

# **Characteristics of articulatory gestures in stuttered speech: a case study using real-time magnetic resonance imaging**

Yijing Lu <sup>a,\*</sup>, Charlotte E. E. Wiltshire <sup>b,1</sup>, Kate E. Watkins <sup>b</sup>, Mark Chiew <sup>c</sup>, Louis Goldstein <sup>a</sup>

<sup>a</sup> Department of Linguistics, University of Southern California, United States

<sup>b</sup> Wellcome Centre for Integrative Neuroimaging, Department of Experimental Psychology, University of Oxford, United Kingdom

<sup>c</sup> Wellcome Centre for Integrative Neuroimaging, Nuffield Department of Clinical Neurosciences, University of Oxford, United Kingdom

yijinglu@usc.edu (corresponding author)

c.wiltshire@phonetik.uni-muenchen.de

kate.watkins@psy.ox.ac.uk

mark.chiew@ndcn.ox.ac.uk

louisgol@usc.edu

---

\* Corresponding author at: 3601 Watt Way, Grace Ford Salvatori 301, Los Angeles, CA 90089, United States.

<sup>1</sup> Charlotte E. E. Wiltshire is now at the Institute for Phonetics and Speech Processing, Ludwig-Maximilians-University (LMU), Munich.

# Characteristics of articulatory gestures in stuttered speech: a case study using real-time magnetic resonance imaging

## Abstract

*Introduction:* Most of the previous articulatory studies of stuttering have focussed on the fluent speech of people who stutter. However, to better understand what causes the actual moments of stuttering, it is necessary to probe articulatory behaviors during stuttered speech. We examined the supralaryngeal articulatory characteristics of stuttered speech using real-time structural magnetic resonance imaging (RT-MRI). We investigated how articulatory gestures differ across stuttered and fluent speech of the same speaker.

*Methods:* Vocal tract movements of an adult man who stutters during a pseudoword reading task were recorded using RT-MRI. Four regions of interest (ROIs) were defined on RT-MRI image sequences around the lips, tongue tip, tongue body, and velum. The variation of pixel intensity in each ROI over time provided an estimate of the movement of these four articulators.

*Results:* All disfluencies occurred on syllable-initial consonants. Three articulatory patterns were identified. Pattern 1 showed smooth gestural formation and release like fluent speech. Patterns 2 and 3 showed delayed release of gestures due to articulator fixation or oscillation respectively. Block and prolongation corresponded to either pattern 1 or 2. Repetition corresponded to pattern 3 or a mix of patterns. Gestures for disfluent consonants typically exhibited a greater constriction than fluent gestures, which was rarely corrected during disfluencies. Gestures for the upcoming vowel were initiated and executed during these consonant disfluencies, achieving a tongue body position similar to the fluent counterpart.

*Conclusion:* Different perceptual types of disfluencies did not necessarily result from distinct articulatory patterns, highlighting the importance of collecting articulatory data of stuttering. Disfluencies on syllable-initial consonants were related to the delayed release and the overshoot of consonant gestures, rather than the delayed initiation of vowel gestures. This suggests that stuttering does not arise from

problems with planning the vowel gestures, but rather with releasing the overly constricted consonant gestures.

**Keywords:** adults who stutter, Articulatory Phonology, articulatory gestures, real-time magnetic resonance imaging, consonant-vowel coarticulation

## 1. Introduction

Developmental stuttering is a neurodevelopmental variation characterized by interruptions to the flow of speech, in the context of language and social interactions. Three types of speech disfluencies are widely accepted as the hallmark characteristics of developmental stuttering: (1) **repetitions** of speech sounds and syllables (e.g., “m-m-m-mom”, “be-be-be-because”); (2) **prolongations** of speech sounds (e.g., “wwwhere”, “llllook”); (3) silent pauses prior to producing a speech sound, known as **blocks** (“–dad”, “ea – ten”). These characteristics are defined from an auditory-perceptual / acoustic point of view, but a growing consideration of developmental stuttering from the perspective of speech motor control (e.g., Büchel & Sommer, 2004; Kent, 2000; Max et al., 2004; Peters et al., 2000) highlights the importance of characterizing stuttering behaviors at the articulatory level. Atypicalities in the neural mechanisms of speech production are expected to be manifested as atypicalities in articulatory behaviors, which in turn give rise to the acoustic signals perceived as disfluencies.

It is worth pointing out that articulatory behaviors cannot be sufficiently inferred from auditory/acoustic output due to the nonlinear mapping between articulation and acoustics. In particular, since phonation is often disrupted or even absent in stuttering, supralaryngeal articulatory activities are not necessarily reflected in acoustics. For example, Harrington (1987) reported that although the vowel target was not acoustically realized in stuttered consonant-vowel (CV) syllables, anticipatory lingual coarticulation did exist, as evidenced by articulatory data. Furthermore, an articulatory study by Didirková et al. (2019) revealed that stuttering disfluencies perceived as the same type of disfluency

(block/prolongation/repetition) are the result of different movement patterns of supralaryngeal articulators. Therefore, a direct examination of articulatory behaviors of stuttering is needed.

Such an undertaking has been difficult because of the limited technology to visualize and track the movements of speech articulators, especially during stuttered speech. Common articulatory kinematic data recording tools used in prior studies on stuttering include electromagnetic articulography (e.g., Didirkova et al., 2019, 2020a, 2020b), Optotrak (e.g., Smith et al., 2010; Kleinow & Smith, 2000; Smith & Kleinow, 2000), X-ray (e.g., Zimmermann, 1980a, 1980b), and ultrasound (e.g., Heyde et al., 2016). Optotrak and electromagnetic articulography (EMA) are point-tracking systems that track the motion of sensors attached to speech articulators. Optotrak uses digital cameras to track the movements of infrared-emitting diodes. Since Optotrak relies on a direct line-of-sight, the infrared-emitting diodes can only be attached to the skin surfaces, and therefore its use is limited to tracking the movements of external articulators like the jaw and lips. EMA, on the other hand, uses magnetic fields to track the positions of transducers, hence allowing the sensors (transducers) to be placed inside the mouth. The sensors can be attached to the jaw, the lips, the anterior part of the tongue, but rarely to the rear of tongue dorsum, the tongue root, and the velum, due to the discomfort that could be caused to participants. Although Optotrak and EMA offer high temporal resolution and high spatial resolution for the positions of sensors, they cannot track the motion of all articulators, especially the posterior ones. The attached sensors might also disturb the speech movements. X-ray and ultrasound are imaging devices that take images of the moving articulators. Ultrasound is typically used to image the surface shape of a portion of the tongue but has a very limited use for imaging other articulators. X-ray can image the entire vocal tract but has been abandoned in research due to safety concerns. Electromyography (EMG) is another commonly used tool although it does not directly measure the kinematics of articulators. Instead, it uses electrodes to measure muscle activation of articulators. Like Optotrak and EMA, EMG can only output information localized to where electrodes are placed, which are typically on the lips, jaw, and neck in studies using surface EMG. Intramuscular EMG, which can be used to probe tongue and larynx muscle activity, is highly invasive.

As indicated by the advantages and disadvantages of the tools discussed above, an ideal vocal tract imaging/tracking tool should have the following qualities: good spatial and temporal resolution (e.g., 3.5 mm and 70 ms, recommended by Lingala et al., 2016); able to probe the movements of multiple articulators, including inner speech organs (e.g., tongue root, velum); non-invasive and safe to use. The recently developed real-time magnetic resonance imaging (RT-MRI) can satisfy all these needs. RT-MRI is a form of structural magnetic resonance imaging that gives a safe, non-invasive solution to visualizing the movements of the entire vocal tract from lips to larynx with good spatial and temporal resolution (e.g., 1.5-2 mm and 20-30 ms; see Niebergall et al., 2013). Videos of vocal tract movements can be recorded in real time using RT-MRI, capturing the fast movements of all articulators simultaneously. The acquired images have good signal-to-noise ratio, are amenable to computerized modeling, and provide excellent structural differentiation. In the current study, we leveraged this state-of-the-art tool to investigate the dynamic vocal tract actions during both perceptually fluent and disfluent speech production of an adult who stutters.

Most of the prior articulatory studies of stuttering have focussed on comparing the articulation of perceptually fluent speech by people who stutter (PWS) with the fluent articulation of people who do not stutter (PWNS). Such comparisons can shed light on the general differences between PWS and PWNS in speech motor control. Yet, little consensus has been reached with regard to whether fluent speech of PWS and PWNS truly differs in various aspects of articulation. For example, some studies found that the articulatory movements of PWS had longer durations relative to those of PWNS (e.g., Caruso et al., 1988; Max et al., 2003; Zimmermann, 1980a), but some other studies did not find the same duration effect (McClellan et al., 2004; McClellan & Tasko, 2004). It is also not clear whether PWS have reduced articulator displacements and velocities (Zimmermann, 1980a) or not (Zimmermann & Hanley, 1983). Although higher articulator movement variability in PWS is the most consistent finding among prior research (e.g., Anonymous, 2020; Kleinow & Smith, 2000; Smith et al., 2010), it was still not replicated in some studies (Leha et al., 2020; Smith & Kleinow, 2000). There is also a lack of agreement regarding whether PWS

and PWNS differ in the sequencing of the movement onset and peak velocity of different articulators in the production of a speech unit, e.g., upper lip, lower lip, and jaw movements in the production of a bilabial stop. Caruso et al. (1988) found the group difference, but some later studies did not (e.g., De Nil, 1995; McClean et al., 1990).

On the other hand, it is also important to know whether the general differences between PWS and PWNS really contribute to the occurrence of dysfluency in the speech of PWS. For this purpose, it is necessary to examine the articulatory manifestation of actual moments of stuttering. Due to technical difficulties, only a limited number of articulatory studies have analyzed disfluent speech. Most of them examined the muscle activities of articulators (Conture et al., 1977; Conture et al., 1985; Denny & Smith, 1992; Fibiger, 1977; McClean et al., 1984; Smith, 1989; Smith et al., 1993; Smith et al., 1996), instead of the kinematics. While muscle activities measured by EMG reflect the amount of tension involved in speech production, they do not directly reveal spatial information about the articulators or their movements. Moreover, since EMG typically applies to lip and larynx muscles, how the tongue—the most important articulator—is activated during stuttered speech was not investigated in these studies.

Zimmermann (1980b), Didirková et al. (2019), Didirková et al. (2020a), and Didirková et al. (2020b) are the few studies that directly investigated the movements of multiple supralaryngeal articulators during stuttering disfluencies. Zimmermann (1980b) is a cinefluorographic study in which the movements of the tongue tip, tongue dorsum, lower lip, and jaw of PWS during perceptually identified repetitions and prolongations were examined and compared to the articulatory movements in the fluent speech of a PWNS. Oscillations and static positioning of the primary articulator (the articulator making the major consonantal constriction), along with the repositioning of the secondary articulator, were observed in repetitions and prolongations. For example, during the repetition of the initial /p/ in /pap/, the lower lip, as the primary articulator, was doing repetitive vertical movements. Meanwhile, the secondary articulator—the jaw—was gradually lowered throughout the repetition. Repositioning of the tongue tip

and tongue dorsum (as secondary articulators) was also found in the production of other disfluent /p/. Zimmermann interpreted it as the reshaping of the tongue towards a “resting” posture.

Didírková and colleagues used EMA to track the movements of the tongue body, tongue tip, lower and upper lips, and mandible in speech. Didírková et al. (2019) summarized four articulatory patterns based on the movements of all these articulators during disfluencies: reiterations of movements, global maintain of postures, anarchical movements, and a combination of the above. The first articulatory pattern was always related to repetition, whereas the other three were involved in all three types of stuttering disfluencies. Didírková et al. (2020a) and Didírková et al. (2020b) focussed on the contextual effect on articulation. Didírková et al. (2020a) examined whether articulatory activities before the stuttering disfluency are in anticipation of the stuttered speech sound. They found that most of the disfluent sounds were correctly anticipated, suggesting that stuttering disfluencies are not necessarily caused by a lack of anticipatory coarticulation. Didírková et al. (2020b) further included “typical” disfluencies produced by PWNS (e.g., word repetitions, filled pauses) into the analysis. Compared to PWNS, PWS tended to maintain the articulatory postures for the preceding speech sound during disfluencies. Consequently, anticipatory movements for the subsequent sound started later.

