranting interventions based on the principals of motor learning.

A final implication is that, as shown by Maxwell and Weismer (1982) and Forrest et al. (1990), children who showed evidence of a covert contrast who then underwent therapy acquired target productions better than children who did not show a covert contrast. Therefore, clinicians can look for evidence of a covert contrast in their assessments before customising a targeted therapy program.

5.7 Limitations

The following limitations in the present study have been identified:

- Instrumental limitations with UTI (poor tongue images due to synchronisation problems, poor positioning of the probe).
- Possibility of inaccurate speech material data due to using a low number of repetitions (e.g., SG_4 only had three repetitions of /aka/ & /ata/).
- The lack of quantitative (t-test, KTmax) and qualitative measures (Ultrasound tongue image splines) for t/k singleton minimal pairs for CG participants due to limited time availability. Therefore, only tongue splines between the SG and CG participants' in /aka/ & /ata/ context speech material data were compared.
- PCC was not calculated, and speech error types not identified for t/k singleton minimal pairs. Not having speech error analysis for t/k singleton minimal pairs makes it difficult to compare speech error analysis vs non-lexical speech error analysis.

5.8 Future Work

5.8.1 Increasing sample size

The present study consisted of only six participants in SG ($n = 5$) and ten participants in CG ($n = 11$ with SG_3) with SG having lesser participants than CG. Having a lesser number of participants presents a risk of having biased data. For example, only one participant out of six in SG demonstrated covert contrast, that provided articulatory evidence for children with persistent velar fronting. A higher sample size for SG would have given data that is more representative of that group. Having a larger number of participants as well as having equal numbers in SG and CG reduces the risk of accidentally having outliers, or groups with bias.

5.8.2 Conducting a longitudinal study

In the present study most SG participants (e.g., SG_2, SG_5 and SG_6) showed complete mergers in UTI splines, with near KTmax measures and there was no significant difference in p-values in both speech material data (i.e., non-lexical and single words). By conducting the same study longitudinally, we expect these participants to improve their scores in line with the range of KTmax of TD children.

Therefore, we hypothesized that a categorical shift from KTmax of near zero to KTmax within the normal range for TD children would suggest a phonological basis to the error, as investigated in Cleland and Scobbie (2021). However, a gradual or variable change in KTmax would align with the view of Gibbon (1990) and Byun (2012) that VF may be due to motor constraints.

Cleland and Scobbie (2021) conducted a longitudinal study involving participants who demonstrated gradual gradient acquisition of velars during the intervention, with the intervention of speech therapy. Their findings suggest an articulatory/motor cause to persistent velar fronting. Therefore, by extending the present study longitudinally we can attempt to explore whether the present participants presented motor/articulatory related difficulty. Furthermore, conducting the same study longitudinally would reveal whether real word productions substituting [t] for /k/ in broad phonetic transcriptions can also show a covert contrast. This is because real words productions which involve top-to-bottom processing are difficult to produce compared to productions that use the bottom-up processing method. However, it is unclear whether an articulatory based method like UTI study, which mainly looks at articulatory gestures, could reveal phonological SSDs. Therefore, further research should be carried out along both pathways of the psycholinguistic framework as a longitudinal study to reveal answers to these questions.

5.8.3 A study combining with all motor related SSDs

Similar studies could be conducted to compare children with CAS childhood dysarthria and children with articulatory difficulties with unknown origin to determine how covertness differs according to the diagnosis.

5.8.4 Conducting the same study using both types of t-tests

The study used a high significance level or lower p-value (i.e., $p \leq 0.001$) to identify covert contrast among participants. The intention of using a stricter p-value criterion was to reject results that have happened due to chance, when attempting to identify a significant difference. Although this would not be a problem for participants who showed an overt contrast or the CG participants, employing such a lower p-value criterion could have resulted in failing to identify covert contrasts among other SG participants.

When using $p \leq 0.001$, SG_4 showed a covert contrast in /ata/ & /aka/ (non-verbal), but not in t/k singleton minimal pairs (verbal). However, in speech material data where SG_4 showed a covert contrast (e.g., /aka/ & /ata/) the participant's KTmax (KTmax = 9.27 mm) was nearly double the KTmax of t/k singleton minimal pairs (KTmax = 4.33 mm). However, it could be argued that conducting the same study using a significance level of $p \leq 0.05$ could have given different results for SG_4. When comparing SG_4's UTI splines visually (see section 4.2.4.2) without the aid of a statistical analysis such as a t-test, we can see that there is a critical difference between /t/ and /k/ tongue splines of /aka/ & /ata/ and also a slight difference between tongue splines also for t/k singleton minimal pairs, which did not indicate a covert contrast using a t-test with $p \leq 0.001$ criterion. Therefore, conducting the same study with $p \leq 0.05$ is recommended as future work.

