Published in final edited form as:

Clin Linguist Phon. 2016 ; 30(3-5): 249–276. doi:10.3109/02699206.2015.1056884.

Covert contrast in velar fronting: An acoustic and ultrasound study

Tara McAllister Byun¹, Adam Buchwald¹, and Ai Mizoguchi²

¹New York University, New York, NY

²City University of New York Graduate Center, New York, NY

Abstract

There is growing evidence that speech sound acquisition is a gradual process, with instrumental measures frequently revealing *covert contrast* in errors perceived to involve phonemic substitution. Ultrasound imaging has the potential to expand our understanding of covert contrast by showing whether a child uses different tongue shapes while producing sounds that are perceived as neutralized. This study used an ultrasound measure (Dorsum Excursion Index) and acoustic measures (VOT and spectral moments of the burst) to investigate overt and covert contrast between velar and alveolar stops in child speech. Participants were two children who produced a perceptually overt velar-alveolar contrast and two children who neutralized the contrast via velar fronting. Both acoustic and ultrasound measures revealed significant differences between perceptually distinct velar and alveolar targets. One child with velar fronting demonstrated covert contrast in one acoustic and one ultrasound measure; the other showed no evidence of contrast. Clinical implications are discussed.

Covert contrast in child phonology

Both typically developing children and children with phonological delay or disorder produce speech sounds that deviate systematically from adult patterns of production. These patterns are often described in terms of substitution of one phoneme for another, or neutralization of a contrast between two phonemes. For instance, a child with the phonological pattern of *velar fronting* may be perceived to neutralize /k/ and /t/ sounds so that words such as ‘tea’ and ‘key’ sound the same ([ti]). However, analyses based on impressionistic transcription may not provide a sufficiently detailed characterization of child speech patterns. An adult listener with a fully developed phonology is predisposed to transcribe a child's productions using segments and contrasts familiar from his/her own phonology, but this transcription may in fact stand at quite a distance from phonetic reality (e.g. Amorosa, von Benda, Wagner, & Keck, 1985).

With these concerns in mind, a number of instrumental analyses have been undertaken to gather the precise phonetic details that are omitted from transcription studies of child speech

Address correspondence to: Tara McAllister Byun, Department of Communicative Sciences and Disorders, New York University, 665 Broadway, Room 914, New York, NY 10012, USA, Phone: 212-992-9445, Fax: 212-995-4356, tara.byun@nyu.edu.

Declaration of Interest Statement: The authors report no conflicts of interest.

(e.g. Young & Gilbert, 1988; Tyler, Edwards, & Saxman, 1990; Tyler, Figurski, & Langsdale, 1993; Edwards, Gibbon, & Fourakis, 1997; Scobbie, 1998). With surprising frequency, these studies have shown that where broad transcription indicates a categorical error, such as omission of a segment or neutralization of a contrast, instrumental analysis can detect traces of the correct target. For example, multiple studies have found that children who are perceived to neutralize voiced and voiceless stops do maintain a statistically reliable distinction in voice onset time (VOT) between voiced and voiceless targets (e.g., Macken & Barton, 1980; Hitchcock & Koenig, 2013). These measurable but perceptually subtle phonetic distinctions are termed *covert contrasts*.¹

Covert contrast in child phonology is of more than academic interest for two major reasons. First, the presence or absence of covert contrast can offer evidence regarding the level of processing at which a child's speech errors apply, with covert contrast interpreted as evidence for errors at a more peripheral or articulatory level (Gibbon, 1999). Second, instrumental studies of child speech have reported that covert contrast is a positive prognostic indicator in child speech development: children who make a covert distinction between two targets often progress to an overt contrast without any intervention (Forrest, Weismer, Hodge, Dinnsen, & Elbert, 1990). If these children do enter treatment, they tend to make more rapid progress than children who make a complete neutralization (Tyler et al., 1990; Tyler et al., 1993; Gibbon, Dent, & Hardcastle, 1993).

One limitation of studies of covert contrast is the inconclusive nature of negative results. A phonetic distinction could be realized in any of a large number of acoustic dimensions, and children acquiring speech often mark contrast in phonetically unusual ways (Scobbie et al., 2000). Thus, if a given study reports no evidence of contrast in the limited set of parameters selected for measurement, it remains possible that contrast was marked in an unmeasured domain. For certain distinctions, notably in place of articulation, it is possible that articulatory measures could provide a more direct means to identify covert contrast than acoustic measures. The study reported here makes use of ultrasound tongue imaging to look for covert contrast in children's apparent neutralization of velar and coronal place in the process of velar fronting.

Velar fronting

Velar fronting is a child speech pattern in which target velar sounds (e.g. /k, g/) are perceived to be produced with an alveolar place of articulation. Velar fronting is commonly observed in children with articulatory-phonological disorders, but it is also well-attested in typical development. Grunwell (1981) proposed that velar fronting should be considered typical in children up to three years, six months of age. Examples of velar fronting are provided in (1). Velar fronting may be 'across-the-board' (applying across all positions in the syllable), or it may manifest as 'positional velar fronting', which characteristically affects velars in onset

¹Characterization of a contrast as "covert" does not entail that it is impossible to perceive. A sufficiently fine-grained perceptual task, such as narrow transcription by a trained listener, or quantification of gradient degrees of perceived difference using visual analog scaling (e.g., Munson, Schellinger, & Urberg-Carlson, 2012) can succeed in identifying a contrast that is not represented in broad transcription. Throughout this paper, contrasts will be characterized as "neutralized" if they would not be distinguished by a casual listener or in broad transcription; it remains possible that an expert listener or fine-grained perceptual scale could tease out a contrast.

but not coda position, as exemplified in (1d) below. Note that the actual productions transcribed below may also reflect errors other than fronting, such as vocalization of liquids or voicing errors.

1. Examples of velar fronting produced by a 4-year-old child with speech sound delay/disorder (McAllister Byun, 2012)
 - a. [tʌp] ‘cup’
 - b. [dæso] ‘castle’
 - c. [dʊtao] ‘guitar’
 - d. [dajak] ‘kayak’

A number of studies have looked for acoustic evidence of covert contrast in children's perceptually neutralized realizations of alveolar and velar place (Young & Gilbert, 1988; Forrest et al., 1990; Tyler et al., 1990; Tyler et al., 1993, Edwards et al., 1997). For example, in typical production, different places of articulation are associated with different average voice onset time (VOT) values. As a general rule, VOT values are longer for more posterior places of articulation, such that labial stops have the shortest average VOT, followed by alveolar stops and then by velar stops (Lisker & Abramson, 1964; Klatt, 1975). Thus, if the target velars produced by children with velar fronting are more ‘velar-like’ than target alveolars, they might be produced with a measurably longer VOT. In a study comparing VOT values in perceptually neutralized velar and alveolar targets, two out of four velar-fronting children were found to make a consistent distinction in the unexpected direction, such that target velars had significantly shorter VOTs than target alveolar stops (Tyler et al., 1990). However, another study found no significant difference in VOT between alveolar stops and target velar stops produced with fronting (Young & Gilbert 1988).

Locus equations represent a linear regression over F2 frequencies, with F2 at onset plotted as a function of F2 at midvowel. In adults, the slope and y-intercept values of this equation have been found to vary predictably as a function of consonant place (Lindblom 1963). Sussman, Hoemeke, & McCaffrey (1992) found that locus equations can also be used to differentiate consonant place in child speakers. Tyler et al. (1993) used locus equations to analyse perceptually indistinguishable stops produced by three children who fronted velars to alveolar place in all positions in the syllable. However, their measures returned no evidence of covert contrast.

