Journal of Fluency Disorders 88 (2026) 106224

Contents lists available at ScienceDirect

Journal of Fluency Disorders

journal homepage: www.elsevier.com/locate/jfludis

Temporal patterns in articulation underlying repetitions, prolongations and blocks

Yijing Lu [a,*], Louis Goldstein [b], Christina Hagedorn [c], Shrikanth Narayanan [b,d]

[a] Department of Linguistics, University of Potsdam, Germany
[b] Department of Linguistics, University of Southern California, Los Angeles, CA, United States
[c] Department of English, College of Staten Island, City University of New York, New York, NY, United States
[d] Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA, United States

ARTICLE INFO

Keywords:
Real-time magnetic resonance imaging
Articulatory patterns
State stuttering

ABSTRACT

Purpose: Repetitions, prolongations, and blocks, as canonical categories of stuttering disfluencies, have long been characterized on auditory-perceptual grounds, while the articulatory behaviors giving rise to them remain poorly understood. This study investigates the articulatory basis of these disfluencies, and assesses how the underlying articulatory patterns relate to perceptual disfluency types (i.e., repetitions, prolongations, blocks) and the phonological features of stuttered segments.

Method: Speech production of seven adults who stutter was recorded using real-time MRI. We extracted the time-varying constriction trajectories of the primary oral articulator for stuttered tokens of labial, alveolar, and velar consonants, as well as the constriction trajectories of the velum. K-medoids clustering with dynamic time warping distance was used to identify recurrent patterns.

Results: Two constriction patterns of the primary oral articulator were observed consistently across participants: fixation, characterized by an extended constriction period, and oscillation, characterized by repeated constriction-release cycles. Velic constriction trajectories did not show stuttering-related patterns but simply reflected the nasality of the stuttered segment. The mapping between articulatory constriction patterns and perceptual disfluency types was many-to-one and one-to-many. Crucially, this mapping was influenced by the manner of the stuttered segment.

Conclusion: State stuttering is characterized by two stable articulatory patterns: fixation and oscillation of the primary oral articulator. Their mappings to perceptual disfluency types are not in a simple one-to-one correspondence and also depend on the phonological task being executed. These findings point to a principled organization in the articulation of stuttering, enriching our understanding of the phenomenon beyond its perceptual surface.

1. Introduction

Despite a vast and longstanding literature on stuttering, one area that has received little attention yet deserves considerably more is the articulatory characteristics of stuttered speech. Stuttering has long been characterized and described based on observers' perceptions, i.e., what stuttering sounds and looks like when a person speaks. However, the articulatory behaviors underlying perceived

* Correspondence to: University of Potsdam, Campus Golm, Karl-Liebknecht-Str. 24/25, House 14, Room 3.18/19, Potsdam 14476, Germany.
E-mail address: yijing.lu@uni-potsdam.de (Y. Lu).

https://doi.org/10.1016/j.jfludis.2026.106224
Received 14 January 2026; Received in revised form 24 April 2026; Accepted 20 May 2026
Available online 23 May 2026
0094-730X/© 2026 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

stuttering events remain greatly understudied. This lack of research matters because the mapping between articulation and acoustics is complex, meaning that these behaviors cannot be sufficiently inferred from acoustic output or auditory perception alone. Studying articulatory behavior during stuttering provides a more direct window into the speech motor processes involved. Such knowledge could also be clinically useful as it may allow stuttering events to be characterized more precisely than by perceptual labels alone.

Though in recent years there has been a shift towards studying stuttering as a disorder of speech motor control (see a review in Packman & Attanasio, 2017), most articulatory or speech motor studies of stuttering have focused on the fluent speech of people who stutter rather than their disfluent speech *per se* (Table 1). This gap is in no small measure due to the technical difficulty of collecting and analyzing data on stuttered speech, but it is also to some extent a deliberate choice. Stuttered speech is often seen as a complex mix of core stuttering behaviors and the reactions thereto, with considerable variability recognized within and between individuals. In contrast, fluent speech is believed to be less affected by learned reactions or management, and it is further assumed that even perceptually fluent speech of people who stutter may be motorically atypical. Consequently, researchers have tended to compare the fluent speech of people who stutter (PWS) to the fluent speech of people who do not stutter (PWNS) in order to identify trait-level differences between the two groups, which are often considered causally related to stuttering. This broader view of stuttering, encompassing more than just moments of overt stuttering, is known as "trait stuttering" (Belyk et al., 2015). In contrast to it, the notion of "state stuttering" refers to the act of stuttering itself. Studies of state stuttering examine stuttered speech directly, often in comparison to the speaker's own fluent productions.