Likewise, the current study aims to uncover the articulatory characteristics of stuttered speech, but it extends the previous studies in two ways. First, the choice of the baseline. In Zimmermann (1980b) and Didírková et al. (2020b), the speech of PWNS was used as the baseline. No baseline was used in Didírková et al. (2019) and Didírková et al. (2020a). The current study uses the fluent speech of PWS as the baseline. Articulatory characteristics of disfluent speech are summarized by comparing articulatory events in fluent and disfluent speech of the same PWS. This kind of within-subject comparison has already been applied to studying the neural characteristics of stuttering states (see Connally et al., 2018; see also Belyk et al., 2014 for a meta-analysis).

The current study also differs in its choice of kinematic measures. While prior research examined only the displacement of articulators during stuttering, the current study examines how well the

articulatory goals are achieved through such displacement. The theoretical framework underlying the current research is the gestural framework of Articulatory Phonology (Browman & Goldstein, 1989, 1992) and Task Dynamics (Saltzman & Munhall, 1989). It understands that the basic tasks in speech production (termed articulatory gestures) are for speech articulators to control the aperture between the upper and lower surfaces of the vocal tract in various regions. The aperture goal of a gesture can be a constriction in the vocal tract, e.g., full closure or narrowing in the oral cavity (for plosives and fricatives respectively). It can also be an opening, e.g., different degrees of opening between the tongue body and palate (for vowels), or velic opening (for nasals). The magnitude of an implemented gesture is the degree of constriction or opening achieved, depending on the what the aperture goal is. For example, the articulatory gesture for /p/ has a goal of forming a lip closure. In addition to observing the repetitive vertical movement of the lower lip during the repetitions of /p/, as in Zimmermann (1980b), we quantify how closely the lips are approximated to each other (the magnitude of lip constriction) throughout the repetition. The examination of coarticulation during stuttering disfluencies is also expanded from only examining the presence/absence of coarticulation, as Didirková et al. (2020a) did, to quantifying the degree of coarticulation by measuring the magnitude of coproduced consonant and vowel gestures.

Various analysis techniques have been developed to extract information about vocal tract apertures from RT-MRI data (see Ramanarayanan et al., 2018 for a review). The technique used in the current study, region-of-interest (ROI) based analysis, is a method that is relatively simple to implement, compared to other common MRI analysis techniques (e.g., grid-based methods and contour-based methods). It analyzes pixel intensity in specific regions of the vocal tract in the MRI image. The variation of pixel intensity across time reflects the presence/absence of soft tissues in the regions, thereby allowing the change of the vocal tract aperture in those regions to be estimated. Aperture kinematics obtained using this method have informed important questions in both linguistic studies and clinical research. For example, it provides insights into the articulation of speech sounds in various languages, including geminate consonants in Italian (Hagedorn et al., 2011), liquid consonants in Korean (Lee et al., 2015), and



nasal vowels in Portuguese (Teixeira et al., 2012). It also reveals important aspects of atypical speech production that may not be accessed by traditional methods, for example, non-audible articulatory gesture intrusions in apraxic speech (Hagedorn et al., 2017); compensatory articulatory behaviors in post-glossectomy speech (Hagedorn et al., 2014). Therefore, applying the ROI-based method to analyzing RT-MRI data of stuttered speech has great potential to yield meaningful insights.

To summarize, the current study had two goals: first, to show that RT-MRI is a suitable tool for collecting articulatory kinematic data for stuttered speech. Second, to demonstrate that vocal tract aperture information obtained from RT-MRI reveals potentially important articulatory kinematic differences between stuttered and fluent speech of PWS, which in turn have broader implications for the causes of stuttering moments.

## **2. Methods**

### **2.1. Data acquisition**

The data analyzed in the present study are a subset of the data collected in a RT-MRI study reported in Anonymous (2020). Data were collected on a 3T MRI scanner (Prisma, Siemens Healthineers, Erlangen, Germany) with a 64-channel head and neck receiver array. Mid-sagittal images of the head and upper airways were acquired with in-plane spatial resolution of 2 mm  $\times$  2 mm and 5-mm slice thickness over a 200-mm nominal field-of-view. A radial FLASH sequence (TE/TR = 1.4/2.5 ms), flip angle 5°, bandwidth 1929 Hz/px, with golden angle spoke ordering was used. Images were reconstructed at 33.3 frames per second (12 spokes per frame) using a second-order spatiotemporal total generalized variation constraint (Knoll et al., 2011).

### **2.2. Participant**

The data included here were from one participant in the original study, who frequently stuttered during RT-MRI scanning. The participant was a 23-year-old man with “severe” stuttering (SSI-4 = 35), as assessed by the second author (a postdoctoral researcher with a doctorate focussed on developmental

stuttering and trained by a qualified Speech Language Pathologist to do SSI scoring) using the Stuttering Severity Instrument-Fourth Edition (SSI-4; Riley, 2009). The participant reported normal vision and hearing, no audiological or neurological disorders, and no speech-language disorder other than persistent developmental stuttering. The participant reported the onset of stuttering around 4 years old and had received speech and language therapy around age 6 to 12 years old. The participant does not use any learned fluency-enhancing techniques in his day-to-day speech.

### **2.3. Experimental procedure and stimuli**

During scanning, the pseudowords were displayed on a screen one by one and the participant read them aloud at his natural speaking rate. The pseudoword stimuli, varying in length and phonological complexity, were taken from Smith et al. (2010): “mab” (/mæb/), “mabshibe” (/mæb.ʃaɪb/), “mabfieshabe” (/mæb.fai.ʃeɪb/), “mabshaytiedoib” (/mæb.ʃeɪ.taɪ.dɔɪb/), “mabteebeebee” (/mæb.ti.bi.bi/). Prior to scanning, the participant received instructions on how the pseudowords should be pronounced and practised until the pronunciation was accurate. Each pseudoword was displayed 10 times in a random order. Two trials of “mabfieshabe”, one trial of “mabshaytiedoib”, and one trial of “mab” were excluded from the analyses due to audio corruption or rushed speech. In total, 46 trials were analyzed.

### **2.4. Disfluency analysis**

The perceptual rating of stuttering disfluencies was conducted by two raters—the second author and a Speech Language Pathologist—both postdoctoral researchers with doctoral training focussed on developmental stuttering. The first rater performed the rating twice, six months apart. The raters examined the audio and RT-MRI video recordings of the participant’s production of the pseudowords to identify all instances of disfluencies. The identified instances of disfluencies were classified as repetition, prolongation, block, or the combination of any of these three types, based primarily on the auditory perception, with reference to the vocal tract videos. Repetitions and prolongations can be easily identified from the audio. The identification of blocks, on the other hand, involves using both the audio and the RT-

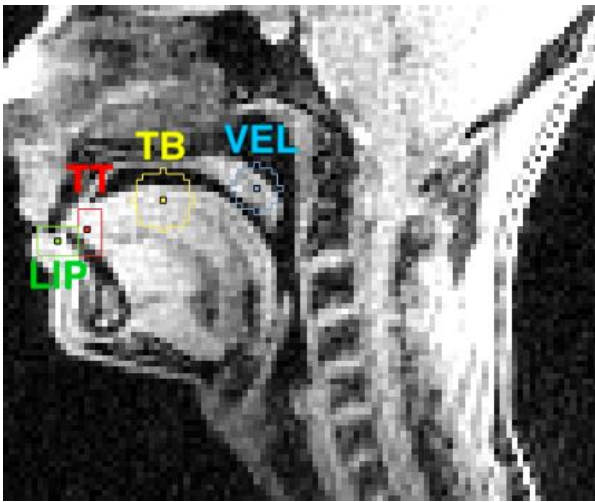
MRI video. When a silent interval was perceived as an interruption to speech, or when slightly audible laryngeal tension together with visible lip tension were perceived, the disfluency was rated as a block. If more than one type of disfluency was perceived in one instance, it was rated as the combination of corresponding types. The audio and video recordings were also manually segmented and annotated with the onset and offset of perceived disfluencies and the approximate temporal locations of phonetic units in fluent speech.

## **2.5. Articulatory analysis**

Vocal tract activity was analyzed into articulatory gestures of distinct proximal articulators, i.e., effectors. An analytical method of estimating vocal tract apertures (i.e., constriction or opening) based on average pixel intensity within regions of interest (ROIs) on RT-MRI frame sequences was used. As mentioned in the introduction section, this ROI-based method is one of the commonly used image analysis methods in RT-MRI studies and has been used to investigate both typical and atypical speech production. By averaging pixel intensities across all the pixels within an ROI, the results are robust to image noise (Lammert et al., 2010). Aperture information estimated from variations of mean pixel intensity within ROIs has been shown to be very similar to the direct measurements of aperture from traced vocal tract air-tissue boundaries in the same MRI image sequences (Proctor et al., 2011). When a complete closure is formed in an ROI, further changes in pixel intensity can indicate the degree of compression between the active and passive articulators, as such compression can increase the amount of tissue present in the ROI. Estimating the compression between articulators based on pixel intensity has been shown to generate results that are consistent with observations made using electropalatography (Hagedorn et al., 2011).

In the present study, four ROIs were manually defined by the second author according to anatomical landmarks and confirmed by the first author. The ROIs were placed at locations in the vocal tract where constrictions were proximally made by lips (LIP), tongue tip (TT), tongue body (TB), and where the velum (VEL) is lowered to when the nasopharynx is open (see Fig. 1). Specifically, the LIP

ROI was placed between the upper lip and lower lip to capture the labial constriction for /m/ and /f/. The TT ROI was placed immediately below the alveolar ridge to capture the alveolar constriction in the production of the alveolar stop /t/ and the palato-alveolar fricative /ʃ/. The TB ROI was placed at the end of the hard palate to detect the tongue body gestures involved in the production of palato-alveolar /ʃ/ and the production of vowels. The VEL ROI was placed at a location where the velum is lowered to detect the velic opening for the nasal consonant /m/ and the velic closing for the oral stops and fricatives. The ROIs were kept fixed across all frames.



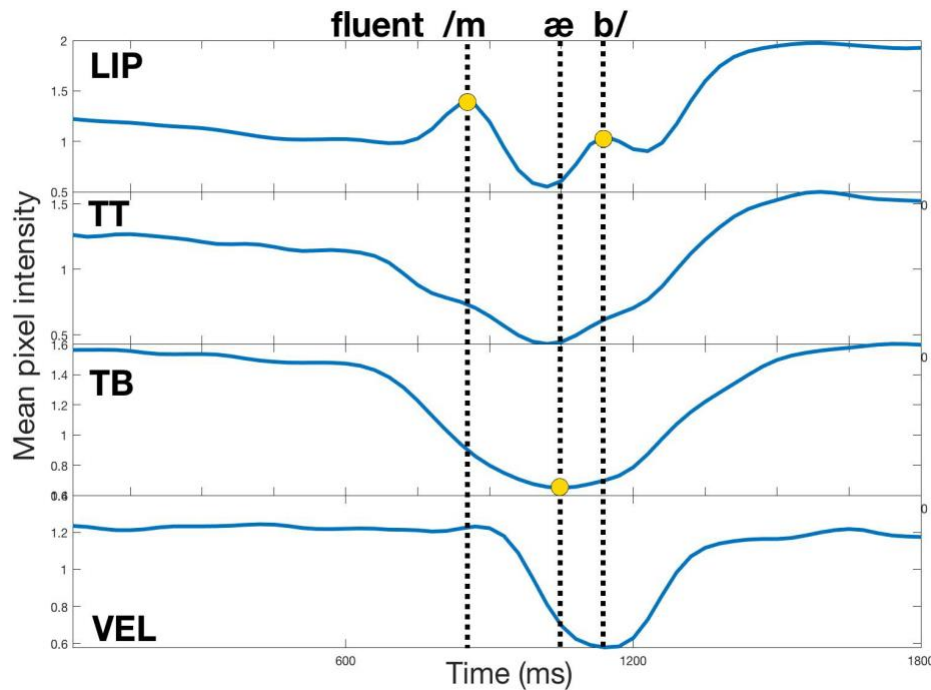
**Fig. 1.** Four regions of interest (ROIs) on the mid-sagittal slice of the speaker's vocal tract. LIP (green): lips; TT (red): tongue tip; TB (yellow): tongue body; VEL (blue): velum.