A final acoustic measure that has been used to look for covert contrast in children's fronted velars involves analysis of the spectral moments of the stop burst. This involves treating the spectrum of energy generated in the stop release as a random distribution and summarizing it in terms of mean (centroid), variance, skewness, and kurtosis. In typical adult speech, /t/ is associated with a higher centroid frequency, a negative skewness (indicating a greater concentration of energy in high frequencies), and a moderate level of kurtosis (indicating that energy is spread fairly broadly through the spectrum). On the other hand, /k/ is typically produced with a centroid frequency in the middle range, minimal skewness, and a high kurtosis indicating a more compact spectrum. Forrest et al. (1990) conducted a spectral moments analysis of velar and alveolar targets produced by four children perceived to front

velars in all positions in the syllable. For three children, Forrest et al. found no significant difference between /t/ and /k/ tokens in any spectral parameter. However, the fourth child produced alveolar and velar targets with a statistically reliable difference in both kurtosis and skewness.

Overall, the existing literature looking for acoustic evidence of covert contrast in velar fronting shows that a wide range of parameters have been measured, with variable results. Because the target contrast involves a major place of articulation distinction, it is possible that articulatory measures could provide a more direct and sensitive indicator of contrast than acoustic measurements. A promising body of research has used electropalatography (EPG) to supplement acoustic and perceptual measures of child speech. EPG research by Gibbon and colleagues (Gibbon, 1990, 1999; Gibbon, Dent, & Hardcastle, 1993) suggests that 'undifferentiated lingual gestures' are widespread in child speech. These productions are characterized by midsagittal contact extending from alveolar into palatal and sometimes velar regions of the palate, indicating simultaneous elevation of tongue tip/blade and body regions. Speech sounds produced with undifferentiated linguopalatal contact may be transcribed as phonemic substitutions, in which case electropalatographic data may reveal covert contrast. For instance, Gibbon (1990) found significantly different patterns of tongue-palate contact for velar and alveolar targets produced by a child who was perceived to neutralize the velar-alveolar contrast in a coronal backing pattern.

However, there are several factors that tend to limit the utility of EPG as a means of collecting data on child speech. It requires that the speaker be fitted with a custom-made pseudopalate, which introduces a delay and expense; furthermore, young children may not tolerate the pseudopalate. Ultrasound imaging of speech has the potential to represent a quicker, easier source of information about articulation. Because it does not require fabrication of individualized materials like the EPG pseudopalate, it can be used the first time a child is seen. However, it is not yet known whether ultrasound is an effective way to reveal covert contrast between velar and alveolar targets in children with velar fronting. The present study represents a preliminary attempt to answer this question.

Ultrasound imaging of speech

During ultrasound imaging of articulation for examining tongue contours during stop production, the ultrasound probe is placed in a midsagittal position beneath the chin, roughly equidistant between the mentum and the neck. When the high-frequency waves emitted by the transducer encounter a change in density at the boundary between the tissues of the tongue and the air above the tongue, they reflect and are captured by the probe. The reflected sound energy is used to create an image of the surface of the tongue. Rotating the probe by 90 degrees around its longitudinal axis makes it possible to shift between sagittal and coronal views of the tongue (e.g., Slud, Stone, Smith and Goldstein, 2002). Sagittal tongue shapes captured during a 4;2-year-old child's production of perceptually typical alveolar and velar stops are provided in figure 1(a)-(b). The tongue surface is visible as a white line on the black background. The right side of each figure represents the anterior direction (tongue tip). Two shadows are typically seen flanking the image of the tongue on either side; the left

shadow represents refraction of the ultrasound beam by the hyoid bone, and the right shadow represents refraction by the mandible.

One limitation of ultrasound imaging for speech is the lack of a fixed point of reference on the ultrasound display. First, note that the ultrasound image does not reveal the position of the tongue relative to the palate: the sound waves emitted by the ultrasound probe typically do not reach the palate because they are reflected by the air in the oral cavity above the tongue. Second, unless the ultrasound probe is fixed in place relative to the head or the jaw, it is prone to displacement in vertical, horizontal, and/or angular dimensions. Thus, there is no guarantee that images of the tongue are occupying the same coordinate space from trial to trial, or even from frame to frame, which limits the inferences that can be drawn from these measures. One solution involves stabilizing the position of the ultrasound transducer relative to the head using specialized helmets, probe holders, or braces to position the subject.² However, this approach is not well-tolerated by children in the age range that characteristically exhibits velar fronting.

A more child-friendly alternative would make use of measurements that can differentiate velar versus alveolar tongue shapes without requiring head-to-transducer stabilization. Such measures were devised by Zharkova (2013) for use in the clinical analysis of abnormal lingual stops produced by children with cleft palate. Zharkova emphasized the importance of using a measure that can quantify tongue shape with data from a single tongue curve, since it cannot be assumed that the unstabilized probe will remain in the same coordinate space across multiple frames. Zharkova's measures are similar to the single-frame measures that Ménard, Aubin, Thibeault, and Richard (2012) characterized as robust across vertical and horizontal probe translations, as well as pitch rotation in the same plane.

The contrast between tongue shapes used for velar and alveolar stops can be sensitively characterized by the measure that Zharkova terms Dorsum Excursion Index (DEI). Calculating DEI, which can be accomplished using a script in R or Matlab (Zharkova, unpublished), involves tracing a straight line from the most posterior to the most anterior point of the tongue curve. This line is labelled N in figure 2 below. A line perpendicular to N is then generated at the midpoint of N, and the distance from N to the tongue surface is calculated. This line segment is labelled D in figure 3. DEI is then calculated as the ratio of D to N, i.e. the magnitude of excursion of the tongue dorsum relative to front-to-back extension of the tongue front and back. A higher DEI is indicative of a greater degree of dorsum excursion. Zharkova reports that 0.5 would be considered a typical DEI for a velar stop, versus roughly 0.26 for an alveolar stop. DEI can be influenced by how much of the tongue is included in the ultrasound image; therefore, only images that show the entire tongue from hyoid shadow to mandibular shadow should be analysed with this measure. Even when the tongue is visible from shadow to shadow, for high and anterior sounds the tip of the tongue may be obscured by the mandibular shadow, which could also affect DEI.

²Because stabilization is never perfect, it is often paired with video recording or optical tracking to detect changes in probe placement, which can be used to correct or discard problematic trials (e.g., Whalen et al., 2005). These techniques were not used in the present study.

However, calculations reported by Zharkova (2013) suggest that the impact on DEI in such cases is relatively minor.

Hypotheses

The present study explored the utility of ultrasound imaging as a means to detect covert contrast in children's perceptually neutralized velar and alveolar stop productions. Near-minimal pairs featuring velar and alveolar targets were elicited from two preschool-aged children with velar fronting and two matched children who produced a perceptually appropriate velar-alveolar contrast. Velar and alveolar targets were compared using one ultrasound measure (dorsum excursion index) and two acoustic measures that have succeeded in detecting covert contrast in previous studies of children with velar fronting (VOT and spectral moments of the stop burst). It was hypothesized that both acoustic and ultrasound measures would differ significantly in a comparison between children's velar and alveolar stop targets produced with an overt place distinction. In a comparison between velar and alveolar stop targets that are perceived to feature neutralization to alveolar place, it was hypothesized that ultrasound would serve as a more sensitive marker of covert contrast than acoustic measures. That is, it was predicted that in at least some speakers and/or contexts, ultrasound measurements would reveal contrast where acoustic measures failed to do so.

Methods

Participants

Participants were 2 children with velar fronting (2 males) and 2 children who produced an accurate velar-alveolar contrast (1 male). The two participants with velar fronting were referred from the New York University Speech-Language-Hearing clinic, where they were receiving treatment for speech sound errors including velar fronting. One participant exhibited across-the-board velar fronting and one exhibited positional fronting affecting only syllable-initial velars. These classifications of participants' speech patterns were determined based on the first author's transcriptions of a combination of spontaneous and imitative utterances from each child. Participants who produced an appropriate velar-alveolar contrast were recruited from the surrounding community. By parent report, participants had no history of major behavioural, neurological, or hearing impairment. Ages and other participant details are provided in table 1. All names used in table 1 and throughout the paper are pseudonyms.