Chapter 6

Conclusion

In summary, by looking at two previous studies Cleland et al. (2017), Byun et al. (2016b) and the present study, we can conclude that different factors such as type of speech material data, speech material data collection method, speech material data processing method, age of participants and other related diagnoses (e.g., swallowing difficulties) of the participants all may have led to different results. Furthermore, we can see that speech material data and speech material data processing methods, which are more related to either articulatory plans or automatise cognitive systems, may have influenced the production or detection of a covert contrast in the present study.

The aim of the study (see section 2.7) was to investigate whether school-aged children with SSDs of unknown origin and persistent velar fronting show a covert contrast at the articulatory level, suggesting an articulatory rather than phonological cause to their SSD. The present study, where one participant (SG_4) showed a covert contrast in non-lexical speech material data (e.g., /aka/ & /ata/ context speech material data), which is more related to the articulatory route, proves that this covert contrast can exist in some children. This finding in an articulatory UTI study, may suggest that the existence of a covert contrast is evidence of a phonetic cause in some children.

However, it is questionable whether SG_4 presented with an articulatory difficulty of unknown origin or a motor speech disorder such as CAS or childhood dysarthria originating from a neurological problem.

The overall conclusion is that children who are perceptually persistent velar fronters with articulatory difficulties can show covert contrasts and these can be measured using UTI. The overall conclusion is that children with persistent velar fronters with articulatory difficulties can show covert contrasts and these can be measured using UTI.

References

AAA (2012). Articulate assistant advanced user guide: Version 2.14.

Aggarwal, G., Gochhayat, S. P., and Singh, L. (2021). Chapter 10 - parameterization techniques for automatic speech recognition system. In Singh, K. K., Elhoseny, M., Singh, A., and Elngar, A. A., editors, *Machine Learning and the Internet of Medical Things in Healthcare*, pages 209–250. Academic Press.

Alighieri, C., Bettens, K., Bruneel, L., Sseremba, D., Musasizi, D., Ojok, I., and Van Lierde, K. (2020). Comparison of motor-phonetic versus phonetic-phonological speech therapy approaches in patients with a cleft (lip and) palate: a study in uganda. *International Journal of Pediatric Otorhinolaryngology*, 131:109849.

AllAboutLinguistics.com (n.d.). Articulatory phonetics.

Anderson, C. (2022). *Essentials of linguistics*. eCampusOntario, 2 edition.

ASHA (2019). Speech sound disorders.

Ashby, M. (2006). *Phonetic Classification*, pages 364–372.

Ashby, M. (2009). Sounds of the world's languages (vowels and consonants).

Baum, S. R. and McNutt, J. C. (1990). An acoustic analysis of frontal misarticulation of /s/ in children. *Journal of Phonetics*, 18(1):51–63.

Bernthal, J. E. and Bankson, N. W. (1997). *Articulation and Phonological Disorders*. Pearson, Upper Saddle River, NJ, 4 edition.

Bernthal, J. E., Bankson, N. W., and Flipsen, P. (2016). *Articulation and phonological disorders*. Pearson, Upper Saddle River, NJ, 8 edition.

Bishop, D. V. M., Nation, K., and Patterson, K. (2014). When words fail us: insights into language processing from developmental and acquired disorders. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1634):20120403.

Broomfield, J. and Dodd, B. (2011). Is speech and language therapy effective for children with primary speech and language impairment? report of a randomized control trial. *International Journal of Language and Communication Disorders*, 46(6):628–640.

Buhr, R. D. (1980). The emergence of vowels in an infant. *Journal of Speech, Language, and Hearing Research*, 23(1):73–94.

Burton, P., Deng, J., McDonald, D., and Fewtrell, M. S. (2013). Real-time 3D ultrasound imaging of infant tongue movements during breast-feeding. *Early Hum. Dev.*, 89(9):635–641.

Byun, T. M. (2012). Positional velar fronting: An updated articulatory account. *Journal of Child Language*, 39(5):1043–1076.

Byun, T. M., Buchwald, A., and Mizoguchi, A. (2016a). Covert contrast in velar fronting: An acoustic and ultrasound study. *Clinical Linguistics & Phonetics*, 30(3-5):249–276. PMID: 26325303.

