CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Using ultrasound tongue imaging to identify covert contrasts in children's speech
Authors	Zharkova, Natalia;Gibbon, Fiona E.;Lee, Alice S.
Publication date	2017
Original Citation	Zharkova, N., Gibbon, F. E. and Lee, A. (2017) 'Using ultrasound tongue imaging to identify covert contrasts in children's speech', Clinical Linguistics and Phonetics, 31(1), pp. 21-34. doi: 10.1080/02699206.2016.1180713
Type of publication	Article (peer-reviewed)
Link to publisher's version	https://www.tandfonline.com/doi/full/10.1080/02699206.2016.1180713 - 10.1080/02699206.2016.1180713
Rights	© 2017, Taylor & Francis. This is an Accepted Manuscript of an article published by Taylor & Francis in Clinical Linguistics and Phonetics, on 20 June, 2016, available online: https://www.tandfonline.com/10.1080/02699206.2016.1180713. It is deposited under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. - https://creativecommons.org/licenses/by-nc/4.0/
Download date	2026-11-17 06:35:36
Item downloaded from	https://hdl.handle.net/10468/6545

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

This is an Accepted Manuscript of an article published by Taylor & Francis Group in Clinical Linguistics & Phonetics on 20/06/2016, available online: http://www.tandfonline.com/doi/full/10.1080/02699206.2016.1180713.

Using ultrasound tongue imaging to identify covert contrasts in children's speech

Natalia Zharkova
Clinical Audiology, Speech and Language Research Centre
Queen Margaret University, Edinburgh, UK

Fiona E. Gibbon
Department of Speech and Hearing Sciences
University College Cork, Cork, Ireland

Alice Lee
Department of Speech and Hearing Sciences
University College Cork, Cork, Ireland

Correspondence to:
Natalia Zharkova
Clinical Audiology, Speech and Language Research Centre
Queen Margaret University
Queen Margaret University Drive
Musselburgh EH21 6UU
East Lothian
United Kingdom

Tel.: +44 (0) 131 4740000
Fax: +44 (0) 131 4740001
Email: nzharkova@qmu.ac.uk

Abstract

Ultrasound tongue imaging has become a promising technique for detecting covert contrasts, due to the developments in data analysis methods that allow for processing information on tongue shape from young children. An important feature concerning analyses of ultrasound data from children who are likely to produce covert contrasts is that the data are likely to be collected without head-to-transducer stabilisation, due to the speakers' age. This paper is a review of existing methods applicable to analysing data from non-stabilised recordings. The paper describes some of the challenges of ultrasound data collection from children, and analysing these data, as well as possible ways to address those challenges. Additionally, there are examples from typical and disordered productions featuring covert contrasts, with illustrations of quantifying differences in tongue shape between target speech sounds.

keywords that are not in the title: child, speech production measurement, lingual articulation, tongue shape, subtle differences

Introduction

The term covert contrast describes a phenomenon whereby instrumentally measurable differences are detected between target phonemes that are apparently neutralised in listeners' perceptions (Maken and Barton, 1980; Hewlett, 1988). Previous studies have mostly used acoustic analysis in order to identify covert contrasts (for reviews, see Scobbie et al., 2000, and Li et al., 2009). Methodologies involving direct recording of articulator movements have been used in fewer studies (e.g., 20% of 30 studies reviewed in Gibbon, 2002). Because the populations likely to produce covert contrasts are often young children and/or speakers who have disordered speech, it may be logistically challenging to employ articulatory techniques. However, direct articulatory measurements can be helpful in demonstrating how phonological contrasts are realised despite the lack of perceptual difference (Gibbon and Crampin, 2001; Gibbon, 2003). Compared to acoustic analysis, an advantage of using articulatory techniques to detect covert contrast is that the latter provide much more specific information about the features of articulation underlying the production of covert contrasts, for example, greater details about the place of articulation, tongue position, or groove width (e.g., Gibbon and Wood, 2003). One of the promising articulatory techniques for investigating covert contrasts is ultrasound tongue imaging.

Ultrasound has been used in linguistic phonetic research for a few decades, and it has also been increasingly applied as a visual biofeedback technique in speech and language therapy (see Cleland et al., 2016). The technique provides information on most of the tongue contour between the tip and the root of the tongue, making it possible to quantify subtle changes in tongue shape between two different speech sounds. More specifically, ultrasound data could potentially identify covert contrasts that underlie frequently occurring phonological processes, such as gliding, fronting velars, backing alveolars, stopping fricatives and affricates, cluster reduction, and fronting fricatives. Critical to identification of covert

contrasts in articulatory data is having robust measurements that reveal differences between target sounds. Quantitative measures using ultrasound data have the potential to identify articulatory patterns than cannot be described with acoustic analysis, and to provide information that would not be available from other physiological methods such as electropalatography. To the authors' knowledge, ultrasound has only been used in one published study to date to identify covert contrasts in children (McAllister Byun et al., 2015). However the number of ultrasound studies of speech production in children has been steadily growing over the last decade, and methodological advances have led to younger age groups being targeted in research studies (e.g., Song et al., 2013; Magloughlin, 2016; Abakarova et al., 2015; Rubertus et al., 2015; Yip et al., 2015), as well as using quantification methods that could be employed to identify covert contrasts (e.g., Ménard and Noiray, 2011; Noiray et al., 2013; Klein et al. 2013; Zharkova et al., 2015a, 2015b).

The present paper is a review of available ultrasound-based measures suitable for studying covert contrasts, describing some basic considerations that apply to studying covert contrasts with ultrasound tongue imaging, and providing illustrations of covert contrasts using ultrasound data, with example measurements. The paper is targeted at phoneticians and clinical researchers who would like to do, or are already doing, ultrasound speech analysis of covert contrasts in young children. While some background information on ultrasound imaging is provided, for further technical details on the technique the readers are referred to Stone (2005, 2010), and Lee et al. (2013).