The mean intensity of pixels in an ROI reflects the amount of soft tissue in the region: a higher value indicates more soft tissue, and a lower value indicates less tissue. Based on the findings in Proctor et al. (2011) and Hagedorn et al. (2011) introduced above, the amount of tissue present in the ROI is thought to indicate the vocal tract aperture and the degree of compression between active and passive articulators, and therefore is used to estimate the magnitude of the articulatory gesture in terms of achieving its goal. For the lip gestures for /m/ and /f/, the tongue tip gesture for /t/, and the tongue tip and tongue body gestures for /ʃ/, their goal is to make a constriction at the locations where the ROIs are placed. Hence, a higher pixel intensity in the corresponding ROIs indicates that the lips/tongue tip/tongue body has entered the ROI, resulting in reduced aperture, more tissue contact (or compression), and a higher magnitude of constriction. The goal of a velum gesture depends on the nasality of the speech

sound. For producing nasals like /m/, the goal is to lower the velum. Since the VEL ROI is placed at where the velum is maximally lowered, a higher pixel intensity in the VEL ROI indicates a greater magnitude of velum lowering. The opposite cases are the oral stops and fricatives, which require the velum to be raised. Raising the velum results in a lower pixel intensity in the VEL ROI. The lower the pixel intensity is, the more raised the velum is. As for vowel gestures, high vowels require the tongue body to be raised, whereas the low vowels require the tongue body to be lowered. Since the TB ROI is placed right below the palate (for capturing tongue body raising), a higher pixel intensity indicates a higher tongue body position, and a lower pixel intensity indicates a lower tongue body position.

The time function of the mean pixel intensity in each ROI, smoothed by locally-weighted linear regression (Atkeson et al., 1997) to reduce the influence of image noise, are dubbed articulatory feature trajectories. These trajectories provide information regarding tissue movement into and out of the ROIs, reflecting the formation and release of gestures made by each effector of interest. The articulatory feature trajectories for a fluent trial of /mæb/ are shown in Fig. 2. The mean pixel intensity (y-axis), as mentioned above, provides an estimate of the kinematics of lip closure and opening, tongue movement into and out of the alveolar and palatal regions, or velum raising and lowering. For each speech sound in /mæb/, the kinematics of the composing gestures can be observed from these articulatory trajectories. Both /m/ and /b/ involve a lip closing gesture, which is seen as local peaks in the LIP trajectory, with the moments of maximal labial closure for each gesture marked in Fig. 2. The nasal sound /m/ also requires a velum lowering gesture, which is opposite to the velum raising gesture required for /b/. With the low-positioned VEL ROI, low pixel intensity in the VEL trajectory means a raised velum. Hence, we see a smooth lowering of the VEL trajectory, indicating that the velum kept raising throughout the production of the entire word /mæb/. When the word was finished the velum was then lowered again to its rest position, which is seen as the VEL trajectory rising again at the end. The low vowel /æ/ requires a tongue body lowering gesture, which is seen as the decrease of pixel intensity in the TB trajectory. The trough in the TB trajectory indicates the moment of maximal tongue body lowering for the vowel. Note that the

lowering of the tongue body started approximately at the same time with the closing of the lips, suggesting that the consonant gesture and the vowel gesture were initiated relatively synchronously. This pattern was also observed in PWNS's production of CV syllables (Löfqvist & Gracco, 1999).



**Fig. 2.** Time functions of mean pixel intensity in four ROIs: lips (LIP), tongue tip (TT), tongue body (TB), and velum (VEL) for a fluent trial of /mæb/. The dotted lines respectively indicate the moment of maximal lip closure for /m/, maximal

tongue body lowering for /æ/, and maximal lip closure for /b/.

We were interested in the movement of the primary effector(s), which are effector(s) deployed for making the critical gesture(s) for a given speech sound. Therefore, the corresponding articulatory feature trajectories were examined. Specifically, we examined the VEL trajectory for all the consonants, the LIP trajectory for /m/, the LIP trajectory for /f/, the TT trajectory for /t/, the TT and TB trajectories for /ʃ/ (the production of palato-alveolar /ʃ/ require the tongue blade to approach the back of the alveolar ridge, which means parts of the tongue will enter both TT and TB ROIs), and the TB trajectories for vowels.

Articulatory feature trajectories for disfluent trials were analyzed in two ways. First, we qualitatively examined the trajectories of oral gestures for disfluent speech sounds. The upward and downward slopes in trajectories in Fig. 2 can be characterized by single-peaked velocity time functions, indicating a smooth formation and release of gestures in perceptually fluent speech of this PWS, as is

typical in fluent speech of PWNS (examples can be found in e.g., Lammert et al., 2010; Proctor et al., 2011). Yet, when disfluencies occur, different articulatory kinematics like oscillation (multiple velocity peaks) and static posturing are expected to occur. Hence, we characterized the articulatory patterns of disfluent gestures and evaluated whether or not they could be matched up with the perceptual types of disfluencies.

Second, we estimated the magnitude of both consonant and vowel gestures during disfluencies and compared them to the magnitude of matched gestures in fluent speech. The purpose is to identify any systematic differences between disfluent and fluent gestures and any differences in degree of CV coarticulation (indicated by the magnitude of V gestures during consonant production). As shown in Fig. 2, a peak or a trough in the articulatory feature trajectories indicate the maximum magnitude achieved by an articulatory gesture (e.g., the maximum constriction of the lips, the maximum lowering/raising of the velum, the maximum lowering/raising of the tongue body). Therefore, we measured the pixel intensity at these peaks and troughs. When multiple peaks or troughs (due to articulator oscillation) occurred during one instance of disfluency, we took the average value. Assuming that the magnitude of fluent gestures represents the typical goal, if disfluent gestures turn out to have a greater magnitude compared to fluent gestures, it suggests that gestures are overshoot during disfluencies. The opposite case would be gestural undershoot.

Additionally, since oscillatory movement and relatively static posturing of the primary effector have been observed during disfluencies (Zimmermann, 1980b), in order to examine whether the constriction magnitude changes through repeated constriction gestures or a sustained constriction state, we compared the constriction magnitude at the beginning and the end of these articulatory events. In the case of articulator oscillation, the beginning and the end are defined as the first peak and the last peak of the articulatory feature trajectory. In the case of sustained constrictions, the beginning and the end are the points in the articulatory feature trajectory between which the velocity falls below 20% of the total velocity range within the disfluency.

## 2.6. Statistical analysis

Cohen's kappa ( $\kappa$ ) with an alpha level of .05 was used to measure the intra- and inter-rater reliability in the perceptual ratings of disfluencies. Since individual ROIs differ in size and hence the average pixel intensity in each ROI has different ranges, the pixel intensity measurements of all trials (fluent and disfluent) for a given ROI were first normalized using z-scores calculated based on the mean and standard deviation for that ROI. By doing so, gestures involving different ROIs could be considered in the same analysis. The means of pixel intensity values associated with instances of a fluent gesture and their disfluent counterparts are compared using two-sample *t*-tests ( $\alpha = .05$ ). The 95% confidence intervals (CIs) of the differences in means were also calculated. Paired sample *t*-tests ( $\alpha = .05$ ) were used to compare the constriction magnitude at the beginning and the end of the repeated or sustained constrictions during disfluencies. All statistical analyses were done in R (R Core Team, 2020).

## 3. Results

### 3.1. Perceptual rating of stuttering disfluencies

In this dataset, different types of stuttering disfluencies were commonly perceived to have co-occurred in the production of one single speech sound. For example, disfluencies rated as a combination of block and prolongation/repetition are cases where a block is perceived before a prolongation/repetitions of the speech sound. A combination of prolongation and repetition are cases where each repetition of the speech sound is prolonged. A combination of block, prolongation, and repetition are cases where a block is perceived first, and then repetitions of the prolonged speech sound.

The intra-rater reliability between the two ratings of the first rater was very high,  $\kappa = .97$ , 95% CI: [.93, 1], meaning “almost perfect agreement” (Landis & Koch, 1977). Only two instances of disfluencies were rated differently, and they were errors made in the first rating and were therefore corrected in the second rating. The inter-rater reliability between the corrected rating of the first rater and the rating of the second rater is also high,  $\kappa = .74$ , 95% CI [.64, .83], meaning “substantial agreement” (Landis & Koch,



1977). Most of the disagreement between the two raters was on whether there was a block in combination with other disfluencies. This is not surprising because the participant's facial expressions, which are important for identifying blocks, were not available in the audio or mid-sagittal vocal tract video. Further analyses were based on the corrected rating of the first rater (shown in supplementary material, Table 1), which was more likely to identify a block in mixed disfluencies.

### **3.2. Overview of stuttering disfluencies**

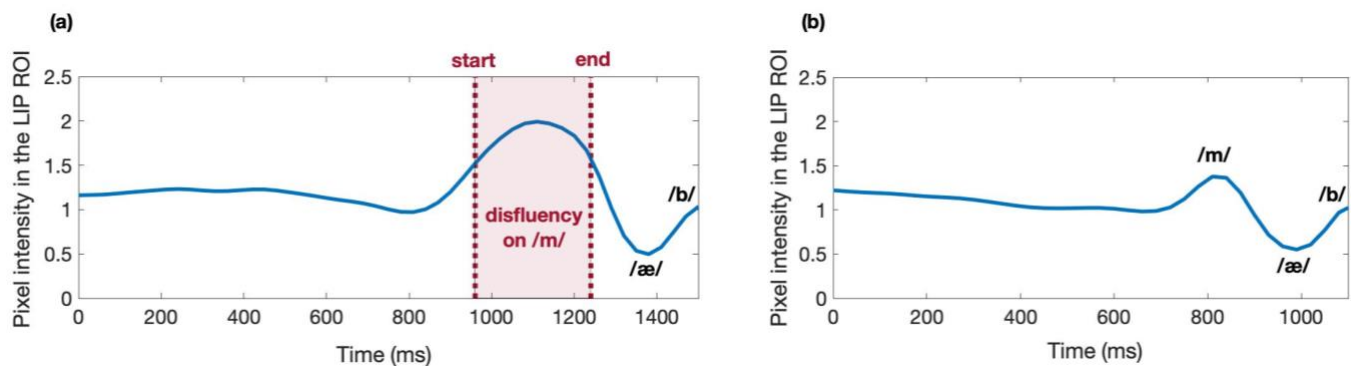
A total of 42 instances of disfluencies were identified in 24 disfluent trials. They all occurred in the production of syllable-initial consonants, some of which were also word-initial consonants. Twenty-two instances occurred on /m/ in the word-initial syllable /mæb/, 5 on /f/ in /faɪ/, 6 on /ʃ/ in /ʃeɪ/ or /ʃeɪb/, 9 on /t/ in /ti/ or /taɪ/. From the distribution of disfluent speech sounds, we see that syllable-initial consonants, especially word-initial consonants, were more likely to be subject to stuttering, compared to syllable-final consonants. This is consistent with the position factor and the phonetic factor known for a long time to influence the occurrence of disfluency (discussions date back to Brown's work in the 1930s and 1940s, e.g., Johnson & Brown, 1935; Brown, 1938).

Twenty-two trials were identified as fluent (see supplementary material, Table 2). All trials of shorter words /mæb/ and /mæbʃaɪb/ were identified as fluent. As for longer words /mæbʃaɪʃeɪb/, /mæbʃeɪtɑɪdɔɪb/, and /mæbtɪbɪbɪ/, only one fluent trial was identified for each word. This is consistent with the observation that stuttering is more frequent when utterances are long or phonologically complex or both (e.g., Brown, 1945; Smith et al., 2010; Max et al., 2019).