Byun, T. M., Buchwald, A., and Mizoguchi, A. (2016b). Covert contrast in velar fronting: An acoustic and ultrasound study. *Clinical Linguistics & Phonetics*, 30(3-5):249–276. PMID: 26325303.

Byun, T. M. A. and Tessier, A.-M. (2016). Motor influences on grammar in an emergentist model of phonology. *Language and Linguistics Compass*, 10(9):431–452.

Carmona, M. B., Ángel Rivière Gómez, and González, J. M. I. (1992). Psicología del lenguaje: investigación y teoría. *Psicolingüística*, pages 439–449.

Cleland, J., Lloyd, S., Campbell, L., Crampin, L., Palo, J.-P., Sugden, E., Wrench, A., and Zharkova, N. (2020). The impact of real-time articulatory information on phonetic transcription: Ultrasound-aided transcription in cleft lip and palate speech. *Folia Phoniatr. Logop.*, 72(2):120–130.

Cleland, J. and Preston, J. L. (2020). Biofeedback interventions. In Williams, A., McLeod, S., and McCauley, R., editors, *Interventions for Speech Sound Disorders in Children*, number 2nd, pages 573–599. Paul Brookes, Baltimore.

Cleland, J. and Scobbie, J. M. (2021). The dorsal differentiation of velar from alveolar stops in typically developing children and children with persistent velar fronting. *Journal of Speech, Language, and Hearing Research*, 64(6S):2347–2362.

Cleland, J., Scobbie, J. M., Heyde, C., Roxburgh, Z., and Wrench, A. A. (2017). Covert contrast and covert errors in persistent velar fronting. *Clinical Linguistics & Phonetics*, 31(1):35–55. PMID: 27610938.

Cleland, J., Scobbie, J. M., Roxburgh, Z., and Heyde, C. (2015a). The ultraphonix project: Ultrasound visual biofeedback for heterogeneous persistent speech sound disorders.

Cleland, J., Scobbie, J. M., and Wrench, A. A. (2015b). Using ultrasound visual biofeedback to treat persistent primary speech sound disorders. *Clinical Linguistics & Phonetics*, 29(8-10):575–597. PMID: 25751614.

Cleland, J., Wrench, A., Lloyd, S., and Sugden, E. (2018). *ULTRAX2020 : Ultrasound Technology for Optimising the Treatment of Speech Disorders : Clinicians' Resource Manual*. University of Strathclyde, Glasgow.

ClinicalTrials.gov (2018). Correcting residual errors with spectral, ultrasound, traditional speech therapy (c-results). NCT03737318.

Crowe, K. and McLeod, S. (2020). Children's english consonant acquisition in the united states: A review. *American Journal of Speech-Language Pathology*, 29(4):2155–2169.

Curtin, S., Hufnagle, D., Mulak, K., and Escudero, P. (2017a). *Speech Perception: Development*.

Curtin, S., Hufnagle, D., Mulak, K. E., and Escudero, P. (2017b). Speech perception: development. *Reference Module in Neuroscience and Biobehavioral Psychology*, pages 1–7.

Dale, E. W., Plumb, A. M., Sandage, M. J., and Plexico, L. W. (2020). Speech-language pathologists' knowledge and competence regarding percentage of consonants correct. *Communication Disorders Quarterly*, 41(4):222–230.

Dodd, B. (2014). Differential diagnosis of pediatric speech sound disorder. *Current Developmental Disorders Reports*, 1(3):189–196.

Dodd, B., Holm, A., Hua, Z., and Crosbie, S. (2003a). Phonological development: a normative study of british english-speaking children. *Clinical Linguistic Phoetics*, 17(8):617–643.

Dodd, B., Holm, A., Hua, Z., and Crosbie, S. (2003b). Phonological development: a normative study of british english-speaking children. *Clinical Linguistics & Phonetics*, 17(8):617–643. PMID: 14977026.

Donegan, P. and Stampe, D. (2009). Hypotheses of natural phonology. *Poznan Studies in Contemporary Linguistics - POZNAN STUD CONTEMP LINGUIST*, 45:1–31.

Epstein, M. A. and Stone, M. (2005). The tongue stops here: Ultrasound imaging of the palate. *The Journal of the Acoustical Society of America*, 118(4):2128–2131.

Farquharson, K. (2015). Language or motor: reviewing categorical etiologies of speech sound disorders. *Frontiers in Psychology*, 6.