The basics of ultrasound for tongue imaging

Ultrasound tongue imaging makes it possible to view and record tongue movements in real time. Research studies of speech production with ultrasound typically use two-dimensional scanning. Midsagittal or coronal plane images can be recorded, though in practice most

studies to date have used midsagittal scans (see a review in Lee et al., 2013; for examples of studies using coronal images, see Stone, 1990; Stone et al., 1992; Bressmann et al., 2005). For imaging the tongue, the transducer, which emits high frequency waves and receives echoes when the sound is reflected back to the source, is placed beneath the speaker's chin. Figure 1 shows an ultrasound recording of a 3-year-old child, with the participant sitting in the lap of his carer, and the transducer held by the experimenter.

Insert figure 1 about here

Examples of midsagittal ultrasound images of the tongue can be found in figure 2, with the bright white line corresponding to the interface of the tongue surface with the air above it. The tongue outlines at mid-/s/ from /sa/ and mid-/ʃ/ from /ʃa/, from productions by a typically developing boy aged 3;8 [years;months], are presented in the left and the right panel, respectively. Adjacent to the end of each curve on the right is the acoustic shadow of the mandible, while at the other end of the tongue curve there is the acoustic shadow of the hyoid bone. A difference in tongue shape between the two consonants is noticeable, in that the most excursed part of the tongue is further forward along the tongue curve for the postalveolar consonant /ʃ/, reflecting the tongue predorsum raising towards the hard palate.

The palate and the pharyngeal walls are not imaged with ultrasound during speech. The hard palate can be imaged if the speaker has water in the mouth, or during a swallow. The shadows of the hyoid bone and the mandible can partly obscure the tongue root and the tongue tip. Besides, the tip is sometimes not imaged due to the air beneath it, making it challenging to compare tongue shapes where key differences may only be in the tongue tip position.

Insert figure 2 about here

The advantages of ultrasound over other articulatory methods include its relatively low cost, as well as the possibility to obtain interpretable data without the need for the speaker to be enclosed in a scanner (such as with magnetic resonance imaging), or to wear an intraoral device (such as with electropalatography). The fact that ultrasound images of the tongue typically do not contain information on other articulatory organs has implications for studies of covert contrasts. Any contrasts may not always be identified from ultrasound data, such as when physiological differences between target realisations are based on lip, larynx or velum movements. The limitations of ultrasound described above need to be taken into account when planning research studies or treatment. This particularly applies to studying fine phonetic details in the speech of very young children. The next section concentrates on specific technical challenges involved with recording small children, and implcations for quantitative analysis.

Collecting ultrasound data from young children

Any studies of child speech using ultrasound might have data collected without head-to-transducer stabilisation, and therefore may need to meet the challenge of analysing such data. This challenge is likely to apply to ultrasound studies focusing on covert contrasts (see an example in McAllister Byun et al., 2015). Although fixing a headset is not harmful to participants, it is problematic to attempt to do so with young children because they are unlikely to tolerate its presence for the duration of a speech recording session, due to the headset's weight and the requirement that the headset tightly fits the head (an illustration of a headset can be found in Zharkova et al., 2015a). While avoiding the cumbersome process of head stabilisation (see Stone, 2005) is undoubtedly a bonus when recording tongue

movements from young children, the resulting data consist of tongue curves which are not located in the same coordinate space. Every small movement of the transducer in relation to the head, for example, due to the speaker's head turn, or the experimenter's hand slipping under the chin, affects the picture reflected in the field of view of the transducer. Consequently, different tongue curves recorded from the same speaker, even within a single session, cannot be quantitatively analysed in absolute coordinates in relation to each other, unless the recording includes capturing external landmarks enabling post-processing of the signal (cf. Whalen et al., 2005; Mielke et al., 2005; Noiray et al., 2015). If no such techniques are employed, then the scope of possible analysis measures is limited to those capable of extracting quantitative information from a single tongue curve. Such information can then be compared across multiple curves, whether from contrasting speech targets, from different time points within the same token (such as onset versus offset of a segment or a syllable), or from different recording sessions of the same speaker.

Even though during recordings without head-to-transducer stabilisation there is a possibility to image the hard palate, without the post-processing described above, it is not possible to place the palate image in the same reference frame as tongue images. Consequently, due to the lack of visible reference structures external to the tongue in the ultrasound image, the main reference points in studies using non-stabilised data are the ends of the tongue curve that have to be determined by the researcher. Therefore it is important to ensure some consistency in imaging the curve ends during the recording process. If during the recording of midsagittal tongue images the shadow of the mandible and the shadow of the hyoid bone are consistently present in the ultrasound field of view (as in figure 2), they provide some reference points for determining the ends of the curve.

A brief review of existing measures that can be applied to single curves is provided in the next section (for measures based on comparing sets of curves within the same coordinate

system, see, e.g., Davidson, 2006; Zharkova and Hewlett, 2009; Carignan, 2014; Mielke, 2015; Recasens and Rodríguez, 2015).

Analysis of ultrasound data recorded without head-to-transducer stabilisation

This section describes existing measures that can be applied to a single tongue curve, without reference to any landmarks external to the tongue contour. A number provided by each measure characterises one curve, and comparisons using inferential statistical tests are based on repeated productions of each target. Such measures quantify aspects of the curve shape, such as how flat versus "bunched" it is, where the location of bunching is, and whether the curve has more than one location of bunching. For example, the midsagittal tongue shape for the vowel /i/ typically has one very pronounced bunching location towards the front of the tongue, while the tongue for the vowel /a/ tends to be very flat midsagitally, with the location of bunching further back than that for /i/. More than one bunching location can be found in consonants with two lingual constrictions, such as liquids (e.g., Gick et al., 2006, 2008). The measures described below have mostly been designed to apply to midsagittal tongue shapes (the extent of tongue grooving from coronal ultrasound data has been quantified, for example, in Stone et al., 1988, 1992; Bressmann et al., 2005, 2007). Some of the measures could potentially be applied to coronal data as well.