### **3.3. Articulatory patterns during disfluencies**

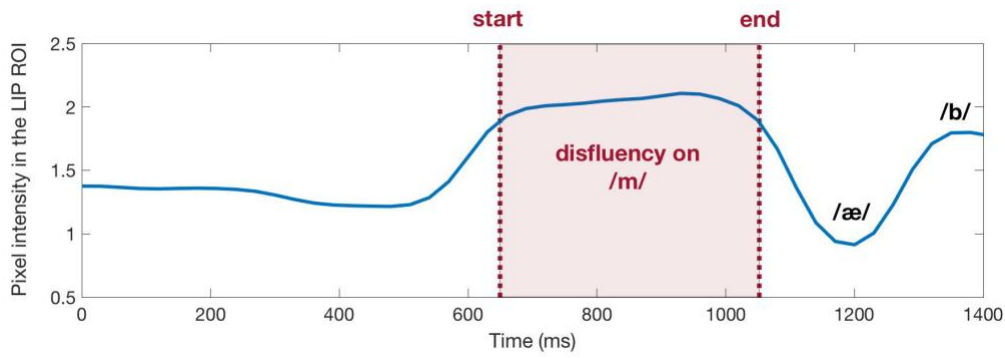
We first qualitatively characterized the articulatory feature trajectories of the primary oral effectors for the disfluent consonants. Three articulatory patterns during disfluencies were identified based on the overall shapes of these trajectories. The first pattern resembles the spatiotemporal pattern of fluent gestures. In Fig. 3, an example of a labial gesture during disfluent /m/ is juxtaposed with a labial

gesture for a fluent /m/. Both LIP trajectories have one single peak for /m/, indicating a smooth formation and release process of the lip closure. One may notice that the peak during disfluency is associated with a higher pixel intensity compared to the peak for a fluent /m/, which means the lip closure was more extreme when /m/ was stuttered. The magnitude of disfluent gestures and fluent gestures are systematically compared in subsection 3.4. All instances of articulatory pattern 1 are listed in supplementary material, Table 3. They were found in disfluencies perceived as block, or prolongation, or block followed by prolongation.



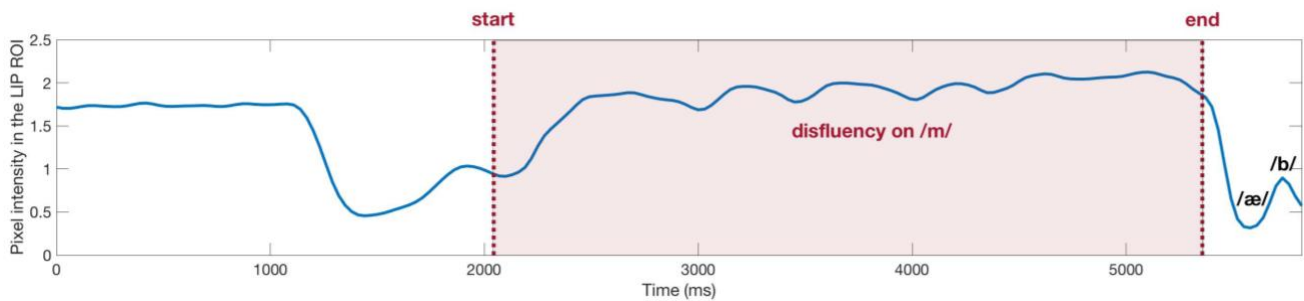
**Fig. 3.** Lip trajectories for (a) an example of disfluent /m/ (from trial AO in supplementary material, Table 1), characterized by articulatory pattern 1, and (b) an example of fluent /m/ (from trial AG in supplementary material, Table 1).

The second articulatory pattern features a plateau in the trajectory of primary oral effector, which indicates that the constriction action of the effector was sustained during disfluencies. For example, the trajectory shown in Fig. 4 suggests that the lips were fixated at an intense closure position during the disfluency. The subtle rise in the plateau indicates that the lips became even more closed and compressed towards the end of disfluency. More analysis regarding this tendency is in subsection 3.5. Articulatory pattern 2 was observed in disfluencies perceived as prolongation or block followed by prolongation (see supplementary material, Table 3).



**Fig. 4.** Lip trajectory for an example of disfluent /m/ (from trial AP in supplementary material, Table 1), characterized by articulatory pattern 2.

The third articulatory pattern features the oscillation of the primary oral effector, as shown in Fig. 5. Like articulatory pattern 2, pattern 3 also shows a subtle upward trend towards a more extreme constriction throughout the oscillation. Articulatory pattern 3 was found in repetitions integrated with other types of disfluencies. But four instances of disfluencies were perceived as repetition in combination with prolongation and/or block were not characterized by pattern 3, because a plateau was found along with the oscillation. Their articulatory patterns were categorized as pattern 2 plus pattern 1 or 3, depending on whether there was only one peak or multiple successive peaks in the trajectory apart from the plateau (see supplementary material, Table 3).



**Fig. 5.** Lip trajectory for an example of disfluent /m/ (from trial AA in supplementary material, Table 1), characterized by articulatory pattern 3.

The three articulatory patterns summarized from the kinematic profiles of oral gestures for disfluent consonants did not show a strict one-to-one correspondence in relation to the three perceptual

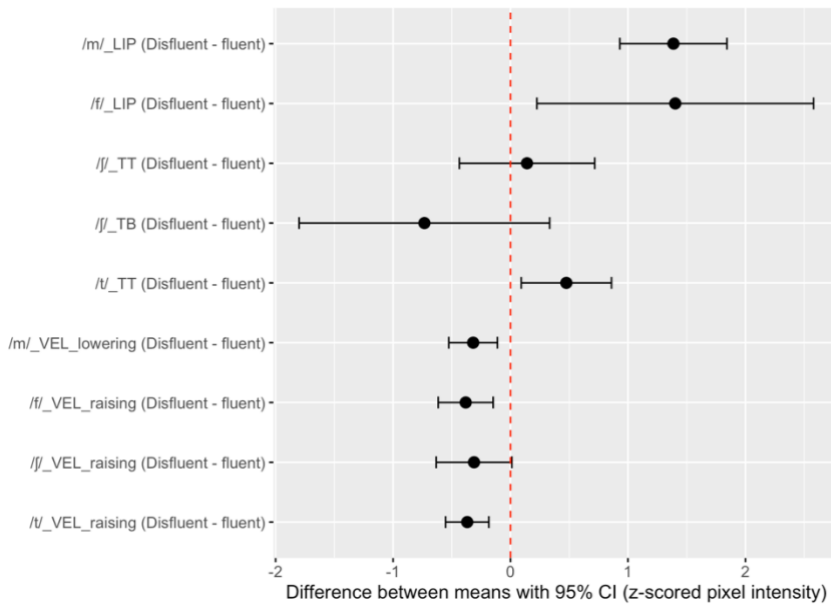
types of disfluency (block, prolongation, repetition), but the mapping between them also did not seem completely irregular. Based on articulatory realizations, a distinction can be drawn between repetition and the other two perceptual types of disfluencies. Repetition was the result of either articulatory pattern 3 or a combination of several articulatory patterns, whereas pure block or pure prolongation were related to articulatory patterns 1 and 2, but not 3. Disfluencies perceived solely as block or prolongation could both have the articulatory realization as pattern 1, but pure prolongation could also result from articulatory pattern 2.

### 3.4. Magnitude of articulatory gestures during disfluencies

As seen in the articulatory trajectories above (Fig. 3), the lip gesture for a disfluent /m/ exhibited a higher constriction magnitude compared to the fluent counterpart. We systematically examined the maximum magnitude of both oral and velum gestures for disfluent consonants and compared that to the maximum magnitude of gestures for fluent consonants. Fluent consonants included the analysis were taken not only from fluent trials but also from the trials in which disfluencies occurred on the non-target consonant. One instance of disfluent /t/ was excluded from the analysis because swallowing was involved.

Two-sample *t*-tests yielded a significant effect of fluency on the magnitude of lip constriction for /m/,  $t(42) = 6.14$ ,  $p < .0001$ , lip constriction for /f/,  $t(6) = 2.91$ ,  $p = .027$ , tongue tip constriction for /t/,  $t(15) = 2.64$ ,  $p = .019$ , velum lowering for /m/,  $t(42) = -3.10$ ,  $p = .003$ , velum raising for /f/,  $t(6) = -3.99$ ,  $p = .007$ , velum raising for /t/,  $t(15) = -4.26$ ,  $p < .001$ , but not for tongue tip constriction for /ʃ/,  $t(14) = 0.52$ ,  $p = .61$ , tongue body constriction for /ʃ/,  $t(14) = -1.47$ ,  $p = .16$ . The difference between the velum raising magnitude for disfluent /ʃ/ and fluent /ʃ/ was marginally significant,  $t(14) = -2.07$ ,  $p = .058$ . The 95% confidence intervals for the mean differences between the fluent and disfluent instances of each gesture are plotted in Fig. 6. Mean differences and confidence limits are given in Table 1. Consistent with the result from *t*-tests, the confidence intervals for mean differences, apart from the ones for /ʃ/ gestures, do not contain zero. Boxplots comparing the magnitude of fluent and disfluent consonant gestures are given in Fig. 7, with oral and velum gestures plotted in separate panels. Descriptive statistics are given in Table

2. The oral gestures, except for the ones for /ʃ/, exhibited a higher constriction magnitude in disfluent speech, as indicated by higher pixel intensity in the ROIs (Fig. 7a). The velum raising gestures for oral consonants (/f/, /ʃ/, /t/) also showed a higher magnitude of raising during disfluencies, as indicated by lower pixel intensity in the VEL ROI (Fig. 7b), hence resulting in a more constricted nasopharynx. The velum lowering gesture for /m/ showed a smaller magnitude of lowering during disfluencies, as indicated by lower pixel intensity in the VEL ROI (Fig. 7b), which means a less opened nasopharynx. Overall, the velum gestures in disfluent speech showed an increased level of constriction in nasopharynx for both nasal and oral consonants, regardless of the aperture goals.

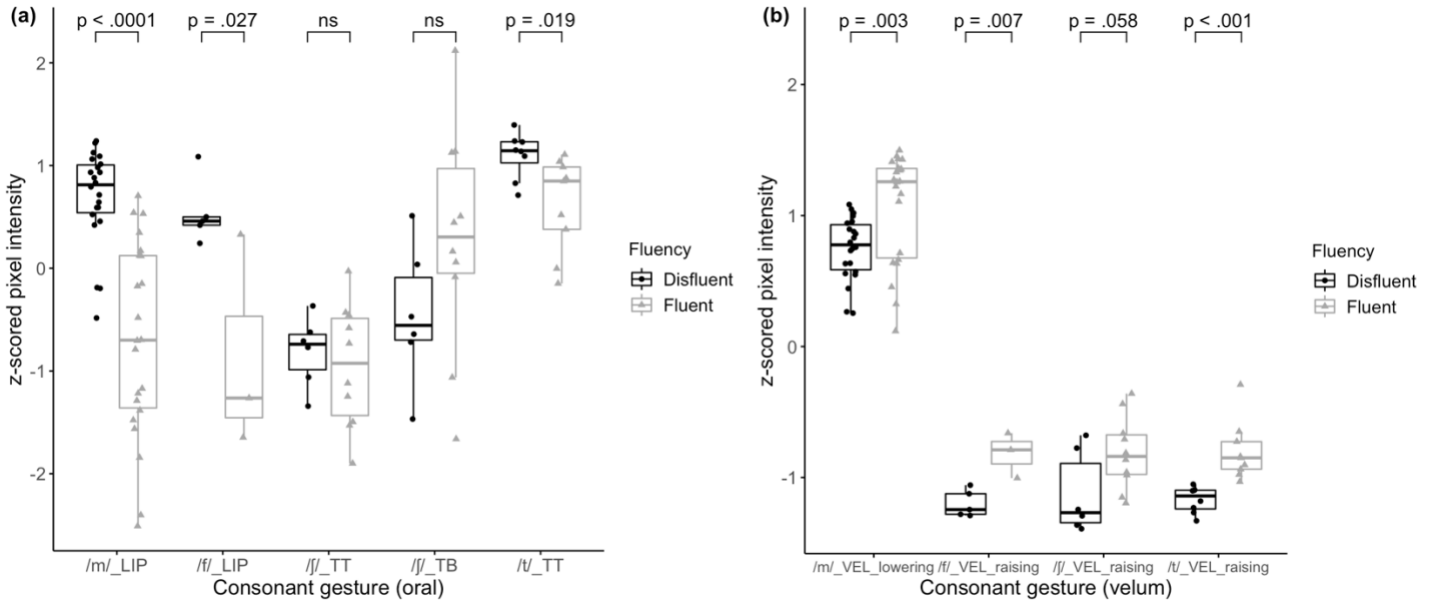


**Fig. 6.** 95% confidence intervals for the differences between the means of z-scored pixel intensity for the fluent and disfluent instances of each consonant gesture. If the interval does not contain zero (indicated by the red dashed line), the corresponding means are significantly different.

**Table 1.** Mean differences of z-scored pixel intensity between the fluent and disfluent instances of each consonant gesture and the corresponding 95% confidence limits.