Feldman, H. M. and Messick, C. (2016). Chapter 13: Language and speech disorders.

Fey, M. E. and Gandour, J. (1982). Rule discovery in phonological acquisition. *Journal of Child Language*, 9(1):71–81.

Flint, P. W., Haughey, B. H., Robbins, K. T., Regan Thomas, J., Niparko, J. K., Lund, V. J., and Lesperance, M. M. (2020). *Cummings otolaryngology*. Elsevier, 7 edition.

Flipsen Jr., P. (2015). Emergence and prevalence of persistent and residual speech errors. *Semin Speech Lang*, 36(04):217–223.

Forrest, K., Weismer, G., Elbert, M., and Dinnsen, D. A. (1994). Spectral analysis of target-appropriate /t/ and /k/ produced by phonologically disordered and normally articulating children. *Clinical Linguistics & Phonetics*, 8(4):267–281. PMID: 22320893.

Forrest, K., Weismer, G., Hodge, M., Dinnsen, D. A., and Elbert, M. (1990). Statistical analysis of word-initial /k/ and /t/ produced by normal and phonologically disordered children. *Clinical Linguistics & Phonetics*, 4(4):327–340.

Gibbon, F. E. (1990). Lingual activity in two speech-disordered children's attempts to produce velar and alveolar stop consonants: Evidence from electropalatographic (epg) data. *International Journal of Language and Communication Disorders*, 25(3):329–340.

Gibbon, F. E. (1999). Undifferentiated lingual gestures in children with articulation/phonological disorders. *Journal of Speech, Language, and Hearing Research*, 42(2):382–397.

Gibbon, F. E. and Crampin, L. B. (2001). An electropalatographic investigation of middorsum palatal stops in an adult with repaired cleft palate. *The Cleft Palate Craniofacial Journal*, 38(2):96–105. PMID: 11294548.

Gibbon, F. E., Hardcastle, W. J., and Dent, H. (1995). A study of obstruent sounds in school-age children with speech disorders using electropalatography. *International Journal of Language and Communication Disorders*, 30(2):213–225.

Gibbon, F. E. and Lee, A. (2017). Electropalatographic (epg) evidence of covert contrasts in disordered speech. *Clinical Linguistics & Phonetics*, 31(1):4–20. PMID: 27267128.

Gibson, J. and Carmichael, L. (1966). *The Senses Considered as Perceptual Systems*. Houghton Mifflin.

Gick, B. (2002). The use of ultrasound for linguistic phonetic fieldwork. *Journal of the International Phonetic Association*, 32(2):113–121.

Gick, B., Bacsfalvi, P., Bernhardt, B. M., Oh, S., Stolar, S., and Wilson, I. (2007). A motor differentiation model for liquid substitutions in children's speech. *Proceedings of Meetings on Acoustics*, 1(1). 060003.

Green, J. R., Moore, C. A., Higashikawa, M., and Steeve, R. W. (2000). The physiologic development of speech motor control: lip and jaw coordination. *Journal of Speech Language Hearing Research*, 43(1):239–255.

Green, J. R. and Wang, Y.-T. (2003). Tongue-surface movement patterns during speech and swallowing. *Journal of Acoustical Society of America*, 113(5):2820–2833.

Grunwell, P. (1981). The development of phonology: a descriptive profile. *First Language*, 2(6):161–191.

Grunwell, P. (1986). John l. locke, phonological acquisition and change. new york: Academic press, 1983. pp. xxii 263. *Journal of Child Language*, 13(3):599–602.

Grunwell, P. (1988). Phonological assessment, evaluation and explanation of speech disorders in children. *Clinical Linguistics & Phonetics*, 2(3):221–252.

Hamed Mozaffari, M. and Lee, W.-S. (2020). Encoder-decoder cnn models for automatic tracking of tongue contours in real-time ultrasound data. *Methods*, 179:26–36. Interpretable machine learning in bioinformatics.

Heald, S. and Nusbaum, H. (2014). Speech perception as an active cognitive process. *Frontiers in Systems Neuroscience*, 8.

Hua, Z. (2002). *Phonological Development in Specific Contexts; Studies of Chinese-speaking Children*. Child Language and Child Development. Multilingual Matters, Bristol, Blue Ridge Summit.

Iuzzini-Seigel, J., Hogan, T. P., and Green, J. R. (2017). Speech inconsistency in children with childhood apraxia of speech, language impairment, and speech delay: Depends on the stimuli. *Journal of Speech, Language, and Hearing Research*, 60(5):1194–1210.