Measures quantifying the extent and location of the most bunched part of the tongue were proposed in Aubin and Ménard (2006). Two measures, Curvature Degree and Curvature Position, are based on fitting a triangle to the tongue shape. The line between the two ends of the curve (e.g., in figure 3 it is the long solid line) serves as a base of the triangle, and the most excursed part of the tongue (in figure 3, the crossing point of the tongue curve and the

other solid line) constitutes the third point of the triangle[1]. Both measures are ratios of straight lines derived from the triangle. Curvature Degree quantifies the extent of maximal excursion within the tongue shape, defined by the ratio of two straight lines (in figure 3, the short solid line divided by the long solid line), with higher values corresponding to more bunched shapes. Curvature Position captures where, along the tongue curve, the excursion takes place (in figure 3, two parts of the long solid line, separated by the crossing with the short solid line, are the numerator and the denominator for the ratio), and higher values represent tongue shapes with more bunching towards the front of the tongue. In Ménard et al. (2012), using an articulatory model and mathematical transformations, these measures were shown to be not affected by rotation and displacement of the transducer. In that study, Curvature Degree and Curvature Position were used to describe tongue shapes of six French vowels generated by an articulatory model. These indices have been used in a number of studies of child and adult speech to characterise tongue shape for consonants and vowels (e.g., Klein et al., 2013; Noiray et al., 2013; Zharkova et al., 2015a; Ménard et al., 2015). Combining the two measures, such as in Klein et al. (2013), may be helpful, as this may provide more complete information on the articulatory mechanisms responsible for producing certain tongue shapes. This applies also to other measures described below, with different indices capable of quantifying complementary aspects of tongue shape.

The performance of Curvature Degree and Curvature Position was compared by Zharkova et al. (2015a) across two conditions, with versus without head-to-transducer stabilisation. The same ten adolescent speakers were recorded producing the same stimuli in

[1] Compare these measures with another way of quantifying tongue curvature, reported in Stone et al. (1987), who used the origin of the ultrasound image, i.e., a transducer-based landmark, as a reference point for curvature calculations using tongue contours from adults' productions. With the reference point for measurements being external to the vocal tract, if such measurements were applied to young children's productions, the values would likely be affected by children's head movements in relation to the transducer during the recording.

those two conditions, and vowel-related effects on tongue shape for the consonants /p/, /t/, /s/ and /ʃ/ were assessed in both conditions. There were different results across the two conditions for Curvature Position with alveolar consonants /t/ and /s/, and for Curvature Degree with /p/. An effect from contrasting vowels on Curvature Degree for /p/ was observed when the transducer was hand-held, but not when the speakers wore the headset. This difference was possibly due to less of the tongue curve imaged without head stabilisation for /p/ in the context of /a/ than when the stabilisation was used, because of less pressure on the chin from the transducer in the hand-held condition. The fact that for the alveolar consonants /t/ and /s/ Curvature Position demonstrated vowel-related coarticulatory effects on the consonants when the speakers wore the stabilising headset, but not when the transducer was hand-held, means that this index is less sensitive to the changes in tongue shape when there is no head-to-transducer stabilisation. Therefore if the index was used with hand-held data on /t/ and /s/, it might not capture small differences in tongue shape that could be responsible for covert contrasts.

While Curvature Degree, in combination with Curvature Position, provide information characterising the overall tongue shape, they do not aim to address the behaviour of specific parts of the tongue, such as the tongue dorsum, which is the principal articulator for producing velar consonants, or the predorsal area involved in forming a tongue-to-palate constriction for postalveolar fricatives. Focussing on areas of the tongue responsible for creating certain phonological contrasts may be important when the aim is to identify covert contrasts involved in producing the target phonemes. A measure of tongue shape called Dorsum Excursion Index was proposed in Zharkova (2013a). This measure quantifies the extent of excursion of the tongue dorsum, conceptually represented by the point on the tongue curve located opposite the middle of the straight line between two curve ends. Unlike Curvature Degree, which shows the extent of bunching in tongue shape wherever it occurs,

Dorsum Excursion Index quantifies the extent of bunching in the middle part of the tongue. Dorsum Excursion Index has been shown to robustly differentiate between tongue shapes for /k/ and several consonants with other places of articulation in Scottish English speaking adults (Zharkova, 2013b), with the highest values for the velar consonant. Also, /k/ was the only consonant in that study not affected on this measure by contrasting vowels /a/ versus /i/. Another measure aimed to describe the tongue shape with reference to the tongue dorsum is Tongue Constraint Position Index (Zharkova, 2013a). For this index, the dorsum is defined in the same way as for Dorsum Excursion Index, described above. Tongue Constraint Position Index assesses the location of the most excursed part of the tongue and returns positive versus negative values depending on whether it is further forward or further back along the tongue curve than the dorsum, respectively. In the study described above, Zharkova et al. (2015a) compared the performance of these two indices in head-stabilised versus non-stabilised conditions, and found different results across the two conditions for Tongue Constraint Position Index with alveolar consonants, and for Dorsum Excursion Index with /p/ and /ʃ/.

For Tongue Constraint Position Index, similarly to Curvature Position described above, there were vowel-related coarticulatory effects on /t/ and /s/ when the speakers wore the headset, but not in the hand-held condition, suggesting that the index might not detect small changes in alveolar consonant tongue shapes in ultrasound data recorded without head-to-transducer stabilisation.

Dorsum Excursion Index was used in McAllister Byun et al. (2015) to compare /k/ and /t/ productions in four American English speaking children aged between 3;6 and 4;5. Two of the children had a /k/-/t/ contrast, and two other children were fronting velar consonants. Higher Dorsum Excursion Index values were found for /k/ than for /t/ in correct productions. Also, using the index, covert contrasts were identified in the productions by one child. The authors' hypothesis that the articulatory index would be more sensitive than

acoustic measures was not confirmed, since Dorsum Excursion Index was not more successful in predicting the vowel context in which an overt contrast would emerge than the fourth spectral moment. The authors explained this finding in part by a rather limited sample in the study, warranting more research comparing articulatory and acoustic measures of velar-alveolar contrasts.