	/m/_LIP	/f/_LIP	/ʃ/_TT	/ʃ/_TB	/t/_TT	/m/_VEL_lowering	/f/_VEL_raising	/ʃ/_VEL_raising	/t/_VEL_raising
Mean difference	1.39	1.4	0.14	-0.73	0.48	-0.31	-0.38	-0.31	-0.37
Lower confidence limit	0.93	0.22	-0.44	-1.8	0.09	-0.52	-0.62	-0.63	-0.55

Upper confidence limit	1.84	2.58	0.71	0.33	0.86	-0.11		-0.15	0.01	-0.18
------------------------	------	------	------	------	------	-------	--	-------	------	-------



**Fig. 7.** Boxplots showing the normalized magnitude of (a) oral gestures and (b) velum gestures for disfluent and fluent consonants, estimated by the z-scored pixel intensity in individual ROIs.

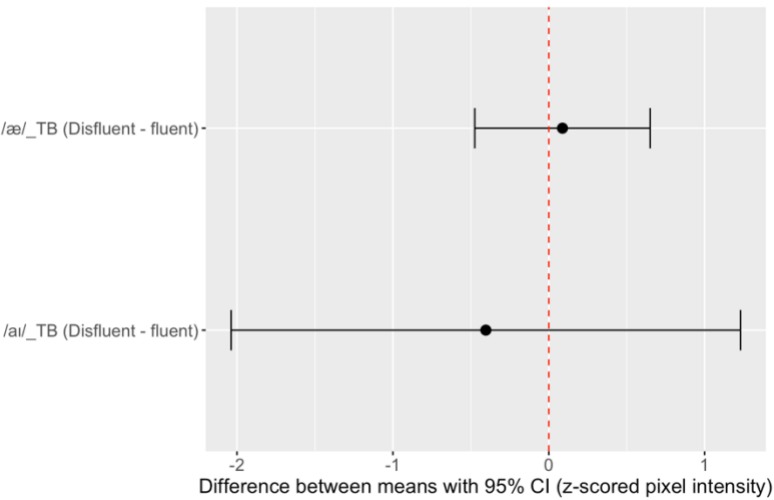
**Table 2.** Number of observations (*N*), mean (*M*) and standard deviation (*SD*) of z-scored pixel intensity in individual ROIs, representing the normalized magnitude of disfluent and fluent consonant gestures.

		/m/_LIP	/f/_LIP	/j/_TT	/ʃ/_TB	/t/_TT	/m/_VEL_lowering	/f/_VEL_raising	/j/_VEL_raising	/t/_VEL_raising
Disfluent	<i>N</i>	22	5	6	6	8	22	5	6	8
	<i>M</i>	0.69	0.54	-0.81	-0.46	1.10	0.75	-1.20	-1.12	-1.17
	<i>SD</i>	0.47	0.32	0.34	0.68	0.22	0.24	0.10	0.31	0.10
Fluent	<i>N</i>	22	3	10	10	9	22	3	10	9
	<i>M</i>	-0.70	-0.86	-0.95	0.28	0.62	1.07	-0.82	-0.81	-0.80
	<i>SD</i>	0.95	1.05	0.60	1.09	0.46	0.42	0.17	0.28	0.23

Given that disfluencies all occurred on syllable-initial consonants, we also examined the magnitude of tongue body gestures during disfluencies to see if the vowel gesture, which is coproduced

with the consonant gesture in the fluent speech of PWNS (Löfqvist & Gracco, 1999), was still initiated during stuttering disfluencies, despite the atypicality of the consonant gesture. The comparison between vowel gestures in disfluent and fluent speech was possible for only two syllables: /mæb/ and /faɪ/. Syllables that begin with /ʃ/ or /t/ were excluded from this analysis because first, both /ʃ/ and following vowel can affect the position of tongue body during disfluencies, thus introducing confounds; second, only one instance of fluent /ti/ was available in the dataset, which was not sufficient to be compared to multiple instances of disfluent /ti/. Since the vowels in /mæb/ and /faɪ/ both involve a low vowel position, we estimated the maximum magnitude of the tongue body lowering by measuring the minimum pixel intensity in the TB trajectory.

From the result of two-sample *t*-tests, there is no significant effect of fluency on the magnitude of tongue body lowering gesture for /æ/,  $t(42) = 0.32, p = .75$ , nor for /aɪ/,  $t(6) = -0.60, p = .56$ . Fig. 8 and Table 3 show the confidence intervals for the differences in means, which are consistent with the result of *t*-tests. Boxplots comparing vowel gestures in disfluent and fluent speech are given in Fig. 9. Descriptive statistics are given in Table 4.

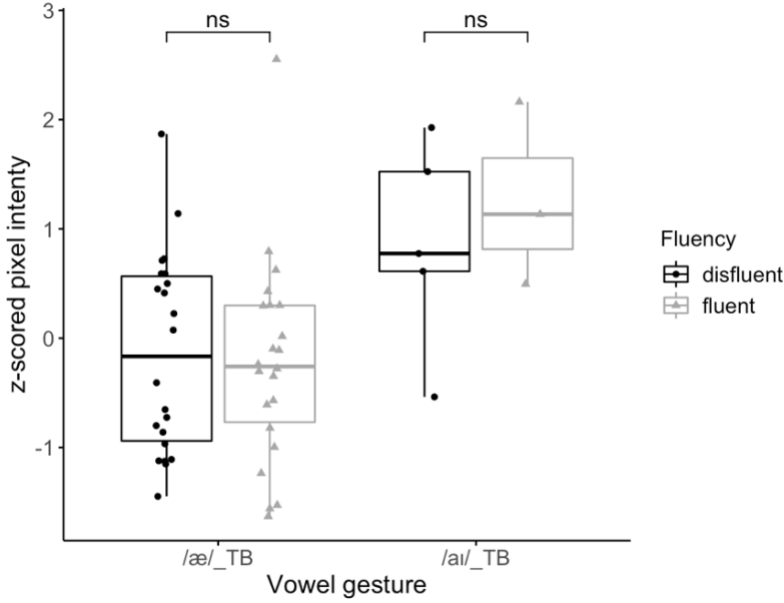


**Fig. 8.** 95% confidence intervals for the differences between the means of z-scored pixel intensity for the fluent and disfluent instances of each vowel gesture. If the interval contains zero (indicated by the red dashed line), the corresponding means are not significantly different.

**Table 3.** Mean differences of z-scored pixel intensity between the fluent and disfluent instances of each vowel gesture and the corresponding 95% confidence limits.

	/æ/_TB	/aɪ/_TB
--	--------	---------

Mean difference	0.09	-0.4
Lower confidence limit	-0.48	-2.04
Upper confidence limit	0.65	1.23



**Fig. 9.** Boxplots showing the normalized magnitude of vowel gestures in disfluent and fluent speech, estimated by the z-scored pixel intensity in the TB ROI.

**Table 4.** Number of observations (*N*), mean (*M*) and standard deviation (*SD*) of z-scored pixel intensity in the TB ROI, representing the normalized tongue body lowering magnitude for vowel gestures in disfluent and fluent speech.

		/æ/_TB	/a/_TB
Disfluent	<i>N</i>	22	5
	<i>M</i>	-0.14	0.86
	<i>SD</i>	0.91	0.95
Fluent	<i>N</i>	22	3
	<i>M</i>	-0.23	1.26
	<i>SD</i>	0.94	0.84

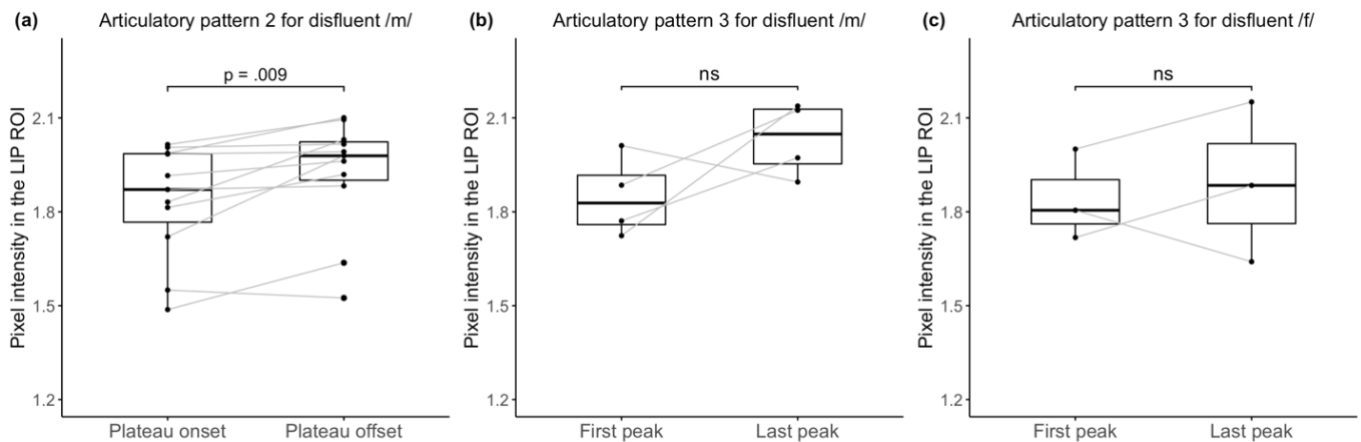
To summarize, comparisons between the maximum magnitude of gestures in disfluent and fluent speech revealed that most of the consonant gestures, except for the tongue tip and tongue body gestures for /j/, and the velum lowering gesture for /m/, exhibited more extreme magnitude during disfluency. But the less extreme velum lowering for /m/ fits the pattern that the consonant gestures tended to produce a



more constricted airway. The tongue body lowering magnitude of vowel gestures showed no difference across disfluent and fluent speech.

### 3.5. Correction of articulatory gestures during disfluencies

Since articulatory patterns 2 and 3 involve repeated or sustained oral constrictions, we also tested whether there was any correction of the extreme constriction towards the end of these articulatory patterns. Due to the small size of the dataset, we were only able to examine patterns 2 and 3 for disfluent /m/, and pattern 3 for disfluent /f/. Paired sample *t*-tests were used to compare the pixel intensity in the LIP ROI (labial constriction magnitude) at the onset and offset of the plateau in articulatory pattern 2 (plateau onset and offset were determined by the threshold of 20% velocity range, see the methods section for details), and at the first peak and last peak of the oscillation in articulatory pattern 3. As shown in Fig. 10, for disfluent /m/, pixel intensity was significantly less for the plateau onset ( $N = 11$ ,  $M = 1.83$ ,  $SD = 0.18$ ) than the offset ( $N = 11$ ,  $M = 1.92$ ,  $SD = 0.18$ ) in articulatory pattern 2,  $t(10) = -3.24$ ,  $p = .009$ , indicating the opposite of correction; and no significant difference was found in the pixel intensities for the first ( $N = 4$ ,  $M = 1.85$ ,  $SD = 0.13$ ) and last peak ( $N = 4$ ,  $M = 2.03$ ,  $SD = 0.12$ ) in articulatory pattern 3,  $t(3) = -1.67$ ,  $p = 0.19$ . Likewise, for disfluent /f/, no significant difference was found in the pixel intensities for the first ( $N = 3$ ,  $M = 1.84$ ,  $SD = 0.15$ ) and last ( $N = 3$ ,  $M = 1.89$ ,  $SD = 0.26$ ) peak in articulatory pattern 3,  $t(2) = -0.47$ ,  $p = 0.68$ . Hence, no evidence for correction of the extreme oral constrictions was found during the production of disfluent consonants.



**Fig. 10.** Boxplots showing comparisons between the beginning and ending magnitudes of labial constrictions for disfluent /m/ and /f/ characterized by articulatory patterns 2 and 3.

#### **4. Discussion**

To summarize, in this dataset from a single adult man who stutters, all stuttering disfluencies occurred in the production of syllable-initial consonants. We found that the supralaryngeal articulatory gestures in these disfluencies could be characterized as follows:

(1) Movements of the primary oral effector for making the consonant constriction showed three kinematic patterns, but these patterns did not necessarily give rise to distinct perceptual types of disfluencies.

(2) Oral gestures for disfluent consonants typically exhibited a constriction magnitude more extreme than the matched gestures in the speaker's fluent speech. Such extreme constrictions were rarely corrected during disfluencies and sometimes showed even more constriction towards the end of disfluencies. Likewise, velum gestures for disfluent consonants were also more constricted during disfluencies.

(3) Gestures for the vowel following the disfluent consonant were already initiated and executed during the disfluencies perceived as stuttered onset consonants, achieving a tongue body position no different from vowel gestures in fluent speech.