James, M. S., Gibbon, F. E., and William, J. (2000). Covert contrast as a stage in the acquisition of phonetics and phonology. *Acquisition and the lexicon*, pages 194–207.

Joffe, V. and Pring, T. (2008). Children with phonological problems: a survey of clinical practice. *International Journal of Language and Communication Disorders*, 43(2):154–164.

Kabakoff, H., Beames, S. P., Tiede, M., Whalen, D. H., Preston, J. L., and McAllister, T. (2023). Comparing metrics for quantification of children's tongue shape complexity using ultrasound imaging. *Clinical Linguistics & Phonetics*, 37(2):169–195. PMID: 35243947.

Keating, P. A. and Lahiri, A. (2005). Articulatory and acoustic differences between palatal and velar stops. *The Journal of the Acoustical Society of America*, 88(S1):S80–S80.

Kent, R. D. (1992). *Phonological development: Models, research, implications*, chapter The biology of phonological development, pages 65–90. York Press.

Klein, H. B., McAllister Byun, T., Davidson, L., and Grigos, M. I. (2013). A multidimensional investigation of children's /r/ productions: perceptual, ultrasound, and acoustic measures. *American Journal of Speech and Language Pathology*, 22(3):540–553.

Klein, K. E., Walker, E. A., Kirby, B., and McCreery, R. W. (2017). Vocabulary facilitates speech perception in children with hearing aids. *Journal of Speech, Language and Hearing Research*, 60(8):2281–2296.

Knight, R.-A. (2012). *Phonetics: A Coursebook*. Cambridge University Press.

Ladefoged, P. and Disner, S. (2012). *Vowels and Consonants*. Wiley.

Law, J., Garrett, Z., and Nye, C. (2003). Speech and language therapy interventions for children with primary speech and language delay or disorder. *Cochrane Database of Systematic Reviews*, (3).

Lawson, E., Stuart-Smith, J., and Scobbie, J. (2008). Articulatory insights into language variation and change: Preliminary findings from an ultrasound study of derhoticization in scottish english. *University of Pennsylvania Working Papers in Linguistics*, 14.

Levelt, W. J. (1999). Models of word production. *Trends in Cognitive Sciences*, 3(6):223–232.

Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1):1–38.

Li, F., Edwards, J., and Beckman, M. E. (2009). Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in english and japanese toddlers. *Journal of Phonetics*, 37(1):111–124.

Liu, L. and Kager, R. (2018). Monolingual and bilingual infants' ability to use non-native tone for word learning deteriorates by the second year after birth. *Frontiers in Psychology*, 9.

Lulich, S. M., Berkson, K. H., and de Jong, K. (2018). Acquiring and visualizing 3d/4d ultrasound recordings of tongue motion. *Journal of Phonetics*, 71:410–424.

Mackay, D. (1956). The epistemological problem for automata. In Shannon, C. E. and Mccarthy, J., editors, *Automata Studies*, pages 235–251. Princeton University Press, Princeton, NJ.

Macken, M., Ferguson, C. A., Nelson, K. E., and van Kleek, A. (1983). *Children's Language*, volume 4 of *Children's Language Series*. Lawrence Erlbaum Associates.

MacNeilage, P. F., Davis, B. L., Kinney, A., and Matyear, C. L. (2000). The motor core of speech: A comparison of serial organization patterns in infants and languages. *Child Development*, 71(1):153–163.

Manan, H. A., Yahya, N. A., and Yusoff, A. N. (2021). Chapter 37 - aging, babble noise, and the processing of speech perception. In Martin, C. R., Preedy, V. R., and Rajendram, R., editors, *Factors Affecting Neurological Aging*, pages 427–437. Academic Press.

Mat Zin, S., Rasib, S., Suhaimi, F., and Jaafar, M. (2021). The technology of tongue and hard palate contact detection: a review. *BioMedical Engineering OnLine*, 20.

Maxwell, E. M. and Weismer, G. (1982). The contribution of phonological, acoustic, and perceptual techniques to the characterization of a misarticulating child's voice contrast for stops. *Applied Psycholinguistics*, 3(1):29–43.

May, L., Byers-Heinlein, K., Gervain, J., and Werker, J. F. (2011). Language and the newborn brain: does prenatal language experience shape the neonate neural response to speech? *Front. Psychol.*, 2:222.