A common limitation of the measures described above is that their calculation requires establishing the length of the line between two curve ends. The length of this line is dependent on assessing the location of the ends of the tongue curve, which, as explained above, may be problematic when the data are based on recordings without head-to-transducer stabilisation. Therefore when analysing ultrasound data from non-head-stabilised recordings, it might be useful to have tongue shape measures not involving the length of the line between curve ends. Such a measure was proposed in Zharkova et al. (2015a). The index, called $LOC_{a\text{-}i}$, quantifies the location of the most excursed part of the tongue in relation to the overall tongue curve. This index, unlike the ones described above, is not directly based on the straight line between the two ends of the tongue curve. $LOC_{a\text{-}i}$ is a ratio of two perpendiculars to that straight line, the "front" one located further forward along the tongue curve, and the "back" one located further back, so that the two perpendiculars divide the line between the curve ends into three equal intervals. Thus, the index compares the extent of excursion of the front versus the back of the tongue. The two perpendiculars used for calculating $LOC_{a\text{-}i}$ can be seen in figure 4, with the further forward located perpendicular being longer than the other perpendicular in both panels. In typical adult and child speech, the index tends to have higher values for consonants in the context of /i/ than in the context of /a/, though this has been shown to differ across consonants and across age groups (Zharkova et al., 2015a, 2015b). Perhaps because this measure is not as strongly related to the line between the ends of the curve, $LOC_{a\text{-}i}$ was the only measure in the study by Zharkova et al. (2015a) that consistently

had the same results on vowel-on-consonant coarticulation, across head-stabilised versus non-stabilised conditions, for all four consonants: /p/, /t/, /s/ and /ʃ/.

The measures described above are suitable for assessing the extent and location of bunching on the tongue with a single main bunching location, when the aim of the analysis does not include capturing the extent of prominence of two bunching locations within a given tongue curve. If there is a reason to expect two distinct constrictions demonstrated as two separate bunching locations, such as with a constriction of the dorsum and another constriction in the pharyngeal area in the English postalveolar approximant, then these indices may not capture the relevant details for a certain overt or covert contrast. Several measures have been designed that are sensitive to changes in the number of bunching locations, and aim to quantify the extent of tongue shape differentiation, or complexity. Gick et al. (2008) described an index that allows for quantifying the level of differentiation in the tongue curve, that can distinguish between tongue shapes with one versus two constrictions. Curvature Index was proposed by Stolar and Gick (2013), with a similar aim; Modified Curvature Index, based on Stolar and Gick's measure, was described in Dawson et al. (2015). The latter study also described two other procedures, Procrustes analysis and Discrete Fourier transform, that could potentially be applied to ultrasound data collected without head stabilisation. All these indices could be useful in a study of covert contrasts in rhotic consonant production. Also, such metrics would be suitable to analyse coronal tongue shapes. In the present paper, illustrations of highly differentiated tongue shapes can be found in figure 3. For example, the bottom right panel has two excursions in the tongue, while the other two curves in the bottom row have only one excursion each. Modified Curvature Index was calculated for these three tongue curves using a Python (Python Language, 2013) script provided in Dawson et al. (2015). The index for a given tongue curve is obtained by taking the integral of the absolute curvature with respect to the arc length of the curve; higher index

values characterise more differentiated tongue shapes. The tongue curve on the right, i.e., where the tongue front and back are excursed more than the tongue middle, has the highest value (2.67), followed by the middle curve (2.18), and the curve on the left has the smallest value (1.87).

Ultrasound examples of convert contrasts

Figure 3 shows example tongue curves from a girl aged 5;10, with typically developing speech (top row), and from a girl aged 5;5, who consistently produces sibilant fricatives as dental stops (bottom row). Both children speak Scottish Standard English. Ultrasound images for a token of /s/ from /sa/ (left) and two tokens of /ʃ/ from /ʃa/ (middle and right), taken at mid-consonant (mid-closure for the consonants produced as dental stops), are displayed in the figure. The consonant-vowel target syllables were produced in a short carrier sentence, "It's a …, Pam". In the productions by the typically developing child, there is a noticeable bunching towards the front of the tongue for /ʃ/, but not for /s/. For the other child, who produced all three tokens as unaspirated dental stops auditorily indistinguishable from each other by the three authors[2], tongue contours for /ʃ/ realisations also appear to be different from the contour for /s/, in that the tongue looks somewhat more bunched towards the front. In one realisation of /ʃ/ (right panels in figure 3), both children seem to have another constriction towards the back of the tongue, creating a "saddle" shape for the whole tongue contour. This then is an example of a covert contrast, where the child with the speech disorder seems to the listener to

[2] Perceptual similarity was assessed by each of the three authors individually by listening to audio recordings of the realisations of the carrier sentences containing the target consonant-vowel syllables. There was no fixed order of presentation of the three sentences, and no limitation on the number of times that the stimuli could be played back. The task was to determine whether the three target consonants sounded the same to each other, or whether there were any audible differences.

be neutralising target /s/ and /ʃ/ sounds, with both targets produced as /t/. However, on examination of the ultrasound images, the /t/ that is a substitution for /s/ is quite different from the /t/ that is a substitution for /ʃ/. Futhermore, one of this child's attempts to produce /ʃ/ is approaching an appropriate saddle shaped tongue contour similar to the typical child's contour. In summary, the ultrasound evidence suggests that the child with the disorder is making an articulatory distinction in the form of different tongue contours between target sounds /s/ and /ʃ/, both realised perceptually as /t/, although this distinction is not readily detectable to listeners.

Insert figure 3 about here

A measure of tongue shape described above, $LOC_{a\text{-}i}$, makes it possible to quantify the differences between alveolar and postalveolar fricative tongue shapes produced by the typically developing child. $LOC_{a\text{-}i}$ is an index showing to what extent the front part of the tongue is excursed in relation to the back part of the tongue. $LOC_{a\text{-}i}$ was calculated in R (R Development Core Team, 2013). The measure is illustrated in figure 3. The straight line between two ends of the curve is shown in each panel. Two perpendiculars from this line to the tongue curve are shown as dashed lines, plotted at one third and two thirds of the line that connects the curve ends. The lengths of the two perpendiculars represent the extent of excursion of the tongue front and the tongue back. From a visual inspection of the plots in the top row of figure 3 it is clear that in the typically developing child the "front" perpendicular is noticeably longer than the "back" one for both realisations of /ʃ/, but not for /s/. $LOC_{a\text{-}i}$, the ratio of the two perpendiculars, confirms this observation, returning higher $LOC_{a\text{-}i}$ values for both tokens of /ʃ/ than for the token of /s/ (the values are printed on each panel). This pattern,

similar to that reported for older children and adults (Zharkova, 2016), illustrates the fact that the child raises the tongue predorsum in order to produce the required constriction for /ʃ/, as opposed to the production of /s/, where the constriction is formed by the tongue tip at the lower teeth, with no need to raise the predorsum in the context of a following low vowel.