There are important reasons to spotlight the investigation of actual stuttered speech. First, it remains unclear whether PWS and PWNS exhibit consistent articulatory differences during fluent speech. Contradictory findings have been reported regarding whether they truly differ in movement duration, displacement, velocity, and sequencing (Table 1). The most consistent finding is that PWS exhibit higher speech movement variability than PWNS, but even this finding is not replicated in Smith and Kleinow (2000). Second, even if a stable trait-level difference is identified, it does not necessarily explain the specific manifestations of state stuttering. For example, establishing that PWS have less stable speech motor control (as evidenced by higher variability) does not automatically explain why they exhibit repetitions, prolongations, and blocks in speech, and why these disfluencies predominantly occur in syllable-initial positions. To explain these specific characteristics of stuttering, it is important to understand the nature of the speech production behaviors during stuttering moments (e.g., whether they are disruptions, repairs, or stalling tactics). Third, although it is reasonable to consider behaviors during stuttering moments as resulting from both core disruption and learned reactions, this is not a sufficient reason to avoid such investigations. Such ambiguity can be reduced, at least in part, through theoretically informed analysis and explicit analytic principles, an approach we take in Section 1.1.

Hence, we argue that articulatory characteristics of stuttered speech represent a fundamental phenomenon of stuttering that merits greater attention from both scientists and clinicians. Trait and state perspectives should ideally be brought together to promote a more complete understanding of stuttering. Notably, some brain studies have started to adopt this broader perspective (e.g., Connally et al., 2018; Belyk et al., 2015). The same approach could also be extended to the articulatory domain. The present study, however, does not attempt to address trait and state stuttering equally within a single design, but instead focuses specifically on state stuttering, thereby helping to address its relative underrepresentation in articulatory research.

1.1. The infrastructure for articulatory studies of stuttered speech

Carrying out articulatory studies on stuttered speech poses significant challenges and requires advancements in laboratory instrumentation, experimental design, and data analysis methods. The first two aspects have begun to be addressed by recent methodological developments. Real-time structural MRI of the vocal tract has proven to be a powerful tool for studying speech production (Kumar et al., 2024; Lim et al., 2024; Narayanan et al., 2004; Niebergall et al., 2013; Uecker et al., 2010) and has recently been applied to people who stutter (Lu et al., 2022; Ponssen et al., 2024; Wiltshire et al., 2021). In addition, Lu, Goldstein, et al. (2025) introduced variegated speech tasks designed to increase the likelihood of eliciting stuttering in the scanner. Together, these

Table 1
Summary of findings from previous studies on fluent speech of people who stutter. Unless otherwise noted, studies were conducted with adults.

Compared to fluent speech of PWNS:	Yes	No
Longer movement durations and/or longer temporal intervals between articulatory events	Max et al. (2003); Sasisekaran (2013); Smith et al. (2010); Wiltshire et al. (2021); Zimmermann (1980a)	McClean, Tasko, and Runyan (2004); McClean and Tasko (2004) Smith et al. (2012) [children]; Usler et al. (2017) [children]
Reduced articulator displacements and velocities	Zimmermann (1980a)	Zimmermann and Hanley (1983)
Greater movement variability across repeated utterances	Frisch et al. (2016); Jackson et al. (2016); Kleinow and Smith (2000); MacPherson and Smith (2013); Sasisekaran (2013); Smith et al. (2010); Wiltshire et al. (2021)	Smith and Kleinow (2000)
Different sequencing of upper lip, lower lip, jaw movement onset and peak velocity	Caruso et al. (1988)	De Nil (1995); McClean et al. (1990)

developments make it increasingly possible to acquire articulatory data from stuttered speech. Regarding data analysis and interpretation, here we argue that the framework of Articulatory Phonology and Task Dynamics (APTD) (Browman & Goldstein, 1992; Saltzman & Munhall, 1989) offers unique advantages for analyzing articulatory data on stuttered speech.