McCormack, J., McLeod, S., McAllister, L., and Harrison, L. J. (2009). A systematic review of the association between childhood speech impairment and participation across the lifespan. *International Journal of Speech-Language Pathology*, 11(2):155–170.

McLeod, S. and Baker, E. (2016). *Children's speech.* Pearson, Upper Saddle River, NJ.

McLeod, S. and Crowe, K. (2018). Children's consonant acquisition in 27 languages: A cross-linguistic review. *American Journal of Speech-Language Pathology*, 27(4):1546–1571.

McLeod, S., Verdon, S., and Bowen, C. (2013). International aspirations for speech-language pathologists' practice with multilingual children with speech sound disorders: Development of a position paper. *Journal of Communication Disorders*, 46(4):375–387.

Mefferd, A. S. (2017). Tongue- and jaw-specific contributions to acoustic vowel contrast changes in the diphthong /ai/ in response to slow, loud, and clear speech. *Journal of Speech, Language, and Hearing Research*, 60(11):3144–3158.

Menn, L. (2013). *Development of articulatory, phonetic, and phonological capabilities*, page 168–214. Cambridge University Press.

Munson, B., Edwards, J., Schellinger, S. K., Beckman, M. E., and Meyer, M. K. (2010). Deconstructing phonetic transcription: Covert contrast, perceptual bias, and an extraterrestrial view of vox humana. *Clinical Linguistics & Phonetics*, 24(4-5):245–260. PMID: 20345255.

Munson, B., Schellinger, S. K., and Carlson, K. U. (2012). Measuring speech-sound learning using visual analog scaling. *Perspectives on Language Learning and Education*, 19(1):19–30.

Namasivayam, A. K., Coleman, D., O'Dwyer, A., and van Lieshout, P. (2020). Speech sound disorders in children: An articulatory phonology perspective. *Frontiers in Psychology*, 10.

Nathan, E. A. (2001). The development of speech processing skills in children with and without speech difficulties.

Nittrouer, S. and Studdert-Kennedy, M. (1987). The role of coarticulatory effects in the perception of fricatives by children and adults. *Journal of Speech, Language, and Hearing Research*, 30(3):319–329.

Noiray, A., Ménard, L., and Iskarous, K. (2013). The development of motor synergies in children: ultrasound and acoustic measurements. *Journal of Acoustical Society of America*, 133(1):444–452.

Ohala, J. J. (1974). Experimental historical phonology. pages 353–389. North-Holland Publishing Company, Amsterdam.

Otomo, K. and Stoel-Gammon, C. (1992). The acquisition of unrounded vowels in english. *Journal of Speech, Language, and Hearing Research*, 35(3):604–616.

Pandeli, H., Eska, J. F., Ball, M. J., and Rahilly, J. (1997). Problems of phonetic transcription: The case of the hiberno-english slit-t. *Journal of the International Phonetic Association*, 27(1-2):65–75.

Potter, N. L., Nievergelt, Y., and VanDam, M. (2019). Tongue strength in children with and without speech sound disorders. *American Journal of Speech-Language Pathology*, 28(2):612–622.

Preston, J. L., Leece, M. C., and Maas, E. (2016). Intensive treatment with ultrasound visual feedback for speech sound errors in childhood apraxia. *Frontiers in Human Neuroscience*, 10.

Preston, J. L., McAllister Byun, T., Boyce, S. E., Hamilton, S., Tiede, M., Phillips, E., Rivera-Campos, A., and Whalen, D. H. (2017). Ultrasound images of the tongue: A tutorial for assessment and remediation of speech sound errors. *JoVE*, (119).

Preston, J. L., McCabe, P., Tiede, M., and Whalen, D. H. (2019). Tongue shapes for rhotics in school-age children with and without residual speech errors. *Clinical Linguistics & Phonetics*, 33(4):334–348. PMID: 30199271.

Priestly, T. M. S. (1980). Homonymy in child phonology. *Journal of Child Language*, 7(2):413–427.

Ramsdell, H. L., Oller, D. K., and Ethington, C. A. (2007). Predicting phonetic transcription agreement: Insights from research in infant vocalizations. *Clinical Linguistics & Phonetics*, 21(10):793–831. PMID: 17882695.

Ramus, F., Hauser, M. D., Miller, C., Morris, D., and Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, 288(5464):349–351.

Reinstein, D. Z., Gobbe, M., Archer, T. J., Silverman, R. H., and Coleman, D. J. (2010). Epithelial, stromal, and total corneal thickness in keratoconus: Three-dimensional display with artemis very-high frequency digital ultrasound. *Journal of Refractive Surgery*, 26(4):259–271.