One participant, Max, was recorded at two points in his process of velar acquisition. At the first time point, he had not yet received intervention specifically targeting the velar-alveolar place contrast, and he was consistent in fronting velar targets in initial but not in final position. At the second time point, roughly three months later, he was enrolled in intervention targeting the velar-alveolar contrast and had begun to show progress. He was observed to produce perceptually accurate velar place in initial position, but in an inconsistent fashion. When his fronted outputs were called to his attention, he was generally able to self-correct to a perceptually appropriate velar place.

Data collection

The study involved simultaneous ultrasound and audio recording during elicitation of multiple tokens of near-minimal word pairs featuring velar and alveolar target sounds. Target words *cup-tub* and *key-T* (letter) were used to elicit the targets in initial position; *back-bat* and *bug-mud* elicited the contrast in final position. During the first part of the study, the participant was familiarized with the experimental protocol. First the stimulus images were presented to elicit all target words; prompts and feedback were provided to ensure that the participant named all items in the intended fashion. The participant also practiced producing each word multiple times in succession, repeating the same word until the experimenter gave a visual cue to stop. The participant was then introduced to the recording environment, a large sound booth in the Department of Communicative Sciences and Disorders at New York University. The ultrasound equipment (GE LogiqE ultrasound with 8C transducer) was presented, and the participant was given opportunities to hold the ultrasound probe and observe movements of the tongue during his/her own speech and that of others. During the experimental task, the child was seated in a padded elevated chair facing a laptop screen that displayed the stimulus images, and ultrasound images were obtained while the child produced each word in a single block of five consecutive trials. Words were pseudo-randomly ordered across blocks such that the two members of a minimal pair were not presented consecutively. Because the child had practiced naming the stimulus images in the first part of the study, verbal models were not necessary.

The ultrasound machine was positioned on a table to the side of and slightly behind the child; the display screen was not visible to the child. The ultrasound probe was manually positioned beneath the child's chin by one of the experimenters, and a child-sized sporting helmet with a specialized attachment was used as an additional source of support and stabilization for the ultrasound probe.³ The helmet was not tolerated by one participant with velar fronting ('Max', both time points), and manual positioning alone was used for data collection from this child. The experimenter holding the probe faced the child and viewed the ultrasound display behind the child, monitoring the image for any changes suggestive of probe slippage. Minor probe deviations were corrected between trials in a block. If major deviation was detected, or if background noise or other disruptions occurred, a note was made and the word was collected in an additional block of five trials at the end of the recording protocol. The experimental session elicited a minimum of 40 trials with ultrasound imaging (8 words $\times$ at least 5 repetitions per block).

The ultrasound capture settings were identical for each child. The radial resolution of the probe was 133 degrees, with a 10MHz scan rate and a 35Hz capture rate. The line density was set to 5, and the circumferential resolution was 9cm, a value that is large enough to capture a range of vocal tract sizes. The ultrasound machine contained a single VGA output for video which was passed to a PC in a console room attached to the data collection booth. A Shure 58A unidirectional microphone connected to a Mobile-Pre pre-amp was positioned on a microphone stand at a distance of approximately 6 inches from the participant's mouth.

³This helmet was developed by the authors in collaboration with a mechanical engineer. It is subjectively judged to assist with probe stabilization, particularly to minimize rotational changes in probe angle that can alter the plane of section. However, formal testing of this device has not been undertaken.

⁴ Audio files were recorded with a 44.1kHz sampling rate and 16-bit encoding and were passed to the same PC in the console room. The audio and video signals were synchronized using a PCI card designed to capture and synchronize video from VGA with the audio signal. As described below, the frames for the analysis of articulatory data were identified primarily based on articulatory landmarks, but searching for those frames was guided by the acoustic record. In particular, we located the closure and release in the acoustics and calculated the corresponding ultrasound frame based on the capture rate. We then looked at 3 frames before and after the release in the acoustics to find the frame meeting our definition of the maximal constriction (see below).

Measurement

Acoustic measurements—Acoustic measurements were carried out by the third author using Praat (Boersma & Weenink, 2011). Two measures were selected that have succeeded in detecting covert contrast in previous studies of velar fronting: VOT (e.g., Tyler et al., 1990) and spectral moments of the stop burst (e.g., Forrest et al, 1990). VOT could be calculated only for stops in onset position. Spectral moments of the burst could be computed for any released stops in either initial or final position.

VOT—VOT was measured with reference to both the spectrogram and the waveform, but precise durational measures were collected from the waveform exclusively. The onset of the burst was marked at the first upward-going zero crossing associated with the spike in energy accompanying the stop release. If multiple bursts were present, measurements were taken starting from the first burst (Davis, 1995). The onset of the following vowel was marked in the waveform at the zero crossing preceding the first well-formed glottal pulse associated with the onset of periodicity. The duration of time separating these two points was coded as the VOT.

Spectral moments—For spectral moments analysis, the onset of the burst was identified as stated above. The offset of the burst was visually estimated at a point in the waveform characterized by a marked decrease in the amplitude of aperiodic energy, or if there was no visible offset of the burst, a fixed window of 10 ms was used (Repp, 1984). Spectral moments were not calculated from stops for which the total burst duration was less than 10 ms. The burst interval was segmented out of each utterance for analysis. A Matlab script was used to compute multitaper spectra for each interval, then extract the first four spectral moments (centroid, slope, skewness, and kurtosis) at the midpoint of each interval. Multitaper analysis is reported to provide more accurate estimates than conventional methods of spectral analysis (Lousada, Jesus, & Pape, 2012).

Ultrasound measurements

The ultrasound tongue images were prepared for quantitative analysis by a team of trained research assistants (graduate students in speech-language pathology). First, the ultrasound

⁴The ultrasound machine was in the room with the microphone and thus contributed to background noise, but its contribution was consistent across all recordings. Because this study focuses on within-subject comparisons, no corrections for this consistent background noise were undertaken.

frames of interest were identified with reference to the closure interval in the synchronized acoustic record. Second, a single frame representing the point of maximum constriction was identified within this interval, operationalized as the frame immediately before the tongue contours exhibited a transition into the next sound. Three research assistants independently judged which frame represented the point of maximum constriction, with one of the authors adjudicating cases of disagreement. Tokens were discarded if poor image quality made it impossible to distinguish the actual tongue edge from visual artefacts, or if the full surface of the tongue from hyoid shadow to mandibular shadow was not visible at the point of closure. The selected frames were then processed using EdgeTrak (Li, Kambhamettu, & Stone, 2005), a software that semi-automatically traces the tongue surface visible in an ultrasound image. Each frame was tracked independently by two research assistants. Tracked tongue contour data were extracted as x,y coordinates for the purpose of quantitative analysis. An R script (Zharkova, unpublished) was then used to generate an image of each tongue contour and draw the line segments N and D, described above. Finally, DEI values resulting from each of two students' tracking of each tongue contour were compared as a reliability check. The mean of the absolute value of this difference in DEI was calculated for each participant. Any items that fell more than 2 SD above that participant's mean difference in DEI were discarded, and any items that fell between 1 and 2 SD were inspected by the authors, and the contour that was judged to be less accurate was discarded. In most cases, discrepancies in DEI appeared to stem from between-rater differences in the choice of where to terminate the trace of the tongue contour. Figure 3 depicts two stages in the processing of the tongue image for Emma's perceptually accurate production of *T* that was seen in figure 1A: (a) the ultrasound image with semi-automated tracking in Edgetrak; (b) the tongue contour extracted and annotated using Zharkova's R script. Figure 4 provides the same stages of processing for Emma's perceptually accurate production of *key*, previously seen in figure 1B.

Reliability

To assess the reliability of acoustic measurements, a randomly selected subset of tokens representing 10% of the full sample was independently re-measured by the first author. VOT values were extremely highly correlated across the two raters ($r = .996$; $t = 63.5$, $df = 32$, $p < .001$). In addition, the first author's measurements of stop bursts in the reliability subset yielded spectral measures that were strongly correlated with the original measurements for all four moments (M1: $r = .96$; $t = 20.0$, $df = 32$, $p < .001$; M2: $r = .93$; $t = 14.3$, $df = 32$, $p < .001$; M3: $r = .99$; $t = 37.1$, $df = 32$, $p < .001$; M4: $r = .997$; $t = 76.0$, $df = 32$, $p < .001$).