The APTD framework views linguistic units and motor speech representations as essentially one and the same. The units that subserve the dual roles are articulatory gestures, which are linguistically defined, goal-directed motor tasks for speech articulators to control the constriction degree (broadly understood as aperture) in various regions of the vocal tract. The goals of these tasks are discrete, context-independent, and time-invariant, whereas the movement of individual articulators that coordinate to achieve the goals is continuous, highly variable, and context-dependent. For example, to produce /p/, the goal is always to reduce the lip aperture to 0 (or a slightly negative value, taking into account lip compression), but the specific displacement of the upper lip, lower lip, or jaw varies with speaker, phonological context, and external perturbation. In typically fluent speech, the spatiotemporal unfolding of a gesture has a stable structure, which involves movement toward a constriction target and subsequent release from that target (Browman & Goldstein, 1990; Saltzman & Munhall, 1989).

Guided by this framework, we adopt task-level measurements, which are the time course of constriction degree at specific vocal tract locations, as determined by the phonological task. For example, we focus on the change in lip aperture if the speaker is producing or attempting to produce /p/, or the constriction degree between the tongue tip and the alveolar ridge if it is /t/. This approach differs from methods that extract isolated kinematic measurements, such as the displacement or velocity of individual articulators across the utterance, without linking them to specific phonological tasks. Since task-level measurements operate at a higher level of abstraction, by stripping away lower-level variances, they offer greater potential to reveal stable and consistent features of state stuttering. Accordingly, one interpretive principle we adopt is that if common task-level patterns during stuttering moments can be identified across individuals, these patterns are more plausibly linked to fundamental aspects of stuttering rather than learned reactions, as learned reactions would generally be expected to show greater individual idiosyncrasy.

1.2. Prior research on stuttered speech articulation

Repetitions, prolongations, and blocks are the canonical categories used to describe stuttering disfluencies. Although the labels themselves may imply particular forms of behavior, in practice, they are assigned based on perceptual characteristics. Direct evidence about the underlying articulation remains limited. A central question, therefore, is what articulatory patterns are associated with these perceptually defined disfluency types.

Only a small number of studies have examined the movements of speech articulators during repetitions, prolongations, and blocks. Early work suggested that stuttering involved repetitive movements and static posturing in the articulator engaged in the production of the stuttered segment (Zimmermann, 1980b). Later studies examined more directly how perceptual disfluency types relate to articulatory patterns. Based on the global movements of multiple articulators, Didirková et al. (2019) identified four patterns: reiterative movements, global maintenance of postures, anarchical movements, and combinations thereof. Reiterative movements occurred primarily in repetitions, whereas the other patterns occurred across repetitions, prolongations, and blocks. Focusing on the constriction kinematics of the primary articulator for the stuttered segment, Lu et al. (2022) identified three patterns: fluent-like constriction formation and release but with increased constriction degree, excessively prolonged constriction, and successive repetitions of constriction and release. Block and prolongation were associated with the first two patterns, whereas repetition was associated with the third pattern or with combinations of these different articulatory patterns.

Although these studies offer initial insight into the articulatory patterns associated with perceptual disfluency types, they also share several limitations. First, the small numbers of speakers and tokens limit the generalizability of existing findings. Second, articulatory patterns were identified qualitatively, leaving room for observer bias. Third, they have largely overlooked how phonological properties of the stuttered phoneme may shape the relation between articulatory pattern and perceptual disfluency type.

1.3. Current study

The present study investigates the articulatory patterns underlying repetitions, prolongations, and blocks by focusing on the constriction behavior of the primary oral articulators involved in the intended phoneme, as well as the velum, within the APTD framework. Compared to previous work, the study examines a larger group of speakers, samples a substantial number of stuttering tokens from naturalistic speech tasks, and uses quantitative analysis of constriction kinematics together with unsupervised clustering to identify recurrent patterns in the data. With this combination of features, we ask not only what articulatory patterns occur during stuttering, but also whether these patterns are shared across speakers and how they relate to perceptual disfluency types and the phonological properties of the stuttered phoneme. Specifically, the study addresses the following questions:

(1) During repetitions, prolongations, and blocks, what constriction patterns are observed in the primary oral articulator and the velum for the intended phoneme?
(2) Are these patterns shared across individuals who stutter?
(3) How do constriction patterns relate to repetitions, prolongations, and blocks, and does this relationship depend on the phonological features of the stuttered phoneme?