Roberts, J. E., Burchinal, M., and Footo, M. M. (1990). Phonological process decline from 212 to 8 years. *Journal of Communication Disorders*, 23(3):205–217.

Roxburgh, Z. (2018). *Visualising articulation: real-time ultrasound visual biofeedback and visual articulatory models and their use in treating speech sound disorders associated with submucous cleft palate*. PhD thesis, Queen Margaret Universiy, Edinburgh.

Roxburgh, Z., Cleland, J., Scobbie, J. M., and Wood, S. E. (2022). Quantifying changes in ultrasound tongue-shape pre- and post-intervention in speakers with submucous cleft palate: an illustrative case study. *Clinical Linguistics & Phonetics*, 36(2-3):146–164. PMID: 34496688.

Roxburgh, Z., Scobbie, J. M., and Cleland, J. (2015). Articulation therapy for children with cleft palate using visual articulatory models and ultrasound biofeedback. In *International Congress of Phonetic Sciences*.

Schmidt, R. A. and Lee, T. D. (2005). *Motor control and learning: A behavioral emphasis, 4th ed.* Motor control and learning: A behavioral emphasis, 4th ed. Human Kinetics, Champaign, IL, US.

Schneider, W. and Chein, J. M. (2003). Controlled and automatic processing: behavior, theory, and biological mechanisms. *Cognitive Science*, 27(3):525–559.

Seeingspeech (n.d.). How ultrasound tongue imaging (uti) works.

Seifert, M., Morgan, L., Gibbin, S., and Wren, Y. (2020). An alternative approach to measuring reliability of transcription in children's speech samples: Extending the concept of near functional equivalence. *Folia Phoniatr. Logop.*, 72(2):84–91.

Shiffrin, R. and Schneider, W. (1977). Controlled and automatic human information processing ii. perceptual learning, automatic attending and a general theory. *Psychological Review*, 84:127–190.

Shriberg, L. D., Fourakis, M., Hall, S. D., Karlsson, H. B., Lohmeier, H. L., McSweeny, J. L., Potter, N. L., Scheer-Cohen, A. R., Strand, E. A., Tilkens, C. M., and Wilson, D. L. (2010). Extensions to the speech disorders classification system (sdcs). *Clinical Linguistics & Phonetics*, 24(10):795–824. PMID: 20831378.

Shriberg, L. D. and Kwiatkowski, J. (1982). Phonological disorders ii. *Journal of Speech and Hearing Disorders*, 47(3):242–256.

Shriberg, L. D. and Lof, G. L. (1991). Reliability studies in broad and narrow phonetic transcription. *Clinical Linguistics & Phonetics*, 5(3):225–279.

SimpliLearn (2022). What is p-value in statistical hypothesis?

Singh, L. and Singh, N. C. (2008). The development of articulatory signatures in children. *Developmental Science*, 11(4):467–473.

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., and Bird, A. (1990). The iowa articulation norms project and its nebraska replication. *Journal of Speech and Hearing Disorders*, 55(4):779–798.

Smith, B. L. (1989). Clinical phonology (2nd ed.). pamela grunwell. beckenham, england: Croom helm, 1987. pp. 311. *Applied Psycholinguistics*, 10(2):247–253.

Smith, B. L., Macaluso, C., and Brown-Sweeney, S. (1991). Phonological effects shown by normal adult speakers learning new words: Implications for phonological development. *Applied Psycholinguistics*, 12(3):281–298.

Stackhouse, J. and Wells, B. (1997). *Children's speech and literacy difficulties, Book1.* Exc Business And Economy (Whurr). Whurr, London, England.

Steinschneider, M., Volkov, I. O., Noh, M. D., Garell, P. C., and Howard, M. A. (1999). Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *Journal of Neurophysiology*, 82(5):2346–2357. PMID: 10561410.

Stoel-Gammon, C., Bernhardt, B. H., Gilbert, J., and Ingram, D. (1996). Proceedings of the ubc international conference on phonological acquisition.

Stoel-Gammon, C. and Dunn, C. (1985a). *Normal and Disordered Phonology in Children*. Child language acquisition series. PRO-ED, Incorporated.

Stoel-Gammon, C. and Dunn, C. (1985b). *Normal and Disordered Phonology in Children*. Child language acquisition series. PRO-ED, Incorporated.