##### **4.1. Articulatory patterns and perceptual types of disfluencies**

Three articulatory patterns were summarized based on the kinematic profiles of oral gestures for disfluent consonants. Pattern 1 showed movement trajectories that are similar to those of fluent gestures, whereas patterns 2 and 3 respectively featured the fixation and oscillation of the oral effector, which gave rise to atypical spatiotemporal patterns of gestures. Articulatory patterns 2 and 3 suggest that the difficulty of releasing a consonant constriction constitutes an important component of the articulatory behaviors of stuttering. In pattern 2, the release of the constriction coincided with the end of the disfluency (Fig. 4). In

pattern 3, each cycle in the oscillation can be viewed as an unsuccessful release of the oral constriction. Likewise, the full release coincided with the end of the disfluency (Fig. 5).

From a theoretical point of view, difficulties releasing the constriction can be understood as an issue with terminating the consonant gesture, as Guenther (2016, p. 296)'s hypothesis of stuttering states: "the core deficit in persistent developmental stuttering is an impaired ability to initiate, sustain, and/or terminate motor programs for phonemic/gestural units within a speech sequence due to impairment of the left-hemisphere basal ganglia motor loop." Within the framework of Articulatory Phonology, the release problem can be further related to the positional and phonetic factors that influence the occurrence of stuttering. It is possible that syllable-initial consonants are susceptible to stuttering because the intergestural coordination in the syllable-initial position can be analyzed as involving competition and therefore might pose higher demands on the speech motor control system. Nam (2007) proposed that the closure gesture and the release gesture of the onset consonant are coordinated synchronously with the following vowel, thus forming a competitive coupling relation between them. In contrast, the closure gesture and the release gesture of a coda consonant are hypothesized to be coordinated in a sequential manner starting from the preceding vowel, thus forming no competition. Disfluencies may arise due to the motor complexity of coordinating competing gestures in the syllable-initial position, resulting in difficulties releasing the gesture of the onset consonant. Similar motor disruptions resulting from competitive coupling of gestures have been observed in essential tremor patients, who have shown difficulties coordinating competing consonant gestures in onset consonant clusters when receiving deep brain stimulation (Hermes et al., 2019). Although this hypothesis of gesture competition for onset consonant can account for several features of stuttering, we have to point out that it does not generate the right predictions regarding the distribution and the acquisition of different syllables. It predicts that CV syllables, due to its more complicated gestural organization, should be rarer than VC syllables in languages, and that children should acquire CV syllables later than VC syllables. Neither of the

predictions are true (Nam et al., 2009). Hence, further work is needed to verify this hypothesis and to see if it provides a better account for stuttering.

When it comes to the relationship between the articulatory patterns and the perceptual types of disfluencies, although articulatory pattern 3 was always related to repetition, pure blocks and pure prolongations did not necessarily correspond to distinct articulatory patterns. This supports the practice of combining these two perceptual types as one category—audible or inaudible prolongation (e.g., Van Riper, 1982). In other words, block and prolongation may not differ in terms of the movement of oral effectors. What makes block distinct from prolongation is likely to be the constriction at the velum or larynx that could effectively affect phonation.

Although our method of developing articulatory descriptions of stuttering disfluencies differs from the method in Didirkova et al. (2019) as we focussed on the movement of primary effectors, rather than considering the global movement of multiple articulators, we reached the same conclusion as Didirkova et al. (2019): no strict one-to-one correspondence was found between the perceptual outcomes of disfluencies and their underlying articulatory kinematic events. This highlights the importance of investigating articulatory characteristics of stuttered speech, since the subtleties of articulatory behaviors are not completely reflected in, and therefore, cannot be safely inferred from perceived disfluencies. RT-MRI, compared to other commonly used vocal tract imaging/tracking techniques (e.g., ultrasound, EMA), captures the dynamic information of the entire vocal tract without interfering with the speech movements, and thus can make the task of collecting accurate articulatory kinematic data of disfluent speech more feasible.

#### **4.2. Consonant gestures during disfluencies**

Lip gestures for disfluent /m/ and /f/, and tongue tip gestures for disfluent /t/, were found to have a higher constriction magnitude than their fluent counterparts. However, the constriction magnitude of the tongue tip and tongue body gestures for disfluent /j/ were generally no different from the fluent ones. Assuming that the constriction magnitude of fluent gestures represents the typical goal of a gesture, the

results suggest that most of the oral gestures for disfluent consonants exceeded the typical goal, and therefore, were overshoot. For gestures making a full closure in the vocal tract, such overshoot can be understood as involving greater contact force or greater volume of tissue contact than fluent gestures.

The velum gestures for consonants were also more constricted (raised) during disfluencies, regardless of whether the goal of the velum gesture is to be raised or lowered. In the case of velum gestures for oral consonants, whose goal is to be raised, more constriction means the gestures were overshoot. Yet, as for the velum gestures for /m/, whose goal is to be lowered, a more constricted (less lowered) velum means gestural undershoot.

Considering the gestures that were overshoot during disfluencies, they are gestures whose goal is to create complete closure in the vocal tract, with the exception of the lip gesture for /f/. By contrast, gestures that were not overshoot (i.e., the tongue tip and tongue body gestures for /j/ and the velum lowering gesture for /m/) have goals of creating partial closure or opening in the vocal tract. Taken together, the most consistent finding is that gestures with a goal to form complete closure in the vocal tract were overshoot during stuttering disfluencies.

Gestural overshoot is in line with the hypothesis considering overreliance on sensory feedback as the cause of stuttering (Max et al., 2004; Civier et al., 2010). This hypothesis proposes that PWS have weak feedforward control and hence have to overly rely on feedback control. In Civier et al. (2010)'s modeling work, a high gain on feedback control and a low gain on feedforward control resulted in articulatory overshoots and oscillations, due to delays inherent in afferent feedback processing (Guenther et al., 2006). On the other hand, the hypothesis proposing that overreliance on sensory feedback is a strategy used by PWS to compensate for their limited speech motor skill and to prevent stuttering (Namasivayam & van Lieshout, 2008; Namasivayam & van Lieshout, 2011) is unsupported by the observation of gestural overshoot. According to that line of reasoning, stuttering moments arise because of the insufficient sensory feedback caused by small movement amplitude (gestural undershoot), rather than overshoot. Yet, that hypothesis may still be valid if the triggers of stuttering can be separated from

the characteristics of stuttering. Small movements may trigger stuttering, but once stuttering is triggered, it can manifest as extreme constrictions. In that regard, it will be useful to examine articulatory movements right before the occurrence of disfluencies.

We further observed that extreme constrictions were not corrected during repeated or sustained oral gestures. One may suggest that it is due to neural processing delays, e.g., time lags associated with receiving sensory feedback, preparing and executing feedback-based corrective commands. However, disfluencies involving repeated or sustained constrictions lasted long enough to compensate for such delays and allow feedback-based correction to manifest. Sustained constrictions during disfluencies (articulatory pattern 2) on average contained a 500-ms period in which constriction was greater than the fluent level. Repetitive constricting actions during disfluencies (articulatory pattern 3) involved an even longer over-constricted period, which was on average 1625 ms. By contrast, delays in feedback processing are typically assumed to be up to 150 ms, as suggested by response latencies in sensory feedback perturbation experiments. Hence, the absence of correction should not be ascribed to the disfluencies being too “short”.

Alternatively, the lack of real-time correction could be predicted by the hypothesis that PWS cannot make effective use of sensory feedback, especially kinesthetic information (Archibald & De Nil, 1999; Loucks & De Nil, 2006a, 2006b). Given that auditory feedback is compromised during some disfluencies because of absent or disrupted phonation, the possible inhibition of kinesthetic feedback in PWS may play a prominent role in preventing correction of the extreme constrictions. A recent neuroimaging study by Connally et al. (2018) found a greater activation the parietal operculum, where the secondary somatosensory cortex is located, during disfluent states relative to fluent states in PWS. The overactivity of this region could be an indicator of inefficient somatosensory feedback processing during episodes of stuttered speech (Tourville et al., 2008).

#### **4.3. Vowel gestures during disfluencies**

In fluent speech of typical speakers, the vowel gestures are activated relatively synchronously with the gestures for syllable-initial consonant (Löfqvist & Gracco, 1999). As for stuttered speech of PWS, we also checked whether the gestures for the following vowel were initiated and executed during disfluencies that were perceived as stuttered onset consonants. Results show that for the vowels in our analysis, which all require low tongue positions, the tongue body was lowered to the degree for producing a fluent low vowel during the consonant disfluencies.

These results suggest that the vowel gestures were already initiated and executed together with the syllable-initial consonant gestures during disfluent speech. In other words, CV coproduction was present even when the consonant gestures were interrupted and perceived as disfluent, and the degree of CV coproduction was commensurate with that of fluent speech. This is consistent with other experimental findings showing that fluent and non-fluent adults demonstrate similar efficiency when encoding the medial, stress-bearing vowel in a silent phoneme monitoring task (Jacobs, 2016). Hence, stuttering does not appear to be caused by a CV coproduction issue. From this perspective, the movement of secondary articulators like the jaw and tongue accompanying the oscillation or fixation of the primary articulator observed in Zimmermann (1980b) could be the coproduced V gesture, instead of the repositioning of articulators to a rest position.

Hypotheses proposing that stuttering arises from a phonological encoding issue (e.g., EXPLAN: Howell, 2004; Covert Repair Hypothesis: Postma & Kolk, 1993; Fault-line Hypothesis: Wingate, 1988) are not supported by the existence of CV coproduction during disfluencies. These hypotheses predict that the motor plan of the vowel subsequent to the disfluent consonant would not be initiated during disfluencies, because stuttering on the consonant is assumed as a coping strategy to gain extra time to deal with the delayed or erroneous encoding of the upcoming vowel. However, it is possible to reformulate these hypotheses in a way that would find support for them. In Articulatory Phonology, the release of the consonant constriction and the formation of the tongue shape for vowels are hypothesized to be separate gestures that are planned to occur at different times, which has been supported in a stop-signal experiment

(Tilsen & Goldstein, 2012). It is possible to propose therefore that inefficiencies when planning the release gesture of the onset consonant (but not the tongue shaping gesture for the vowel) can cause stuttering. This is consistent with observations of the coincidence of successful release and the end of disfluency made in subsection 4.1.

#### **4.4. Limitations**

The characteristics of supralaryngeal articulatory gestures in stuttered speech summarized in this paper are based on one single adult who stutters. As pointed out in Van Riper (1982), every PWS is unique in terms of their stuttering behaviors. It is unclear how much of our findings reflect the uniqueness of the single participant in the study and how much can be generalized to other PWS. Future work should be dedicated to collecting more data from more PWS, to investigate individual differences in articulatory behaviors during stuttered speech, as well as the commonalities among these individual differences.

Some of our findings might be biased due to the limitations of the experimental design. The stimuli were designed for the study from which this case study is taken, rather than being designed specifically for investigating disfluent speech. For example, the number of syllables with a low vowel and the number of syllables with a high vowel were not balanced in the stimuli. Furthermore, all the pseudowords began with the same syllable with a low vowel. Given that word-initial and syllable-initial positions are susceptible to stuttering, all these factors contributed to the fact that the vowel gestures from stuttered speech that could be examined were mostly low vowels. One may argue that the tongue body lowering gesture we observed during disfluencies was not a planned gesture for producing the vowel, but instead simply the tongue being lowered to its neutral position. In fact, the only case of high vowel gesture during disfluency in this dataset showed a higher-than-fluent tongue body raising degree, but no generalizations could be made based on that one data point. Future research should employ more balanced stimuli with various vowels and consonants to allow more analyses of consonant and vowel gestures and the degree of CV coarticulation during stuttered speech.



Other potential limitations are posed by the RT-MRI analysis and acquisition method used in this study. First, a caveat about the ROI-based method is that, if an ROI is filled with tissues due to extreme constriction, the maximum value of pixel intensity can be reached in that ROI is constrained by the size of the ROI. Pixel intensity associated with consonant oral constrictions in disfluent speech is closer to the ceiling value and therefore has a much lower standard deviation compared to constrictions in fluent speech (see Table 2). Second, only examining the mid-sagittal view of the vocal tract has its limitation in capturing all possible gestures involved in speech, some of which may be less visible from the mid-sagittal plane (such as the constriction for /l/). Although speech stimuli in this study only involve gestures that are visible from the mid-sagittal plane, acquiring RT-MRI data from other planes of interest such as coronal, axial, oblique (Kim et al., 2012; Lingala et al., 2015) may benefit future work that will include more variety of speech stimuli.