For ultrasound data, interrater reliability was calculated by comparing the DEI value extracted from tongue contours tracked by two separate research assistants. All tokens from all participants were included in this analysis, with the exception of (a) two tokens for which only one assistant's measurement was available; (b) 30 tokens for which the entire tongue shape was not visible. The mean absolute difference in DEI between two raters' measurements was .047 (SD = .048). The two measurements for each token were highly correlated across tokens ($r = .8$; $t = 16.3$, $df = 152$, $p < .001$). An intraclass correlation (ICC) with single random raters was calculated to be .80, representing an adequate level of interrater agreement.

Analyses

Acoustic analyses

VOT—VOT data were analysed with separate *t*-tests for each child. VOT values measured for all target velars in initial position for a given child were compared against VOT values measured for initial target alveolars produced by the same child. Because only voiceless targets were elicited in initial position, it was not necessary to include voicing in the analysis. Vowel context also was not incorporated into this analysis due to the small number of syllable-initial tokens available to represent each target and vowel context in each child. Furthermore, existing literature shows a lack of agreement as to whether and how vowel context influences VOT (e.g. Lisker & Abramson, 1967; Klatt, 1975).

Spectral moments—The relationship between spectral moments of the stop burst and target place of articulation was examined using linear regression. Separate regressions were computed for each child. Due to multicollinearity among the four spectral moments, each parameter was examined in a separate regression with target place of articulation (velar or alveolar) and vowel context (back versus front) as the independent variables. Position (initial or final) was also included in these regressions when possible, but it was excluded from the analyses testing for covert contrast between target alveolars and fronted velars produced by Max, since he never exhibited fronting in final position.

Ultrasound analyses

Ultrasound data were analysed using generalized linear mixed models, with separate models constructed for each child and session. The dependent variable was DEI (Zharkova, 2013), where a higher DEI indicates a greater degree of tongue dorsum activity in producing a particular sound. The fixed effects in the model included the phonological identity of the target (target velar or target alveolar), vowel context (back or front), and position (initial or final), as well as first-order target-by-vowel and target-by-position interactions. Finally, the model included a random effect of rater to adjust for potentially differing biases among the several individuals who contributed to the measurement of ultrasound data.

Results

Acoustic measures

Of the 235 audio files originally reviewed, 42 were discarded: 18 due to competing background noise, 14 due to unreleased stops, and 10 due to other anomalies. In addition, 40 tokens were excluded from the inferential statistical analyses because of their status as perceptually accurate velar stops produced by Max, a participant with velar fronting. However, descriptive statistics (means, standard deviations) are reported and graphed for these tokens.

VOT—Figure 5 depicts mean VOT in initial velar versus alveolar targets for each participant and session, and table 2 reports the numerical values for mean and standard deviation in each case. Emma, who produced an overt contrast, was the only participant to show a statistically significant difference in VOT between velar and alveolar targets ($t = 2.1$, $df =$

16.8, $p = .048$). The other participant who produced an overt contrast, Jayden, showed a difference in VOT of similar magnitude, but the comparison did not reach significance in this case ($t = 2.0$, $df = 13.7$, $p = .06$). Differences were not significant for velar fronting participants Rory ($t = -0.08$, $df = 16.0$, $p = .94$) or Max (session 1: $t = -0.1$, $df = 4.4$, $p = .92$; session 2: $t = .62$, $df = 7.7$, $p = .55$).⁵

Spectral moments—Figure 6 depicts mean values for the first four spectral moments of the stop burst for both velar and alveolar targets produced by participants Emma and Jayden. Table 3 reports numerical results (mean, standard deviation) for all children.

Figure 6 reveals that for all four spectral moments, both typical participants showed a similar pattern of relative magnitude for velar versus alveolar targets. Therefore, data were pooled across these two participants prior to regression analysis.⁶ Each of the four spectral moments was examined in a separate linear regression due to multicollinearity. These regressions revealed that velar versus alveolar target place served as a significant predictor of all four spectral moments (M1: $\beta = -467.34$, $SE = 180.7$, $t = -2.6$, $p = .01$; M2: $\beta = -327.0$, $SE = 124.9$, $t = -2.6$, $p = .01$; M3: $\beta = 2.7$, $SE = .77$, $t = 3.5$, $p < .001$; M4: $\beta = 24.9$, $SE = 9.1$, $t = 2.7$, $p < .01$). Because velar place served as the reference level, negative coefficients indicate that velar targets were associated with lower values for a given spectral moment. Vowel context also served as a significant predictor in one regression: higher mean values of the first spectral moment were observed in front than back vowel contexts ($\beta = 523.2$, $SE = 236.5$, $t = 2.2$, $p = .03$). No model showed a significant effect of position, and there were no significant interactions. Complete results of these regressions are reported in appendix 1A.

Figure 7 shows mean values of the first four spectral moments for alveolar targets and fronted velar targets produced by the two speakers with velar fronting; specific numerical values can be found in table 3. For each child and session, separate linear regressions examined whether any of the four spectral moments differed significantly depending on whether a stop burst was produced in connection with a velar as opposed to an alveolar target, or whether this predictor interacted with vowel context.

For Rory, velar versus alveolar place never served as a significant predictor of any spectral moment, nor did it appear in any significant interactions. Rory did show a significant effect of vowel context for the third and fourth spectral moments, with front vowels associated with higher values of skewness and kurtosis (M3: $\beta = 1.2$, $SE = .57$, $t = 2.2$, $p = .04$; M4: $\beta = 5.8$, $SE = 2.7$, $t = 2.2$, $p = .04$); there was also a significant interaction of vowel context and position for the third moment only ($\beta = -1.4$, $SE = .63$, $t = -2.2$, $p = .04$). Complete regression results for Rory are reported in appendix 1B.

For Max, an effect of position was not included in the regression at either time point, because all final velars were produced with perceptually accurate place. In addition, the

⁵Inclusion of the perceptually accurate velar targets did not affect the significance of the comparison for Max's session 2 ($t = -.57$, $df = 20.9$, $p = .57$).

⁶Note that this is the only analysis in which data from Emma and Jayden, the two children who produced a perceptually overt contrast, were merged. This was for reader convenience: because the four spectral moments were examined in separate regressions, the number of results reported quickly becomes overwhelming. When regressions for these two individuals were computed separately, velar versus alveolar target place remained a robustly significant predictor of all four spectral moments for both children.

regression analysing data from Session 2 did not examine the interaction between target place and vowel context, because Max had ceased to front initial velars in the back vowel context. However, spectral moments measured for these perceptually accurate outputs are included for qualitative inspection in figure 7.⁷ In Session 1, velar versus alveolar target place did serve as a significant predictor of the fourth spectral moment, kurtosis (M4: $\beta = 3.7$, $SE = 1.5$, $t = 2.4$, $p = .02$). The direction of this effect was consistent with that observed in the two speakers who produced a perceptually overt contrast; thus, this can be considered an example of covert contrast. There was also a significant interaction of target place and vowel context for the fourth spectral moment, reflecting the fact that the difference in kurtosis between velar and alveolar targets was prominent in the back vowel context but minimal in the front vowel context. This interaction can be seen as predictive of the results observed in Session 2, where an overt velar-alveolar contrast was realized in the back vowel context. In Session 2, there were no effects consistent with covert contrast between target alveolar and fronted velar stops, but there was a significant effect of vowel context, reflecting significantly higher values of the second spectral moment in the front vowel context ($\beta = 284.0$, $SE = 102.2$, $t = 2.8$, $p = .01$). Complete regression results for Max's two sessions are reported in appendix 1C-D.

Ultrasound measures

Of the 201 ultrasound files originally reviewed, 38 were discarded: 6 due to poor image quality, 11 due to anomalies in way the participant produced the word, 3 due to background noise, 10 because the entire tongue surface was not visible, and 8 due to insufficient agreement between DEI measurements extracted from two research assistants' independent tracking of tongue contours. The remaining 163 items were used in the analyses reported below.