Sugden, E. and Cleland, J. (2022). Using ultrasound tongue imaging to support the phonetic transcription of childhood speech sound disorders. *Clinical Linguistics & Phonetics*, 36(12):1047–1066. PMID: 34605343.

Sugden, E., Lloyd, S., Lam, J., and Cleland, J. (2019). Systematic review of ultrasound visual biofeedback in intervention for speech sound disorders. *International Journal of Language and Communication Disorders*, 54(5):705–728.

Terband, H., Maassen, B., and Maas, E. (2019). A psycholinguistic framework for diagnosis and treatment planning of developmental speech disorders. *Folia Phoniatr. Logop.*, 71(5-6):216–227.

Trujillo, F. (2002). English phonetics and phonology - the production of speech sounds.

Tyler, A. A. and Saxman, J. H. (1991). Initial voicing contrast acquisition in normal and phonologically disordered children. *Applied Psycholinguistics*, 12(4):453–479.

Ultrax (2017). Ultrax: Ultrasound-based techniques for speech therapy.

Ultrax (n.d.). Ultrax: Ultrasound-based techniques for speech therapy.

unknown (2020). Dysarthria.

van den Engel-Hoek, L., de Groot, I. J., de Swart, B. J., and Erasmus, C. E. (2015). Feeding and swallowing disorders in pediatric neuromuscular diseases: An overview. *Journal of Neuromuscular Diseases*, 2:357–369. 4.

Vihman, M. (1978). Consonant harmony: Its scope and function in child language. *Phonology*, 2:281–334.

Weaver, J. (2015). Motor learning unfolds over different timescales in distinct neural systems. *PLOS Biology*, 13(12):1–2.

Wellman, B. L., Case, I. M., Mengert, I. G., and Bradbury, D. E. (1931). Speech sounds of young children. University of Iowa Studies: Child Welfare. University of Iowa Press.

Wood, S. and Scobbie, J. (2003). Evaluating the clinical effectiveness of epg in the assessment and diagnosis of children with intractable speech disorders.

Wren, Y., Miller, L. L., Peters, T. J., Emond, A., and Roulstone, S. (2016a). Prevalence and predictors of persistent speech sound disorder at eight years old findings from a population cohort study. *Journal of Speech, Language, and Hearing Research*, 59(4):647–673.

Wren, Y., Miller, L. L., Peters, T. J., Emond, A., and Roulstone, S. (2016b). Prevalence and predictors of persistent speech sound disorder at eight years old: Findings from a population cohort study. *Journal of Speech, Language, and Hearing Research*, 59(4):647–673.

Yalcinkaya, F., Bayar Muluk, N., and Budak, B. (2010). Speech sounds acquisition evaluated by speech sound development test (ssdt) in turkish-speaking children. *Journal of International Advanced Otology*, 6:60–66.

Yeh, K. (2011). Articulation development: What's normal? (& what isn't).

Young, E. C. and Gilbert, H. R. (1988). An analysis of stops produced by normal children and children who exhibit velar fronting. *Journal of Phonetics*, 16(2):243–246.

Zanichelli, L. and Gil, D. (2011). Percentage of consonants correct (PCC) in children with and without hearing loss. *J. Soc. Bras. Fonoaudiol.*, 23(2):107–113.

Zharkova, D. N. (2013). Using ultrasound to quantify tongue shape and movement characteristics. *The Cleft Palate Craniofacial Journal*, 50(1):76–81. PMID: 22117937.

Appendix A

List of Prompts Recorded by Participants

Non-lexical speech material data (aCa)

Ten repetitions of English obstruents and sonorants in /aCa/ context where phonotactically permissible):

1. ata /ata/
2. ana /ana/
3. ara /aɹa/
4. atha /aθa/
5. asa /asa/
6. asha /aʃa/
7. ala /ala/
8. aya /aja/
9. aka /aka/
10. acha /aʧa/

11. bang /baŋ/
12. apa /apa/
13. ama /ama/
14. afa /afa/
15. awa /awa/

Single word level speech material data

One imitation of minimal pair word sets (t/k singleton minimal pairs):

1. a sock a **t**ore a chore a chip
2. she a **k**ip a sore a chop
3. a chew a shack a ship a **c**ool
4. a **k**eep a shoot a chap a suit
5. a **c**ore a sip a **c**op a **t**ool
6. a **t**ip a **t**op a sea a **t**an
7. a **c**an a cheer a shock a sack
8. a shore a **t**eam