Additionally, the extreme constriction observed during disfluencies may reflect a learned response to anxiety caused by stuttering rather than a behavioral symptom resulting from the primary causes of stuttering. Most of the disfluencies examined in this case study are not fleeting disfluencies according to SSI-4, because they last longer than 0.5 seconds. Usually, the longer the disfluencies are, the more severe the stuttering is, and the more likely the articulatory behaviors we observe during disfluencies are a mix of core characteristics of stuttering and reactions to stuttering. Relatedly, the lack of the correction of extreme constrictions in articulatory patterns 2 and 3 could be simply due to that extreme constrictions were learned reactions rather than errors, given the long durations of these disfluencies. Future studies should examine the articulatory behaviors in stuttered speech produced by children who stutter, as well as adults with different levels of stuttering severity, which hopefully can yield some clues regarding how to disentangle learned reactions from core behaviors.

## **5. Conclusion**

In this case study, we used RT-MRI to investigate the articulatory kinematics of consonant and vowel gestures in the stuttered speech produced by an adult man who stutters, and thereby we demonstrated the feasibility of using RT-MRI to study stuttering. We found delayed release and overshoot of certain consonant gestures during disfluencies that were perceived as stuttered syllable onsets. Gestures for the following vowel were initiated and executed during these disfluencies, even though the auditory consequences might not have occurred. These findings suggest that the cause of syllable-initial disfluencies does not lie in the disrupted initiation of the vowel gesture, but rather is localized in the release of the overly constricted onset consonant gesture. A major contribution of the current study is to present articulatory data of stuttered speech, which has been rarely reported in the literature. RT-MRI, as a research tool extensively used in studies of typical speech production, has great potential to advance studies of stuttering. Another contribution of this study is to show that articulatory behaviors constitute an important aspect of the behavioral characteristics of stuttering and serve as a complement to perceptual descriptions. Articulatory behaviors during stuttered speech can function as a probe or window into the neuromotor processes that cause fluency breakdowns.

## References

- Anonymous. (2020). Details omitted for double-blind reviewing.
- Archibald, L., & De Nil, L. F. (1999). The relationship between stuttering severity and kinesthetic acuity for jaw movements in adults who stutter. *Journal of Fluency Disorders*, 24, 25-42.
- Atkeson, C. G., Moore, A. W., & Schaal, S. (1997). Locally weighted learning. *AI Review 11*: 11-73. Kluwer.
- Belyk, M., Kraft, S. J., & Brown, S. (2015). Stuttering as a trait or state - an ALE meta-analysis of neuroimaging studies. *European Journal of Neuroscience*, 41(2), 275-284.
- Browman, C. P. & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, 6(2), 201-251.

- Browman, C. P., & Goldstein, L. (1990). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics*, 18(3), 299-320.
- Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: an overview. *Phonetica*, 49, 155-180.
- Brown, S. F. (1938). Stuttering with relation to word accent and word position. *The Journal of Abnormal and Social Psychology*, 33(1), 112-120.
- Brown, S. F. (1945). The loci of stuttering in the speech sequence. *Journal of Speech Disorders*, 10(3), 181-192.
- Büchel, C., & Sommer, M. (2004). What causes stuttering?. *PLoS Biology*, 2(2), e46.
- Caruso, A. J., Abbs, J. H., & Gracco, V. L. (1988). Kinematic analysis of multiple movement coordination during speech in stutterers. *Brain*, 111(2), 439-455.
- Civier, O., Tasko, S. M., & Guenther, F. H. (2010). Overreliance on auditory feedback may lead to sound/syllable repetitions: simulations of stuttering and fluency-inducing conditions with a neural model of speech production. *Journal of Fluency Disorders*, 35(3), 246-279.
- Connally, E. L., Ward, D., Pliatsikas, C., Finnegan, S., Jenkinson, M., Boyles, R., & Watkins, K. E. (2018). Separation of trait and state in stuttering. *Human Brain Mapping*, 39(8), 3109-3126.
- Conture, E. G., McCall, G. N., & Brewer, D. W. (1977). Laryngeal behavior during stuttering. *Journal of Speech Language and Hearing Research*, 20(4), 661-668.
- Conture, E. G., McCall, G. N., & Brewer, D. W. (1985). Laryngeal behavior during stuttering: a further study. *Journal of Speech Language and Hearing Research*, 29(2), 233-240.
- De Nil, L. F. (1995). The influence of phonetic context on temporal sequencing of upper-lip, lower-lip, and jaw peak velocity and movement onset during bilabial consonants in stuttering and nonstuttering adults. *Journal of Fluency Disorders*, 20, 115-132.
- Denny, M., & Smith, A. (1992). Gradations in a pattern of neuromuscular activity associated with stuttering. *Journal of Speech and Hearing Research*, 35(6), 1216-1229.

- Didirková, I., Le Maguer, S., Hirsch, F., & Gbedahou, D. (2019). Articulatory behaviour during disfluencies in stuttered speech. In *Proceedings of ICPHS2019* (pp. 2991-2995).
- Didirková I., & Hirsch, F. (2020a) A two-case study of coarticulation in stuttered speech. An articulatory approach. *Clinical Linguistics & Phonetics*, 34(6), 517-535.
- Didirková I., & Hirsch, F. (2020b) An articulatory study of differences and similarities between stuttered disfluencies and non-pathological disfluencies. *Clinical Linguistics & Phonetics*, 35(3), 201-221.
- Fibiger, S. (1971). Stuttering explained as a physiological tremor. *Speech Transmission Lab Quarterly Progress and Status Report*, 2(3), 1-24.
- Guenther, F. H. (2016). *Neural control of speech*. Cambridge, MA: MIT Press.
- Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, 96(3), 280-301.
- Hagedorn, C., Proctor, M., & Goldstein, L. (2011). Automatic analysis of singleton and geminate consonant articulation using real-time magnetic resonance imaging. In *Proceedings of INTERSPEECH2011* (pp. 409–412).
- Hagedorn, C., Lammert, A., Bassily, M., Zu, Y., Sinha, U., Goldstein, L., & Narayanan, S. S. (2014). Characterizing post-glossectomy speech using real-time MRI. In *Proceedings of International Seminar on Speech Production, Cologne, Germany*. (pp. 170-173).
- Hagedorn, C., Proctor, M., Goldstein, L., Wilson, S. M., Miller, B., Gorno-Tempini, M. L., & Narayanan, S. (2017). Characterizing articulation in apraxic speech using real-time magnetic resonance imaging. *Journal of Speech, Language, and Hearing Research*, 60(4), 877-891.
- Harrington, J. (1987). Coarticulation and stuttering: an acoustic and palatographic study. In H. F. M. Peters & W. Hulstijn (Eds.). *Speech motor dynamics in stuttering* (pp. 381-392). Wien: Springer-Verlag.

- Heyde, C. J., Scobbie, J. M., Lickley, R., & Drake, E. K. (2016). How fluent is the fluent speech of people who stutter? A new approach to measuring kinematics with ultrasound. *Clinical Linguistics & Phonetics*, 30(3-5), 292-312.
- Hermes, A., Mücke, D., Thies, T., & Barbe, M. T. (2019). Coordination patterns in Essential Tremor patients with Deep Brain Stimulation: Syllables with low and high complexity. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 10(1), 6, 1-20.
- Howell, P. (2004). Assessment of some contemporary theories of stuttering that apply to spontaneous speech. *Contemporary Issues in Communication Sciences and Disorders*, 31, 123–141.
- Jacobs, A. E. (2016). *Phonological encoding of medial vowels in adults who stutter*. [Unpublished master's thesis]. Louisiana State University.
- Johnson, W., & Brown, S. F. (1935). Stuttering in relation to various speech sounds. *Quarterly Journal of Speech*, 21(4), 481-496.
- Kent, R. D. (2000). Research on speech motor control and its disorders: A review and prospective. *Journal of Communication Disorders*, 33(5), 391–428.
- Kim, Y. C., Proctor, M. I., Narayanan, S. S., & Nayak, K. S. (2012). Improved imaging of lingual articulation using real-time multislice MRI. *Journal of Magnetic Resonance Imaging*, 35(4), 943-948.
- Kleinow, J., & Smith, A. (2000). Influences of length and syntactic complexity on the speech motor stability of the fluent speech of adults who stutter. *Journal of Speech, Language, and Hearing Research*, 43(2), 548-559.
- Knoll, F., Bredies, K., Pock, T., & Stollberger, R. (2011). Second order total generalized variation (TGV) for MRI. *Magnetic Resonance in Medicine*, 65(2), 480–491.
- Lammert, A., Proctor, M., & Narayanan, S. (2010). Data-driven analysis of realtime vocal tract MRI using correlated image regions. In *Proceedings of InterSpeech2010* (pp. 1572–1575). Red Hook, NY: International Speech Communication Association.

- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159–174.
- Lee, Y. J., Goldstein, L., & Narayanan, S. S. (2015). Systematic variation in the articulation of the Korean liquid across prosodic positions. In *Proceedings of the International Congress on Phonetic Sciences, Glasgow*.
- Leha, A., Dickhut, S., Ponssen, D., Primassin, A., Korzeczek, A., Joseph, A. A., Paulus, W., Frahm, J., & Sommer, M. (2020). Iceberg or cut off—how adults who stutter articulate fluent-sounding utterances. *bioRxiv*.
- Lingala, S. G., Zhu, Y., Kim, Y. C., Toutios, A., Narayanan, S., & Nayak, K. S. (2015). High spatio-temporal resolution multi-slice real time MRI of speech using golden angle spiral imaging with constrained reconstruction, parallel imaging, and a novel upper airway coil. *Proc ISMRM 23rd Scientific Sessions*, 689.
- Lingala, S. G., Sutton, B. P., Miquel, M. E., & Nayak, K. S. (2016). Recommendations for real-time speech MRI. *Journal of Magnetic Resonance Imaging*, 43(1), 28-44.
- Löfqvist, A., & Gracco, V. L. (1999). Interarticulator programming in VCV sequences: Lip and tongue movements. *The Journal of the Acoustical Society of America*, 105(3), 1864-1876.
- Loucks, T. M. J., & De Nil, L. F. (2006a). Anomalous sensorimotor integration in adults who stutter: A tendon vibration study. *Neuroscience Letters*, 402, 195–200.
- Loucks, T. M., & De Nil, L. F. (2006b). Oral kinesthetic deficit in adults who stutter: A target-accuracy study. *Journal of Motor Behavior*, 38, 238–246.
- Max, L., Caruso, A. J., & Gracco, V. L. (2003). Kinematic analyses of speech, orofacial nonspeech, and finger movements in stuttering and nonstuttering adults. *Journal of Speech, Language, and Hearing Research*, 46(1), 215-232.
- Max, L., Guenther, F. H., Gracco, V. L., Ghosh, S. S., & Wallace, M. E. (2004). Unstable or insufficiently activated internal models and feedback-biased motor control as sources of dysfluency: A

theoretical model of stuttering. *Contemporary Issues in Communication Science and Disorders*, 31, 105-122.

Max, L., Kadri, M., Mitsuya, T., & Balasubramanian, V. (2019). Similar within-utterance loci of dysfluency in acquired neurogenic and persistent developmental stuttering. *Brain and Language*, 189(2019), 1-9.

McClean, M. D., Goldsmith, H., & Cerf, A. (1984). Lowerlip EMG and displacement during bilabial disfluencies in adult stutterers. *Journal of Speech and Hearing Research*, 27(3), 342-

McClean, M. D., Kroll, R. M., & Loftus, N. S. (1990). Kinematic analysis of lip closure in stutterers' fluent speech. *Journal of Speech, Language, and Hearing Research*, 33(4), 755-760.

McClean, M. D., & Tasko, S. M. (2004). Correlation of orofacial speeds with voice acoustic measures in the fluent speech of persons who stutter. *Experimental Brain Research*, 159, 310-318.

McClean, M. D., Tasko, S. M., & Runyan, C. M. (2004). Orofacial movements associated with fluent speech in persons who stutter. *Journal of Speech, Language, and Hearing Research*, 47(2), 294–303.

Nam, Hosung. (2007). A competitive, coupled oscillator model of moraic structure: Split-gesture dynamics focusing on positional asymmetry. In J. Cole & J. I. Hualde (Eds.) *Laboratory phonology 9* (pp. 483-506). Berlin & New York: Mouton de Gruyter.