Overt contrast—In the generalized linear mixed model analysing perceptually distinct alveolar and velar tokens produced by Emma, target place of articulation was a significant predictor of DEI ($\beta = 0.24$; $SE = 0.02$, $t = 10.7$; $p < 0.001$). In this case, the positive coefficient indicates that velar targets were associated with significantly higher DEI than alveolar targets, consistent with expectations from Zharkova (2013). DEI was also higher in front than back vowel contexts ($\beta = 0.07$; $SE = 0.01$, $t = 4.4$; $p < 0.001$) and slightly higher in initial than final position ($\beta = 0.04$; $SE = 0.01$, $t = 2.4$; $p = 0.02$). Lastly, there was a significant interaction between target and vowel context ($\beta = -0.19$; $SE = 0.02$, $t = -7.9$; $p < 0.001$), reflecting a more extreme difference in DEI between velar and alveolar targets in back than front vowel contexts. These effects are depicted in figure 8, with full numerical data in table 4. Complete results of the regression are provided in appendix 2A.

Broadly comparable effects were observed in participant Jayden, including a significantly higher DEI in velar than alveolar target contexts ($\beta = 0.24$; $SE = 0.02$, $t = 10.7$; $p < 0.001$). In Jayden's case, the effect of vowel could not be examined due to the large number of tokens discarded as a consequence of poor image quality. There was a significant interaction

⁷The spectral moments of Max's perceptually accurate velars were consistent with expectations from the children who produced a typical contrast, with the one exception that accurate velars produced in Session 1 were very similar to fronted velars and alveolars with respect to the second spectral moment, slope.

between target and position ($\beta = -0.08$; $SE = 0.03$, $t = -2.4$; $p = 0.02$), reflecting a more extreme difference in DEI between velar and alveolar targets in final than initial position. These effects can be visualized in figure 9, with full numerical data in table 4 and complete regression results available in appendix 2B. Taken together, these results accord with Zharkova (2013) in finding that the DEI measure can distinguish alveolar and velar stops in speakers in this age range. In addition, the numerical values observed are roughly consistent with the DEI values that Zharkova described as typical for velar and alveolar stops (0.5 and 0.26, respectively).

Covert contrast—As in the acoustic analyses above, only target alveolars and fronted velars were included in the statistical analyses of ultrasound measures for children with velar fronting. In the case of Rory, who exhibited across-the-board velar fronting, no fixed effect emerged as a significant predictor of DEI. Ultrasound measures for Rory are depicted in figure 10, with full numerical data in table 4. Complete results of the regression are provided in Appendix 2C. Final alveolars in the back vowel context could not be included in this analysis due to data loss: the full tongue surface was not visible during Rory's block of *mud* productions.

For Max, who exhibited positional velar fronting, position could not be included as a predictor due to the absence of fronted velars in syllable-final contexts. In the data from session 1, velar versus alveolar target was not a significant predictor of DEI ($\beta = -0.005$; $SE = 0.04$, $t = -0.13$; $p = 0.9$). There was a significant effect of vowel context ($\beta = -0.06$; $SE = 0.02$, $t = -2.6$; $p = 0.01$), indicating that dorsum elevation was higher overall in the context of a back relative to a front vowel. There was also a significant target by vowel interaction ($\beta = 0.13$; $SE = 0.04$; $t = 3.3$; $p < 0.001$). Figure 11A shows that in the front vowel context, velar targets were produced with substantially higher DEI than alveolar targets, suggestive of a covert contrast. In the back vowel context, no difference in DEI is apparent.

In his second session, Max produced perceptually accurate velar targets not only in final position, but also in initial position before a back vowel (*cup*). The model for this session thus could not include either an effect of position or an interaction of target place and vowel context. In the reduced model, there was a significant main effect of target place of articulation: velar targets were associated with a slightly higher DEI than alveolar targets ($\beta = 0.05$; $SE = 0.02$; $t = 2.1$; $p = 0.04$). The effect of vowel context was not significant in this model. Results of Max's session 2 are depicted in Figure 11B. Full numerical data for both sessions are reported in table 4, and complete regression results are provided in appendix 2D-E.

Discussion

Comparison of ultrasound and acoustic measures

The present results can be summarized into three major observations. First, velar and alveolar stops produced by the two children who made a perceptually overt contrast were found to differ with respect to all acoustic measures (VOT and the four spectral moments of the stop burst) and the ultrasound measure (DEI). Second, velar and alveolar stops produced by one child with across-the-board velar fronting did not differ with respect to either

acoustic measures or DEI. The final participant, who exhibited positional velar fronting in his first session and additionally produced perceptually correct velars in initial position before a back vowel in his second session, showed some evidence of covert contrast between outputs that were perceptually neutralized. No covert contrast was found in the acoustic parameter of VOT or in the first three spectral moments. A significant difference was found in the fourth spectral moment (kurtosis) in the first recording session only; this predictor was also found to interact with vowel context. For DEI, there was a significant difference between fronted velars and target alveolars in the second session, and there was a significant interaction between target place and vowel context in the first session. A summary of significant findings across both acoustic and ultrasound measures is reported in table 5.

To further examine the relative utility of different correlates of place contrast in child speech, we focus on the results from participant Max, who did show evidence of covert contrast in one acoustic parameter (the fourth spectral moment, kurtosis) and in the ultrasound measure DEI. These two measures found covert contrast in different sessions: the difference in kurtosis was significant in the first but not the second session, and vice versa for the difference in DEI. The measures did agree in finding a significant interaction between target place and vowel context in the first session, but they did not agree in the direction of the interaction. With respect to kurtosis, velar and alveolar targets showed a much larger difference in the back vowel context than with front vowels; by contrast, the DEI data from the same session showed a sizable difference between velar and alveolar targets in the front vowel context and none in the environment of front vowels. We can compare these two measures with respect to their success in predicting the emergence of overt contrast: based on the findings from Tyler et al. (1990), we might expect perceptually overt contrast to emerge in the environment with covert contrast before appearing in the context with no contrast. By this measure, the fourth spectral moment appears to outperform DEI, since it correctly predicts that overt contrast between alveolar and velar targets should appear in the back vowel context before the front vowel context.

However, the present sample is too limited to draw any firm conclusions about the relative validity of acoustic versus ultrasound measures of covert contrast. In particular, the current study did not provide a clear picture of the influence of vowel context on either acoustic or ultrasound parameters, with no coherent pattern emerging across the children sampled. This is in part because the full range of vowel contexts was not represented for all participants due to data loss; future studies would benefit from more extensive efforts to obtain usable data across all vowel contexts. Follow-up research comparing ultrasound and acoustic measures, particularly the spectral moments of the stop burst, could make a valuable contribution to our understanding of the emergence of velar-alveolar place contrasts in both typical and disordered speech development.

Conclusions

Sensitive measures to detect covert contrast are essential for a complete understanding of both typical and disordered speech development. The present study aimed to improve the detection of covert contrast in a commonly occurring child error process, velar fronting. Previous studies using acoustic measures to look for covert contrast between children's

alveolar and fronted velar sounds have returned inconsistent results. It is possible that an articulatory measure could provide a more direct index of this major place contrast. The present study hypothesized that an ultrasound tongue shape measure, Dorsum Excursion Index (DEI; Zharkova, 2013) might provide a more sensitive indicator of covert contrast in perceptually neutralized velar and alveolar targets produced by preschool-aged children with velar fronting.

In the present study, velar and alveolar targets produced by children with perceptually overt contrast were found to differ significantly with respect to both acoustic parameters (VOT, spectral moments) and the ultrasound measure DEI. Furthermore, evidence of covert contrast was found in the speech of one out of two children with velar fronting. For this child, both an acoustic measure (kurtosis) and DEI were found to differ significantly between target alveolars and fronted velars. However, the two measures were not in agreement with respect to the session in which they detected covert contrast, nor with respect to the influence of vowel context. In the present case, the spectral burst data appear to be more informative than the DEI measure, since the covert contrast in kurtosis was observed in Session 1 in the same context that had achieved a perceptually overt distinction by Session 2. Thus, these findings did not support the hypothesis that DEI would represent a more sensitive index of covert contrast than acoustic measures.