While the typically developing child differentiates between /s/ and both tokens of /ʃ/, the child who produces a dental stop for both fricatives has a somewhat higher value of LOC_{a-i} index for the token of /ʃ/ displayed in the right panel than for /s/, but not for the other token of /ʃ/ (middle panel), whose index value is nearly the same as for /s/ (see the bottom row of figure 3). On another index of tongue shape, however, namely Cuvature Position (Aubin and Ménard, 2006), this child clearly differentiates between /s/ and both tokens of /ʃ/. For the present paper, this index was also calculated in R. Cuvature Position quantifies the location of the most excursed part of the tongue in relation to the rest of the curve. This measure is also illustrated in figure 3. Curvature Position is a ratio of two straight lines which together form the long line that connects the two curve ends. The numerator is the length of the line between the back end of the tongue curve and the intersection with the perpendicular from the long line to the most excursed part of the tongue curve (this perpendicular is shown in the panels as a solid line). The denominator is the length of the line that constitutes the remainder of the long straight line. When the perpendicular is located closer to the tongue front, such as in the right panels of figure 3, Curvature Position value is higher; when it is located further back, such as in the left panels of figure 3, the index takes lower values. From examining the values printed in the panels, it is clear that both children have noticeably higher values for the tokens of /ʃ/ than for /s/.

Examining video recordings simultaneously collected during ultrasound data collection from the child who stopped the target fricatives suggests that the tongue tip tended to touch the lower teeth during the production of the whole sequence including the target consonant, the following vowel /a/ and also the preceding schwa. It appears that the tongue was not independent enough from the jaw to fully lift from the floor of the mouth and to create a groove for both consonants, as well as a sublingual space required for producing a perceptually appropriate /ʃ/. This difference in lingual articulation between the two children was not captured by the ultrasound measures, because of the lack of references external to the tongue in the ultrasound image. However, as demonstrated by the ultrasound data, the child was able to produce different tongue shapes for different target sounds during the stop closure of the substituted sound /t/. If this difference in an ultrasound measure was consistently observed in multiple repetitions, then it could be taken as instrumental evidence of a covert contrast. This paper aims to provide examples rather than to present results of comparisons of multiple repetitions using inferential statistical analyses. For more information on possible statistical techniques for comparing tongue shape indices across groups of speakers and across different phonemes, the reader is referred to research papers that have used such techniques, e.g., Ménard et al. (2012); Zharkova (2013b); Klein et al. (2013); Stolar and Gick (2013); Zharkova et al. (2015a, 2015b); Dawson et al. (2015); McAllister Byun et al. (2015).

Another example, illustrated in figure 4, is from a typically developing Scottish English speaking girl aged 3;4, who produces /ʃ/ in the context of /i/ very similar to /s/ in the same vowel context. The figure shows tongue curves for a target /s/ and a target /ʃ/. The tongue curves are taken from mid-consonant in the context of the following vowel /i/. Each of the two consonants was perceived as /s/ by the three authors. The two curves look quite

similar, in that the most bunched part of the tongue is located towards the front of the tongue, although it appears that the tongue root for /ʃ/ is somewhat flatter than for /s/, thus changing the overall tongue shape somewhat in relation to that for the target alveolar consonant. This visual impression is confirmed by calculating $LOC_{a\text{-}i}$ index. The index value is 1.23 for /s/ and 1.58 for /ʃ/, with the larger value for the postalveolar target consonant reflecting the difference between the two consonants in the posterior tongue shape.

Insert figure 4 about here

Discussion and conclusions

This paper demonstrates how ultrasound tongue imaging could be used for identifying covert contrasts. Because covert contrasts tend to involve small changes in articulator position, measurements need to be sensitive to those small changes. Measures based on ratios of straight lines are not as sensitive as measures taking into account all data points along the tongue curve (see Zharkova et al., 2015a, for a more detailed discussion of this issue). This may have been the reason why McAllister Byun et al. (2015), in an ultrasound study of covert contrasts, did not find an ultrasound-based index to be more sensitive to a covert contrast between /t/ and /k/ than the fourth spectral moment. In developing new measurements, specifically those that could successfully demonstrate covert contrasts, we can expect that the more of the tongue curve is captured by the measure, like with quantifications based on a series of data points along the curve (e.g., Stolar and Gick, 2013; Dawson et al., 2015), the more sensitive the measure will be to small changes, and hence the more able it will be to detect covert contrasts.

Similarly to other articulatory indices, the measures described above can be expected to be highly variable across speakers (cf. Gibbon et al., 1993; Hardcastle and Gibbon, 1997;

Gibbon et al., 2003; Gibbon et al., 2007). Such variability has indeed been reported for some of the measures (Zharkova et al., 2015a). Some of the variability may be brought about by using hand-held ultrasound data. If such measures are applied to tongue images from non-stabilised recordings, particular care needs to be taken when interpreting the measurements, due to potential effects from the transducer rotation and translation, and changes in the pressure of the transducer. Given that not all of the indices described here have so far been compared across stabilised versus non-stabilised conditions, and those that have been do not always provide robust results across all speech sounds, it would be useful to carry out such reliability testing on other measures and more speech sounds.

Detecting covert contrasts is important in disordered speech, as the approach to therapy is affected by whether a given error is phonological or phonetic in nature (Gibbon, 2003). Most clinical studies to date, particularly with children as participants, have used hand-held ultrasound without recording external reference points (e.g., Bacsfalvi and Bernhardt, 2011; Bressmann et al., 2011; Klein et al., 2013; Preston et al., 2015). While results of clinical studies with head-to-transducer stabilisation have been reported (e.g., Cleland et al., 2015), using head stabilisation limits the children's age, as with children under six years old it is generally not possible to use reliable stabilising devices. Thus, the indices described above and similar measures can make it easier to use ultrasound as a diagnostic tool in speech therapy.