Nam, H., Goldstein, L., & Saltzman, E. (2009). Self-organization of syllable structure: A coupled oscillator model. In *Approaches to Phonological Complexity* (pp. 297-328). De Gruyter Mouton.

Namasivayam, A. K., & van Lieshout, P. (2008). Investigating speech motor practice and learning in people who stutter. *Journal of Fluency Disorders*, 33(1), 32-51.

Namasivayam, A. K., & van Lieshout, P. (2011). Speech motor skill and stuttering. *Journal of Motor Behavior*, 43(6), 477-489.

- Niebergall, A., Zhang, S., Kunay, E., Keydana, G., Job, M., Uecker, M., & Frahm, J. (2013). Real-time MRI of speaking at a resolution of 33 ms: Undersampled radial FLASH with nonlinear inverse reconstruction. *Magnetic Resonance in Medicine*, 69(2), 477–485.
- Peters, H. M., Hulstijn, W., & van Lieshout, P. H. (2000). Recent developments in speech motor research into stuttering. *Folia Phoniatrica et Logopaedica*, 52(1-3), 103-119.
- Postma, A., & Kolk, H. (1993). The covert repair hypothesis: Prearticulatory repair processes in normal and stuttered disfluencies. *Journal of Speech, Language, and Hearing Research*, 36(3), 472-487.
- Proctor, M., Lammert, A., Katsamanis, A., Goldstein, L., Hagedorn, C., & Narayanan, S. (2011). Direct estimation of articulatory kinematics from real-time magnetic resonance image sequences. In *Proceedings of INTERSPEECH2011* (pp. 281-284).
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Ramanarayanan, V., Tilsen, S., Proctor, M., Töger, J., Goldstein, L., Nayak, K. S., & Narayanan, S. (2018). Analysis of speech production real-time MRI. *Computer Speech & Language*, 52, 1-22.
- Riley, G. (2009). *SSI-4 stuttering severity instrument fourth edition*. Austin, TX: Pro-Ed.
- Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, 1, 333–382.
- Smith, A. (1989). Neural drive to muscles in stuttering. *Journal of Speech and Hearing Research*, 32(2), 252-264.
- Smith, A., Luschei, E., Denny, M., Wood, J., Hirano, M., & Badylak, S. (1993). Spectral analyses of activity of laryngeal and orofacial muscles in stutterers. *Journal of Neurology, Neurosurgery & Psychiatry*, 56(12), 1303-1311.
- Smith, A., Denny, M., Shaffer, L. A., Kelly, E. M., & Hirano, M. (1996). Activity of intrinsic laryngeal muscles in fluent and disfluent speech. *Journal of Speech Language and Hearing Research*, 39(2), 329-348.



- Smith, A., & Kleinow, J. (2000). Kinematic correlates of speaking rate changes in stuttering and normally fluent adults. *Journal of Speech, Language, and Hearing Research*, 43(2), 521-536.
- Smith, A., Sadagopan, N., Walsh, B., & Weber-Fox, C. (2010). Increasing phonological complexity reveals heightened instability in inter-articulatory coordination in adults who stutter. *Journal of Fluency Disorders*, 35(1), 1-18.
- Teixeira, A., Martins, P., Oliveira, C., Ferreira, C., Silva, A., & Shosted, R. (2012, April). Real-time mri for portuguese. In *International Conference on Computational Processing of the Portuguese Language* (pp. 306-317). Springer, Berlin, Heidelberg.
- Tilsen, S., & Goldstein, L. (2012). Articulatory gestures are individually selected in production. *Journal of Phonetics*, 40(6), 764-779.
- Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*, 39(3), 1429-1443.
- Van Riper, C. (1982). *The nature of stuttering* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Wingate, M. E. (1988). *The structure of stuttering*. New York: Springer.
- Zimmermann, G. (1980a). Articulatory dynamics of fluent utterances of stutterers and nonstutterers. *Journal of Speech, Language, and Hearing Research*, 23(1), 95-107.
- Zimmermann, G. (1980b). Articulatory behaviors associated with stuttering. *Journal of Speech Language and Hearing Research*, 23(1), 108-121.
- Zimmermann, G. N., & Hanley, J. M. (1983). A cinefluorographic investigation of repeated fluent productions of stutterers in an adaptation procedure. *Journal of Speech Language and Hearing Research*, 26(1), 35.

## Supplementary material

Trials in this experiment were coded in the order of AA, AB, ..., AZ, BA, BB, ..., BW. Each trial was randomly assigned with a pseudoword.

**Table 1.** Instances of stuttering disfluencies.

	<b>Trial</b>	<b>Pseudoword</b>	<b>Disfluent consonant</b>	<b>Perceptual type of disfluency</b>
<b>1</b>	AA	/mæbfɛɪtɑɪdɔɪb/	m	block, prolongation, repetition
<b>2</b>	AC	/mæbfɑɪfɛɪb/	m	prolongation
<b>3</b>			f	block, prolongation, repetition
<b>4</b>	AD	/mæbfɛɪtɑɪdɔɪb/	m	block, prolongation
<b>5</b>	AE	/mæbfɛɪtɑɪdɔɪb/	m	block, prolongation
<b>6</b>	AJ	/mæbfɑɪfɛɪb/	m	block, prolongation
<b>7</b>	AK	/mæbtɪbɪbɪ/	m	block, prolongation, repetition
<b>8</b>	AL	/mæbfɛɪtɑɪdɔɪb/	m	block, prolongation
<b>9</b>			ʃ	block
<b>10</b>	AN	/mæbfɛɪtɑɪdɔɪb/	m	prolongation
<b>11</b>	AO	/mæbtɪbɪbɪ/	m	prolongation
<b>12</b>			t	block
<b>13</b>	AP	/mæbfɑɪfɛɪb/	m	prolongation
<b>14</b>			f	prolongation
<b>15</b>			ʃ	prolongation
<b>16</b>	AR	/mæbtɪbɪbɪ/	t	block
<b>17</b>	AS	/mæbtɪbɪbɪ/	t	block, repetition
<b>18</b>	AU	/mæbtɪbɪbɪ/	m	block, prolongation
<b>19</b>			t	block
<b>20</b>	AX	/mæbtɪbɪbɪ/	m	block, prolongation, repetition
<b>21</b>			t	block
<b>22</b>	AY	/mæbfɑɪfɛɪb/	m	prolongation, repetition
<b>23</b>			f	prolongation, repetition
<b>24</b>			ʃ	block, repetition
<b>25</b>	BA	/mæbfɛɪtɑɪdɔɪb/	m	block, prolongation
<b>26</b>			ʃ	block, prolongation, repetition
<b>27</b>			t	block, turned to swallow
<b>28</b>	BB	/mæbfɛɪtɑɪdɔɪb/	m	prolongation
<b>29</b>	BC	/mæbtɪbɪbɪ/	m	block, prolongation
<b>30</b>			t	block
<b>31</b>	BD	/mæbfɑɪfɛɪb/	m	block, prolongation, repetition
<b>32</b>			f	block, prolongation, repetition
<b>33</b>	BE	/mæbtɪbɪbɪ/	m	block, prolongation, repetition

<b>34</b>			t	block
<b>35</b>	BF	/mæbfaiʃeɪb/	m	prolongation
<b>36</b>	BG	/mæbtibibi/	m	prolongation
<b>37</b>			t	block
<b>38</b>	BQ	/mæbfaiʃeɪb/	m	prolongation
<b>39</b>			f	block, prolongation
<b>40</b>			ʃ	prolongation
<b>41</b>	BR	/mæbfɛɪtɑɪdɔɪb/	m	prolongation
<b>42</b>			ʃ	prolongation, repetition

**Table 2.** Perceptually fluent trials.

	<b>Trial</b>	<b>Pseudoword</b>
<b>1</b>	AF	/mæb/
<b>2</b>	AG	/mæb/
<b>3</b>	AM	/mæb/
<b>4</b>	AV	/mæb/
<b>5</b>	AW	/mæb/
<b>6</b>	BK	/mæb/
<b>7</b>	BO	/mæb/
<b>8</b>	BT	/mæb/
<b>9</b>	BV	/mæb/
<b>10</b>	AH	/mæbfaiɪb/
<b>11</b>	AI	/mæbfaiɪb/
<b>12</b>	AQ	/mæbfaiɪb/
<b>13</b>	AT	/mæbfaiɪb/
<b>14</b>	BH	/mæbfaiɪb/
<b>15</b>	BJ	/mæbfaiɪb/
<b>16</b>	BP	/mæbfaiɪb/
<b>17</b>	BS	/mæbfaiɪb/
<b>18</b>	BU	/mæbfaiɪb/
<b>19</b>	BW	/mæbfaiɪb/
<b>20</b>	AZ	/mæbfaiʃeɪb/
<b>21</b>	AB	/mæbfɛɪtɑɪdɔɪb/
<b>22</b>	BI	/mæbtibibi/

**Table 3.** Articulatory pattern and perceptual type of each instance of disfluency (excluding instance no. 27 in Table 1, which was identified as “block turned to swallow”)

Articulatory pattern	Perceptual type	Disfluent consonant	Pseudoword	Trial
1	block	ʃ	/mæbfɛɪtɑɪdɔɪb/	AL
1	block	t	/mæbtɪbɪbɪ/	AO
1	block	t	/mæbtɪbɪbɪ/	AX
1	block	t	/mæbtɪbɪbɪ/	BE
1	block	t	/mæbtɪbɪbɪ/	BG
1	block	t	/mæbtɪbɪbɪ/	AR
1	block	t	/mæbtɪbɪbɪ/	AU
1	block	t	/mæbtɪbɪbɪ/	BC
1	block, prolongation	f	/mæbfɑɪfɛɪb/	BQ
1	prolongation	f	/mæbfɑɪfɛɪb/	AP
1	block, prolongation	m	/mæbfɛɪtɑɪdɔɪb/	AL
1	block, prolongation	m	/mæbfɛɪtɑɪdɔɪb/	AE
1	prolongation	m	/mæbtɪbɪbɪ/	AO
1	prolongation	m	/mæbfɑɪfɛɪb/	BF
1	prolongation	m	/mæbfɛɪtɑɪdɔɪb/	BR
1	prolongation	ʃ	/mæbfɑɪfɛɪb/	BQ
2	block, prolongation	m	/mæbfɛɪtɑɪdɔɪb/	AD
2	block, prolongation	m	/mæbfɛɪtɑɪdɔɪb/	BA
2	block, prolongation	m	/mæbfɑɪfɛɪb/	AJ
2	block, prolongation	m	/mæbtɪbɪbɪ/	AU
2	block, prolongation	m	/mæbtɪbɪbɪ/	BC
2	prolongation	m	/mæbfɑɪfɛɪb/	AC
2	prolongation	m	/mæbfɑɪfɛɪb/	AP
2	prolongation	m	/mæbfɑɪfɛɪb/	BQ
2	prolongation	m	/mæbfɛɪtɑɪdɔɪb/	AN
2	prolongation	m	/mæbfɛɪtɑɪdɔɪb/	BB
2	prolongation	m	/mæbtɪbɪbɪ/	BG
2	prolongation	ʃ	/mæbfɑɪfɛɪb/	AP
3	block, prolongation, repetition	f	/mæbfɑɪfɛɪb/	BD
3	block, prolongation, repetition	ʃ	/mæbfɛɪtɑɪdɔɪb/	BA
3	block, prolongation, repetition	f	/mæbfɑɪfɛɪb/	AC
3	block, prolongation, repetition	m	/mæbfɛɪtɑɪdɔɪb/	AA
3	block, prolongation, repetition	m	/mæbtɪbɪbɪ/	AK
3	block, prolongation, repetition	m	/mæbtɪbɪbɪ/	AX
3	prolongation, repetition	f	/mæbfɑɪfɛɪb/	AY
3	prolongation, repetition	m	/mæbfɑɪfɛɪb/	AY
3	block, repetition	ʃ	/mæbfɑɪfɛɪb/	AY

1+2	block, repetition	t	/mæbtibibi/	AS
1+2	block, prolongation, repetition	m	/mæbfaiʃerb/	BD
1+2	prolongation, repetition	ʃ	/mæbfɛɪtɑɪdɔɪb/	BR
3+2	block, prolongation, repetition	m	/mæbtibibi/	BE