Since the present study did not provide compelling evidence of an advantage for ultrasound over acoustic measures of covert contrast, researchers and clinicians might question whether it is possible to justify the time and resources needed to incorporate ultrasound technology into the study or treatment of child speech. However, this study was limited to a comparison of quantitative measures extracted from ultrasound and acoustic data. Other research has suggested that trained individuals can have a high degree of success in categorizing ultrasound tongue shapes by visual gestalt (Klein et al., 2013). This raises the possibility that trained, experienced clinicians could derive diagnostically useful information from qualitative inspection of ultrasound tongue shapes produced by children with velar fronting or other lingual errors. Further, the clinical utility of ultrasound imaging is not limited to diagnosis; a number of studies have documented the efficacy of ultrasound biofeedback treatment for speech sound errors in children (e.g., Preston, Brick, & Landi, 2011; Preston et al., 2014; McAllister Byun, Swartz, & Hitchcock, 2014), including for velar fronting (Cleland, Scobbie, & Wrench, 2015). If the cost of ultrasound technology continues to decline and this evidence base continues to grow, many more clinicians may have access to ultrasound imaging as a tool of clinical practice. Further documentation of the relative efficacy of ultrasound as a tool for diagnosis and/or treatment of speech sound errors in children is an important goal for future research.

Acknowledgments

This paper benefited from contributions by the following student assistants: Reim Farag, Rachel Kadison, Emily Lerner, and Michelle Marron. We thank Doug Whalen for helpful input on the design of the study. Finally, we are grateful to our participants and their families for their cooperation throughout the study. Aspects of this research were presented at the annual convention of the American Speech-Language Hearing Association in Orlando (2014).

This project was supported by an NYU Steinhardt IDEA grant and by NIH R03DC 012883 and NIH R01DC 013668.

Appendix 1. Coefficients, standard errors, and significance levels for all regressions analyzing spectral moments

1A. Data combined across Emma and Jayden (perceptually overt contrast)

	M1	M2	M3	M4
(Intercept)	1079.91*** (144.86)	930.31*** (100.16)	1.11 (0.62)	2.23 (7.28)
Target (Velar)	-467.34* (180.68)	-326.95* (124.92)	2.74*** (0.77)	24.94** (9.08)
Vowel (Front)	523.24* (236.56)	47.87 (163.56)	-0.30 (1.01)	-1.91 (11.88)
Position (Initial)	221.84 (196.15)	42.45 (135.62)	-0.10 (0.84)	-1.08 (9.85)
Target × Vowel	-291.91 (287.42)	-131.54 (198.73)	-0.39 (1.23)	-5.93 (14.44)
Target × Position	-249.90 (246.22)	-95.42 (170.24)	0.51 (1.05)	-1.08 (12.37)
Vowel × Position	56.02 (297.39)	160.48 (205.62)	-0.53 (1.27)	0.29 (14.94)
Target × Vowel × Position	499.95 (367.58)	13.77 (254.16)	-1.73 (1.57)	-11.27 (18.46)

1B. Rory (velar fronting)

	M1	M2	M3	M4
(Intercept)	1106.95*** (199.79)	893.48*** (178.14)	0.80 (0.40)	0.42 (1.90)
Target (Velar)	-2.27 (282.55)	96.68 (251.92)	0.21 (0.57)	0.05 (2.69)
Vowel (Front)	-234.15 (282.55)	-60.96 (251.92)	1.23* (0.57)	5.84* (2.69)
Position (Initial)	-41.67 (223.38)	80.23 (199.16)	0.33 (0.45)	0.45 (2.13)
Target × Vowel	231.81 (357.40)	-132.31 (318.66)	-0.95 (0.72)	-3.09 (3.40)
Target × Position	62.10 (312.73)	-293.69 (278.83)	-0.54 (0.63)	0.70 (2.98)

1A. Data combined across Emma and Jayden (perceptually overt contrast)

	M1	M2	M3	M4
Vowel × Position	308.60 (312.73)	-58.13 (278.83)	-1.39* (0.63)	-5.01 (2.98)
Target × Vowel × Position	-297.02 (399.23)	325.28 (355.95)	1.60 (0.81)	3.36 (3.80)

1C. Max (velar fronting), time 1

	M1	M2	M3	M4
(Intercept)	1449.04*** (269.17)	960.35*** (134.48)	0.64 (0.59)	1.36 (1.39)
Target (Velar)	-449.65 (297.58)	-115.49 (148.68)	0.93 (0.66)	3.68* (1.54)
Vowel (Front)	10.14 (297.58)	130.13 (148.68)	-0.01 (0.66)	-0.65 (1.54)
Target × Vowel	376.81 (347.50)	95.16 (173.62)	-0.84 (0.77)	-3.78* (1.79)

1D. Max (velar fronting), time 2

	M1	M2	M3	M4
(Intercept)	1416.86*** (187.34)	885.19*** (80.76)	0.58 (0.30)	1.43 (1.14)
Target (Velar)	298.79 (316.27)	41.28 (136.33)	-0.39 (0.50)	-1.71 (1.92)
Vowel (Front)	-19.27 (236.97)	283.99* (102.15)	0.48 (0.37)	0.38 (1.44)

Appendix 2. Full results of mixed models analyzing DEI**2A. Emma (perceptually overt contrast)**

	<i>Estimate</i>	<i>SE</i>	<i>Df</i>	<i>t</i>	<i>p</i>
Intercept	0.23	0.01	29.97	15.96	0
Target (Velar)	0.25	0.02	68.76	10.73	0
Vowel (Front)	0.07	0.01	68.42	4.44	0
Position (Initial)	0.04	0.01	67.79	2.4	0.02
Target × Vowel	-0.19	0.02	68	-7.91	0
Target × Position	0.05	0.02	61.42	1.98	0.05

2A. Emma (perceptually overt contrast)					
	<i>Estimate</i>	<i>SE</i>	<i>Df</i>	<i>t</i>	<i>p</i>
2B. Jayden (perceptually overt contrast)					
	<i>Estimate</i>	<i>SE</i>	<i>df</i>	<i>t</i>	<i>p</i>
Intercept	0.24	0.02	27.94	15.09	0
Target (Velar)	0.24	0.02	27.94	10.72	0
Position (Initial)	-0.02	0.02	27.94	-0.93	0.36
Target × Position	-0.08	0.03	27.94	-2.38	0.02
2C. Rory (velar fronting)					
	<i>Estimate</i>	<i>SE</i>	<i>df</i>	<i>t</i>	<i>p</i>
Intercept	0.3	0.04	27.31	7.98	0
Target (Velar)	0	0.04	28.54	-0.08	0.94
Vowel (Front)	0.01	0.03	39.02	0.42	0.68
Position (Initial)	0.02	0.03	39.98	0.66	0.51
Target × Vowel	0.03	0.04	22.71	0.73	0.48
Target × Position	-0.05	0.03	41.61	-1.38	0.17
2D. Max (velar fronting), time 1					
	<i>Estimate</i>	<i>SE</i>	<i>df</i>	<i>t</i>	<i>p</i>
Intercept	0.23	0.02	36.92	12.27	0
Target (Velar)	0	0.04	36.92	-0.13	0.9
Vowel (Front)	-0.06	0.02	36.92	-2.61	0.01
Target × Vowel	0.13	0.04	36.92	3.33	0
2E. Max (velar fronting), time 2					
	<i>Estimate</i>	<i>SE</i>	<i>df</i>	<i>t</i>	<i>p</i>
Intercept	0.28	0.02	41.92	15.9	0
Target (Velar)	0.05	0.02	41.92	2.09	0.04
Vowel (Front)	-0.01	0.02	41.92	-0.36	0.72