Declaration of Interest

This work was supported by a grant from the Economic and Social Research Council (ES/K002597/1). The authors report no declarations of interest.

References

Abakarova, D., Iskarous, K., Robertus, E., Ries, J., Tiede, M., & Noiray, A. (2015). Development of coarticulation in German children: Mutual information as a measure of coarticulation and invariance. Poster at Ultrafest VII, 8-10 December 2015, University of Hong Kong, Hong Kong.

Aubin, J., & Ménard, L. (2006). Compensation for a labial perturbation: An acoustic and articulatory study of child and adult French speakers. In H. C. Yehia, D. Demolin, & R. Laboissière (Eds.), *Proceedings of the 7th international seminar on speech production* (pp. 209–216). Ubatuba, Brazil.

Bacsfalvi, P., & Bernhardt, B. M. (2011). Long-term outcomes of speech therapy for seven adolescents with visual feedback technologies: ultrasound and electropalatography. *Clinical Linguistics & Phonetics*, *25*, 1034–1043.

Bressmann, T., Ackloo, E., Heng, C.-L., & Irish, J. C. (2007). Quantitative three-dimensional ultrasound imaging of partially resected tongues. *Otolaryngology – Head and Neck Surgery*, *136*, 799–805.

Bressmann, T., Radovanovic, B., Kulkarni, G.V., Klaiman, P. & Fisher, D. (2011). An ultrasonographic investigation of cleft-type compensatory articulations of voiceless velar stops. *Clinical Linguistics & Phonetics*, *25*, 1028–1033.

Bressmann, T., Thind, P., Bollig, C. M., Uy, C., Gilbert, R. W., & Irish, J. C. (2005). Quantitative three-dimensional ultrasound analysis of tongue protrusion, grooving and symmetry: Data from twelve normal speakers and a partial glossectomee. *Clinical Linguistics & Phonetics, 19*, 573–588.

Carignan, C. (2014). TRACTUS (Temporally Resolved Articulatory Configuration Tracking of Ultrasound) software suite. Retrieved from http://phon.chass.ncsu.edu/tractus/.

Cleland, J., Scobbie, J. M., & Wrench, A. A. (2015). Using ultrasound visual biofeedback to treat persistent primary speech sound disorders. *Clinical Linguistics & Phonetics, 29,* 575–597.

Cleland, J., Scobbie, J., & Zharkova, N. (2016). Editorial: Insights from ultrasound: enhancing our understanding of clinical phonetics. *Clinical Linguistics & Phonetics, 30,* 171–173.

Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of the Acoustical Society of America*, *120,* 407–415.

Dawson, K. M., Tiede, M. K., & Whalen, D. H. (2015). Methods for quantifying tongue shape and complexity using ultrasound imaging. *Clinical Linguistics & Phonetics, Early Online*. DOI: 10.3109/02699206.2015.1099164.

Gibbon, F. E. (2002). Features of impaired tongue control in children with phonological disorder. In F. Windsor, M. L. Kelly & N. Hewlett (Eds.), *Investigations in Clinical Phonetics and Linguistics* (pp. 299–309). Lawrence Erlbaum: New Jersey.

Gibbon, F. E. (2003). Using articulatory data to inform speech pathology theory and clinical practice. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, Spain*. Pp. 261–264.

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

Gibbon, F., Hardcastle, W., & Nicolaidis, K. (1993). Temporal and spatial aspects of lingual coarticulation in /kl/ sequences: a cross-linguistic investigation. *Language and Speech, 36,* 261–277.

Gibbon, F., Lee, A., & Yuen, I. (2007). Tongue-palate contact during bilabials in normal speech. *Cleft Palate-Craniofacial Journal, 44*, 87–91.

Gibbon, F. E., McNeill, A. M., Wood, S. E., & Watson, J. M. (2003). Changes in linguapalatal contact patterns during therapy for velar fronting in a 10-year-old with Down's syndrome. *International Journal of Language and Communication Disorders, 38,* 47–64.

Gibbon, F. E., & Wood, S. E. (2003). Using electropalatography (EPG) to diagnose and treat articulation disorders associated with mild cerebral palsy: a case study. *Clinical Linguistics & Phonetics, 17*, 365–374.

Gick, B., Bacsfalvi, P., Bernhardt, B. M., Oh, S., Stolar, S., & Wilson, I. (2008). A motor differentiation model for liquid substitutions: English /r/ variants in normal and disordered acquisition. *Proceedings of Meetings on Acoustics, 1(060003),* 1–9.

Gick, B., Campbell, F., Oh, S., & Tamburri-Watt, L. (2006). Toward universals in the gestural organization of syllables: A cross-linguistic study of liquids. *Journal of Phoneticis, 34*, 49–72.

Hardcastle, W. J., & Gibbon, F. (1997). Electropalatography and its clinical applications. In M. J. Ball, & C. Code (Eds.), *Instrumental clinical phonetics* (pp. 149–193). London: Whurr.

Hewlett, N. (1988). Acoustic properties of /k/ and /t/ in normal and phonologically disordered speech. *Clinical Linguistics & Phonetics*, *2*, 29–45.

Klein, H. B., Byun, T. M., Davidson, L., & 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*, 540–553.

Lee, A., Zharkova, N., & Gibbon, F. (2013). Vowel Imaging. In M. Ball & F. Gibbon (Eds.), *Handbook of vowels and vowel disorders* (2nd ed., pp. 138–159). Hove: Psychology Press.

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

Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in English: a study of voice onset time in word-initial stop consonants. *Journal of Child Language, 7,* 41–74.

Magloughlin, L. (2016). Accounting for variability in North American English /ɹ/: Evidence from children's articulation. *Journal of Phonetics, 54,* 51–67.

McAllister Byun, T., Buchwald, A., & Mizoguchi, A. (2015). Covert contrast in velar fronting: An acoustic and ultrasound study. *Clinical Linguistics & Phonetics*, Early Online.

Ménard, L., Aubin, J., Thibeault, M., & Richard, G. (2012). Comparing tongue shapes and positions with ultrasound imaging: a validation experiment using an articulatory model. *Folia Phoniatrica et Logopaedica, 64,* 64–72.

Ménard, L., & Noiray, A. (2011). The development of lingual gestures in speech: Comparing synthesized vocal tracts with natural vowels. *Faits de langue, 37,* 189–202.

Ménard, L., Turgeon, C., Trudeau-Fisette, P., & Bellavance-Courtemanche (2015). Effects of blindness on production–perception relationships: Compensation strategies for a lip-tube perturbation of the French [u]. *Clinical Linguistics & Phonetics*, Early Online.

Mielke, J. (2015). An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons. *Journal of the Acoustical Society of America, 137,* 2858–2869.

Mielke, J., Baker, A., Archangeli, D., & Racy, S. (2005). Palatron: a technique for aligning ultrasound images of the tongue and palate. In D. Siddiqi & B. V. Tucker (Eds.), *Coyote Papers*, *14*, 97–108.

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

Noiray, A., Ries, J., & Tiede, M. (2015). Sonographic & Optical Linguo-Labial Articulation Recording system (SOLLAR). Oral presentation at Ultrafest VII, 8-10 December 2015, University of Hong Kong, Hong Kong.

Preston, J., Maas, E., Whittle, J., Leece, M., & Mccabe, P. (2015). Limited acquisition and generalisation of rhotics with ultrasound visual feedback in childhood apraxia. *Clinical Linguistics & Phonetics*, Early Online.

Python Language. (2013). Python Software Foundation. Retrieved from www.python.org.

R Development Core Team. (2013). R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org.

Recasens, D., & Rodríguez, C. (2015). An investigation of lingual coarticulaiton resistance using ultrasound. Oral presentation at Ultrafest VII, 8-10 December 2015, University of Hong Kong, Hong Kong.

Rubertus, E., Abakarova, D., Tiede, M., Ries, J., & Noiray, A. (2015). Development of coarticulation in German children: acoustic and articulatory locus equations. Poster at Ultrafest VII, 8-10 December 2015, University of Hong Kong, Hong Kong.

Scobbie, J. M., Gibbon, F., Hardcastle, W. J., & Fletcher, P. (2000). Covert contrast as a stage in the acquisition of phonetics and phonology. M. Broe & J. Pierrehumbert

(Eds), *Papers in Laboratory Phonology V: Language Acquisition and the Lexicon*. Cambridge: Cambridge University Press. Pp. 194-203.

Song, J. Y., Demuth, K., Shattuck-Hufnagel, S., & Ménard, L. (2013). The effects of coarticulation and morphological complexity on the production of English coda clusters: Acoustic and articulatory evidence from 2-year-olds and adults using ultrasound. *Journal of Phonetics, 41*, 281–295.

Stolar, S., & Gick, B. (2013). An index for quantifying tongue curvature. *Canadian Acoustics*, *41*, 11–15.

Stone, M. (1990). A three-dimensional model of tongue movement based on ultrasound and x-ray microbeam data. *Journal of Acoustical Society of America, 87*, 2207–2217.

Stone, M. (2005). A guide to analyzing tongue motion from ultrasound images. *Clinical Linguistics & Phonetics, 19*, 455–502.

Stone, M. (2010). Laboratory techniques for investigating speech articulation. In Hardcastle, W. J., Laver, J. & Gibbon, F. E. (Eds.), *The Handbook of Phonetic Sciences* (2nd ed.) (pp. 9-38). Chichester, West Sussex, U.K.; Malden, MA: Wiley-Blackwell.

Stone, M., Faber, A., Raphael, L. J., & Shawker, T. H. (1992). Cross-sectional tongue shape and linguopalatal contact patterns in [s], [ʃ], and [l]. *Journal of Phonetics, 20*, 253-270.

Stone, M., Morrish, K. A., Sonies, B. C., & Shawker, T. H. (1987). Tongue curvature: A model of shape during vowel production. *Folia Phoniatrica, 39*, 302–315.

Stone, M., Shawker, T. H., Talbot, T., & Rich, A. (1988). Cross-sectional tongue shape during the production of vowels. *Journal of the Acoustical Society of America, 83*, 1586–1596.

Whalen, D. H., Iskarous, K., Tiede, M. K., Ostry, D. J., Lehnert-LeHouillier, H., Vatikiotis-Bateson, E., & Hailey, D. S. (2005). HOCUS, the Haskins Optically-Corrected Ultrasound System. *Journal of Speech, Language, and Hearing Research, 48*, 543–553.

Yip, J., Archangeli, D., & To, C. K. S. (2015). Development of lingual articulations among Cantonese-speaking children. Oral presentation at Ultrafest VII, 8-10 December 2015, University of Hong Kong, Hong Kong.

Zharkova, N. (2013a). Using ultrasound to quantify tongue shape and movement characteristics. *The Cleft Palate-Craniofacial Journal, 50*, 76–81.

Zharkova, N. (2013b). A normative-speaker validation study of two indices developed to quantify tongue dorsum activity from midsagittal tongue shapes. *Clinical Linguistics & Phonetics, 27,* 484–496.

Zharkova, N. (2016, to appear). Ultrasound and acoustic analysis of sibilant fricatives in preadolescents and adults. *Journal of the Acoustical Society of America*.

Zharkova, N., & Hewlett, N. (2009). Measuring lingual coarticulation from midsagittal tongue contours: Description and example calculations using English /t/ and /ɑ/. *Journal of Phonetics, 37*, 248–256.

Zharkova, N., Gibbon, F. E. & Hardcastle, W. J. (2015a). Quantifying lingual coarticulation using ultrasound imaging data collected with and without head stabilisation. *Clinical Linguistics & Phonetics, 29*, 249-265.

Zharkova, N., Hardcastle, W. J., Gibbon, F. E. & Lickley, R. J. (2015b). Development of lingual motor control in children and adolescents. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18 th International Congress of Phonetic Sciences, 10-14 August 2015, Glasgow, UK*. Online ISBN 978-0-85261-941-4.

Figure captions

Figure 1. Recording of a 3-year-old child with ultrasound without head-to-transducer stabilisation.