References

- Amorosa H, Benda Uv, Wagner E, Keck A. Transcribing phonetic detail in the speech of unintelligible children: A comparison of procedures*. *International Journal of Language & Communication Disorders*. 1985; 20(3):281–287.
- Boersma, P.; Weenink, D. Praat: Doing phonetics by computer. Version 5.2.46. 2011. retrieved 7 October 2011 from <http://www.praat.org/>
- Cleland J, Scobbie JM, Wrench AA. Using ultrasound visual biofeedback to treat persistent primary speech sound disorders. *Clinical Linguistics & Phonetics*. 2015;1–23. early online.
- Davis K. Phonetic and phonological contrasts in the acquisition of voicing: Voice onset time production in Hindi and English. *Journal of Child Language*. 1995; 22:275–305. [PubMed: 8550724]
- Edwards J, Gibbon F, Fourakis M. On discrete changes in the acquisition of the alveolar/velar stop consonant contrast. *Language and Speech*. 1997; 40(2):203–210.
- Forrest K, Weismer G, Hodge M, Dinnsen DA, Elbert M. Statistical analysis of word-initial/k/and/t/ produced by normal and phonologically disordered children. *Clinical Linguistics & Phonetics*. 1990; 4(4):327–340.

- Gibbon F. 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 & Communication Disorders*. 1990; 25(3):329–340.
- Gibbon F. Undifferentiated lingual gestures in children with articulation/phonological disorders. *Journal of Speech, Language, and Hearing Research*. 1999; 42(2):382–397.
- Gibbon F, Dent H, Hardcastle W. Diagnosis and therapy of abnormal alveolar stops in a speech-disordered child using electropalatography. *Clinical Linguistics & Phonetics*. 1993; 7(4):247–267.
- Grunwell, P. The nature of phonological disability in children. London: Academic Press; 1981.
- Hitchcock ER, Koenig LL. The effects of data reduction in determining the schedule of voicing acquisition in young children. *Journal of Speech, Language, and Hearing Research*. 2013; 56(2): 441–457.
- Julien HM, Munson B. Modifying speech to children based on their perceived phonetic accuracy. *Journal of Speech, Language, and Hearing Research*. 2012; 55(6):1836–1849.
- Klatt DH. Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech, Language, and Hearing Research*. 1975; 18(4):686–706.
- Klein HB, McAllister Byun T, Davidson L, Grigos MI. A multidimensional investigation of children's /r/ productions: Perceptual, ultrasound, and acoustic measures. *American Journal of Speech-Language Pathology*. 2013; 22:540–553. [PubMed: 23813195]
- Li M, Kambhamettu C, Stone M. Automatic contour tracking in ultrasound images. *Clinical Linguistics and Phonetics*. 2005; 19(6-7):545–554. [PubMed: 16206482]
- Lindblom B. Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America*. 1963; 35(11):1773–1781.
- Lisker L, Abramson AS. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*. 1964; 20:384–422.
- Lisker L, Abramson AS. Some effects of context on voice onset time in English stops. *Language and Speech*. 1967; 10(1):1–28. [PubMed: 6044530]
- Lousada ML, Jesus LM, Pape D. Estimation of stops' spectral place cues using multitaper techniques. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*. 2012; 28(1):1–26.
- Macken MA, Barton D. The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. *Journal of child language*. 1980; 7(01):41–74. [PubMed: 7372738]
- McAllister Byun T. Positional velar fronting: An updated articulatory account. *Journal of child language*. 2012; 39(05):1043–1076. [PubMed: 22225837]
- McAllister Byun T, Hitchcock ER, Swartz MT. Retroflex versus bunched in treatment for /r/ misarticulation: Evidence from ultrasound biofeedback intervention. *Journal of Speech, Language, and Hearing Research*. 2014; 57(6):2116–2130.
- Ménard L, Aubin J, Thibeault M, Richard G. Measuring tongue shapes and positions with ultrasound imaging: a validation experiment using an articulatory model. *Folia Phoniatr Logop*. 2012; 64(2): 64–72. [PubMed: 22212175]
- Munson B, Schellinger SK, Urberg-Carlson K. Measuring speech-sound learning using visual analog scaling. *Perspectives in Language Learning and Education*. 2012; 19:19–30.
- Preston JL, Brick N, Landi N. Ultrasound biofeedback treatment for persisting childhood apraxia of speech. *American Journal of Speech-Language Pathology*. 2013; 22(4):627–643. [PubMed: 23813207]
- Preston JL, McCabe P, Rivera-Campos A, Whittle JL, Landry E, Maas E. Ultrasound visual feedback treatment and practice variability for residual speech sound errors. *Journal of Speech, Language and Hearing Research*. 2014; 57:2102–2115.
- Repp BH. Closure duration and release burst amplitude cues to stop consonant manner and place of articulation. *Language and Speech*. 1984; 27(3):245–254. [PubMed: 6521579]
- Scobbie, JM. Interactions between the acquisition of phonetics and phonology. In: Gruber, K.; Higgins, D.; Olsen, K.; Wysochi, T., editors. *Papers from the 34th Annual Regional Meeting of the Chicago Linguistic Society*. Vol. 2. Chicago: Chicago Linguistics Society; 1998. p. 343-358.

- Scobbie, J.; Gibbon, F.; Hardcastle, W.; Fletcher, P. Covert contrast as a stage in the acquisition of phonetics and phonology. In: Broe, M.; Pierrehumbert, J., editors. *Papers in Laboratory Phonology V*, 194-207. Cambridge, UK: Cambridge University Press; 2000.
- Slud E, Stone M, Smith P, Goldstein M. Principal components representation of the two-dimensional coronal tongue surface. *Phonetica*. 2002; 59:108–133. [PubMed: 12232463]
- Sussman HM, Hoemeke KA, McCaffrey HA. Locus equations as an index of coarticulation for place of articulation distinctions in children. *Journal of Speech, Language, and Hearing Research*. 1992; 35(4):769–781.
- Tyler AA, Edwards ML, Saxman JH. Acoustic validation of phonological knowledge and its relationship to treatment. *Journal of Speech and Hearing Disorders*. 1990; 55(2):251–261. [PubMed: 2329787]
- Tyler AA, Figurski GR, Langsdale T. Relationships between acoustically determined knowledge of stop place and voicing contrasts and phonological treatment progress. *Journal of Speech, Language, and Hearing Research*. 1993; 36(4):746–759.
- Whalen DH, Iskarous K, Tiede MK, Ostry DJ, Lehnert-LeHouillier H, Vatikiotis-Bateson E, Hailey DS. The Haskins optically corrected ultrasound system (HOCUS). *Journal of Speech, Language, and Hearing Research*. 2005; 48(3):543–553.
- Young EC, Gilbert HR. An analysis of stops produced by normal children and children who exhibit velar fronting. *Journal of Phonetics*. 1988; 16(2):243–246.
- Zharkova N. Using ultrasound to quantify tongue shape and movement characteristics. *The Cleft Palate-Craniofacial Journal*. 2013; 50(1):76–81. [PubMed: 22117937]

Figure 1a

Figure 1b

Figure 1.

Tongue shapes representing a perceptually typical alveolar-velar contrast produced by speaker Emma, age 4;2. The right side of the image is anterior.

- a. Sagittal view of the tongue shape during maximal closure for perceptually typical /t/ in *T*.
- b. Sagittal view of the tongue shape during maximal closure for perceptually typical /k/ in *key*.

Figure 2.

Trace of a typical adult's tongue shape during maximal closure for /g/ in *mug*, (Zharkova, unpublished R script). N represents the distance from the most posterior to the most anterior point of the tongue contour. D is a line segment perpendicular to N from the midpoint of N to the surface of the tongue.

Figure 3a **Figure 3b****Figure 3.**

(a) Sagittal view of Emma's tongue shape during maximal closure for /t/ in T, with the tongue contour tracked by EdgeTrak (red dots). (b) The extracted and annotated tongue contour for the same production (Zharkova, unpublished R script). This tongue contour has a DEI value of 0.35.