Figure 2. Midsagittal ultrasound images of the tongue at mid-/s/ from /sa/ (left) and mid-/ʃ/ from /ʃa/ (right), taken from productions by a typically developing three-year-old child. The front of the tongue is on the right in this figure (same for figures 3 and 4). In both panels, the dark area to the left of the visible tongue root is the hyoid bone shadow, and the dark area to the right of the frontmost part of the tongue is the mandible shadow.

Figure 3. Tongue curves for two 5-year-old children: a typically developing (TD) child (top row) and a child who consistently stops her sibilant fricatives (bottom). The left panel, for both children, shows a tongue shape for /s/, and the middle and right panels show tongue shapes for two different realisations of /ʃ/ for the two children. Values of two tongue shape indices, $LOC_{a\text{-}i}$ and Curvature Position, are included in each panel – "CP" stands for "Curvature Position". The straight lines included in each plot illustrate the calculations (see text for details). Note that "tongue position" in the abscissa legend in figures 3 and 4 refers to the position relative to the ultrasound transducer within each token.

Figure 4. Tongue curves for target /s/ and /ʃ/ produced by a typically developing 3-year-old child. Values of an index of tongue shape are printed in each panel, and the straight lines included in each plot illustrate the index calculations.

Figure 1



Figure 2

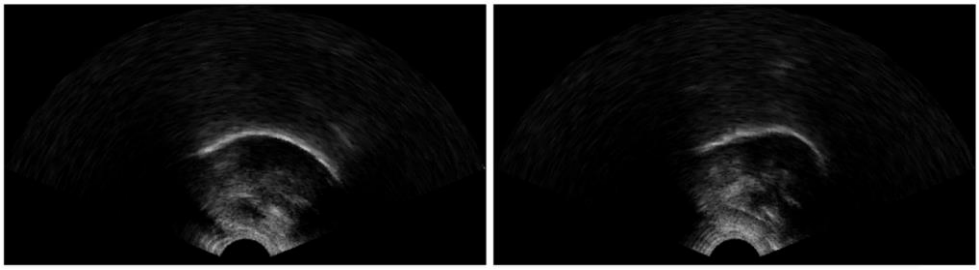


Figure 3

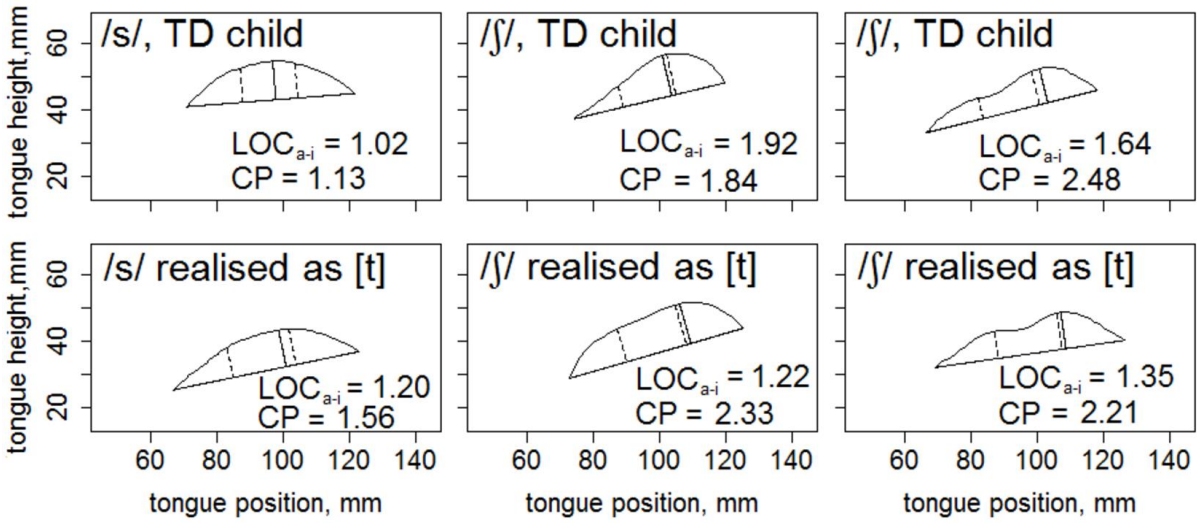
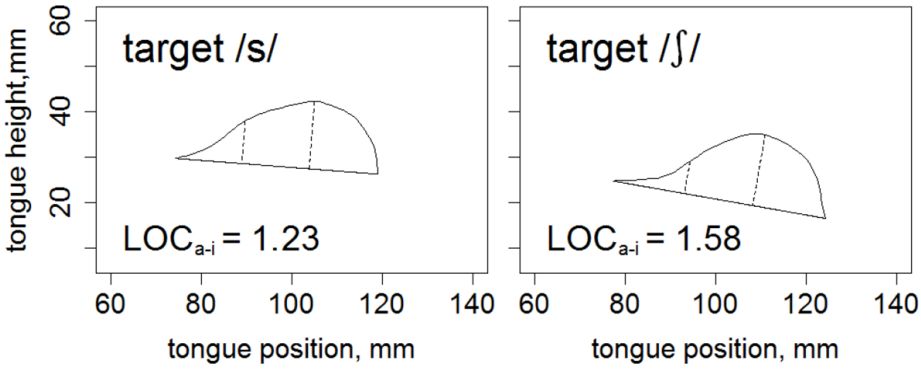

/s/, TD child
LOC$_{a-i}$ = 1.02
CP = 1.13
/ʃ/, TD child
LOC$_{a-i}$ = 1.92
CP = 1.84
/ʃ/, TD child
LOC$_{a-i}$ = 1.64
CP = 2.48
/s/ realised as [t]
LOC$_{a-i}$ = 1.20
CP = 1.56
/ʃ/ realised as [t]
LOC$_{a-i}$ = 1.22
CP = 2.33
/ʃ/ realised as [t]
LOC$_{a-i}$ = 1.35
CP = 2.21
tongue height,mm
20
40
60
60
80
100
120
140
tongue position, mm

Figure 4


target /s/
LOC$_{a-i}$ = 1.23
target /ʃ/
LOC$_{a-i}$ = 1.58
tongue height,mm
20
40
60
60
80
100
120
140
tongue position, mm
tongue position, mm