Figure 4a **Figure 4b****Figure 4.**

(a) Sagittal view of Emma's tongue shape during maximal closure for /k/ in *key*, with the tongue contour tracked by EdgeTrak (red dots). (b) The extracted and annotated tongue contour for the same production (Zharkova, unpublished R script). This tongue contour has a DEI value of 0.43.

Figure 5. VOT means for initial velar and alveolar stop targets for each participant and session. Children who produce an overt contrast and children with velar fronting are plotted separately. Bars represent one standard deviation around the mean.

Figure 6.

Mean value for spectral moments 1-4 calculated for alveolar versus velar stops produced by children with a perceptually overt contrast. Data are partitioned by child and by front versus back vowel contexts. Error bars represent one standard deviation.

Figure 7.

Mean value for spectral moments 1-4 calculated for alveolar versus velar stops produced by children with velar fronting. Data are partitioned by child and by front versus back vowel contexts. Error bars represent one standard deviation.

Figure 8.

Mean value for DEI calculated for alveolar versus velar stops produced with perceptually overt contrast by participant Emma. Data are partitioned by position and by front versus back vowel contexts. Error bars represent one standard deviation.

Figure 9.

Mean value for DEI calculated for alveolar versus velar stops produced with perceptually overt contrast by participant Jayden. Data are partitioned by position; only back vowel contexts are represented due to data loss. Error bars represent one standard deviation.

Figure 10.

Mean value for DEI calculated for alveolar versus velar stops produced by velar fronting participant Rory. Data are partitioned by front versus back vowel contexts and initial versus final position; front vowels in final position are missing due to data loss. Error bars represent one standard deviation.

Figure 11a

Figure 11b

Figure 11. Mean value for DEI calculated for alveolar versus velar stops produced by velar fronting participant Max. Data are partitioned by front versus back vowel contexts. Error bars represent one standard deviation.

- a. Session 1.
- b. Session 2.

Table 1
Characteristics of participants enrolled in the study

Pseudonym	Sex	Age	Velar production	Other relevant history
Max	M	Session 1: 3;6 Session 2: 3;9	Positional fronting	
Rory	M	4;1	Across-the-board fronting	
Jayden	M	4;5	Velar-alveolar contrast	Bilingual (Korean-English)
Emma	F	4;2	Velar-alveolar contrast	

Table 2
VOT values produced for alveolar and velar target sounds in initial position

Group	Child	Output	Mean VOT (ms)	SD of VOT (ms)
Overt Contrast	Emma	Alveolar	91.46	65.2
		Velar (accurate)	168.37	101.81
	Jayden	Alveolar	71.63	63.38
		Velar (accurate)	152.28	94.71
Velar Fronting	Rory	Alveolar	123.65	58.15
		Velar (fronted)	121.62	51.12
	Max, S1	Alveolar	174.6	102.16
		Velar (fronted)	169.89	41.53
	Max, S2	Alveolar	121.26	61.9
		Velar (accurate)	86.5	22.61
		Velar (fronted)	137.4	39.82

Table 3
Spectral moments of the burst (M1—M4) measured for alveolar target sounds and velar target sounds produced with and without velar fronting

Child	Target	mean M1 (SD)	mean M2 (SD)	mean M3 (SD)	mean M4 (SD)
Emma	Alveolar	1545.56 (561.09)	1072.65 (182.17)	0.72 (0.83)	0.95 (2.64)
	Velar (accurate)	975.84 (323.98)	718.28 (234.98)	2.38 (0.81)	8.5 (6.13)
Jayden	Alveolar	1364.1 (457.05)	953.33 (142.56)	0.74 (0.66)	0.35 (1.53)
	Velar (accurate)	788.3 (464.02)	470.58 (199.67)	3.66 (2.47)	27.67 (25.07)
Rory	Alveolar	1085.41 (171.84)	899.45 (160.12)	1.11 (0.43)	1.69 (2)
	Velar (fronted)	1121.25 (189.17)	822.9 (170.47)	1.14 (0.46)	2.42 (2.06)
Max, Session 1	Alveolar	1457.34 (463.83)	1066.83 (203.45)	0.63 (0.76)	0.82 (1.11)
	Velar (accurate)	812.81 (437.7)	904.4 (363.27)	2.96 (1.5)	12.56 (11.34)
	Velar (fronted)	1192.87 (355.23)	957.51 (209.55)	1.14 (0.95)	2.82 (3.19)
Max, Session 2	Alveolar	1404.82 (571.84)	1062.68 (285.83)	0.89 (0.94)	1.67 (3.53)
	Velar (accurate)	531.09 (161.27)	609.84 (248.82)	4.29 (1.68)	26.41 (20.97)
	Velar (fronted)	1696.38 (355.01)	1210.45 (118.16)	0.68 (0.46)	0.1 (1.12)

Table 4
DEI values measured for alveolar and velar target sounds produced by each child, subdivided by position and vowel context

Child	Target	Position	Vowel Context	Mean DEI(SD)	Count of tokens
Emma	Alveolar	Initial	Front	0.35 (0.03)	6
			Back	0.25 (0.03)	7
		Final	Front	0.27 (0.04)	5
			Back	0.26 (0.04)	4
	Velar (accurate)	Initial	Front	0.44 (0.06)	7
			Back	0.56 (0.06)	3
		Final	Front	0.35 (0.05)	4
			Back	0.49 (0.06)	2
Jayden	Alveolar	Initial	Back	0.21 (0.04)	3
		Final	Back	0.24 (0.04)	5
	Velar (accurate)	Initial	Back	0.37 (0.06)	4
		Final	Back	0.47 (0.04)	5
Rory	Alveolar	Initial	Front	0.33 (0.03)	4
			Back	0.31 (0.03)	3
		Final	Front	0.31 (0.07)	2
	Velar (fronted)	Initial	Front	0.29 (0.07)	4
			Back	0.28 (0.05)	5
		Final	Front	0.34 (0.02)	4
			Back	0.28 (0.04)	3
	Max, Session 1 Alveolar	Initial	Front	0.19 (0.03)	3
			Back	0.26 (0.04)	2
		Final	Front	0.16 (0.04)	5
			Back	0.16 (0.03)	1
	Velar (accurate)	Final	Front	0.33 (0.08)	4
			Back	0.4 (0.02)	2
	Velar (fronted)	Initial	Front	0.3 (0.05)	9
			Back	0.22 (0.04)	1

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Child	Target	Position	Vowel Context	Mean DEI(SD)	Count of tokens
Max. Session 2	Alveolar	Initial	Front	0.29 (0.05)	10
			Back	0.28 (0.04)	3
		Final	Front	0.21 (0.03)	4
			Back	0.27 (0.07)	3
	Velar (accurate)	Initial	Back	0.54 (0.04)	3
			Front	0.4 (0.08)	4
		Back	Back	0.32 (0.05)	2
			Front	0.31 (0.03)	4

Table 5

Summary of significant contrasts found in acoustic (VOT, M1-M4) and ultrasound (DEI) measures for each participant. Direction of the effect in parentheses.

	Overt contrast		No overt contrast	
	Emma	Jayden	Rory	Max, Session 1 Max, Session 2
VOT	Target place (velar > alveolar)			
M1	Target place (alveolar > velar)	Vowel context (front > back)		
M2	Target place (alveolar > velar)			Vowel context (front > back)
M3	Target place (velar > alveolar)		Vowel context (front > back)	
M4	Target place (velar > alveolar)		Vowel context (front > back)	Target place (velar > alveolar) Vowel context (back > front) Target-vowel interaction (greater difference in back than front vowel context)
DEI	Target place (velar > alveolar) Vowel context (front > back) Position (initial > final) Target-vowel interaction (greater difference in back than front vowel context)	Target place (velar > alveolar) Target-position interaction (greater difference in final than initial position)	Vowel context (back > front) Target-vowel interaction (greater difference in front than back vowel context)	Target place (velar > alveolar)