2. Method

2.1. Data collection

2.1.1. Participants

Seven adults who stutter (AWS1-AWS7, six men, one woman, mean age = 37.57 years, SD = 14.34 years) took part in the study. All participants were native speakers of English and reported no speech-language hearing disorders apart from persistent developmental stuttering. All participants reported the onset of stuttering in early childhood (i.e., before 10 years of age), and none had received speech therapy within the 12 months prior to participation. Stuttering severity was assessed using the Stuttering Severity Instrument Fourth Edition (SSI-4) (Riley, 2009). Primary SSI-4 scoring was carried out by the first author, a postdoctoral researcher specializing in stuttering research and trained in SSI-4 scoring (rater 1), who evaluated all SSI-4 components. To provide an independent reliability check, the third author, a licensed speech-language pathologist (rater 2), scored stuttering frequency and duration from the audio recordings obtained during the SSI administration. Participant-level SSI-4 scores are summarized in Table 2. The two raters were highly similar in their scoring of the speech-based SSI-4 components. Frequency scores differed by at most 1 point across all participants. Duration scores were identical for six of the seven participants, with the only larger discrepancy observed for AWS6. The study protocol (#HS-21–00712) was approved by the Institutional Review Board of the University of Southern California, and all participants gave informed consent before participating in the study.

2.1.2. Procedure

Before the MRI scan session, participants were given instructions about the study and the speech tasks they would perform inside the scanner. After providing written informed consent, participants took part in an MRI scan session, in which dynamic imaging of their speech production was acquired. The entire scan session, including breaks, lasted approximately 60–75 min.

During scanning, visual stimuli and task instructions were presented to participants via a projector-mirror setup. Oral prompts were delivered through an intercom. Participants wore a pair of MRI-compatible headset plugs (Magna coil TIM headset for Siemens) throughout the scan session, enabling them to hear prompts even while the scanner was running.

2.1.3. Experimental tasks

While lying supine inside the scanner, participants performed three speech production tasks designed to elicit stuttering, incorporating factors known to increase disfluency such as conversational exchange and time pressure (Lu, Goldstein, et al., 2025).

2.1.3.1. Passage reading. Participants read aloud excerpts from three phonetically balanced passages: The Rainbow Passage (Fairbanks, 1960), The Grandfather Passage (Darley et al., 1975), and The North Wind and the Sun (International Phonetic Association., 1999). These passages were read twice — once at the beginning and once at the end of the MRI session — with the two other tasks placed in between to reduce the adaptation effect.

2.1.3.2. Short interview. Participants responded to open-ended questions posed by the experimenter via intercom. Questions covered five topics: weekend and summer activities, music preferences, communication preferences, entertainment preferences, and diet/eating habits. Each topic involved 2–5 questions.

2.1.3.3. Picture description. Participants first saw a single picture and had 30 s to describe it. Then a pair of pictures appeared side by side, with the left one identical to the single picture, and the right one differing from it in two locations. Participants had 30 s to describe the differences between them, without being told the number of differences or that the left picture repeated the previous picture. In both phases, they were encouraged to speak for the full 30 s. In total, 20 sets of pictures were presented to the participants. Stimuli were adapted from the DiapixUK task (Baker & Hazan, 2011), all within a "beach" theme and containing overlapping visual elements (e.g., objects, colors, events), which promoted reuse of the same lexical items across descriptions.

The interview and picture description tasks were organized into blocks (one interview topic per interview block; four picture sets per picture block) and interleaved throughout the scan session, with passage reading at the beginning and end. This alternating task design has been demonstrated to help maintain stable elicitation of stuttering throughout the session (Lu, Goldstein, et al., 2025).

All participants completed the three tasks, except for AWS7, who ended the session early due to fatigue. AWS7 completed the initial reading of The Rainbow Passage and The Grandfather Passage, two sets of interview questions, and eight sets of picture descriptions.

2.1.4. MRI acquisition, reconstruction, and audio acquisition

MRI data were acquired using a whole-body 0.55 T system (prototype MAGNETOM Aera, Siemens Healthineers, Erlangen, Germany) equipped with high-performance shielded gradients (45 mT/m amplitude, 200 T/m/s slew rate). Signal reception was achieved using a custom 8-channel upper airway receiver coil array, with four elements on each side of the participant's cheek. Simultaneously with the MRI acquisition, acoustic speech signals were captured inside the scanner using a noise-canceling fiber-optic microphone (OptoAcoustics FOMRI-III). Scanner noise in the speech recordings was further reduced using STFT-based Wiener filtering.

RT-MRI acquisition was conducted using a 13-interleaf spiral-out balanced steady-state free precession (bSSFP) sequence. The imaging parameters were as follows: repetition time = 5.03 ms, echo time = 0.7694 ms, field-of view (FOV) = 240 × 240 mm^2, slice thickness = 6 mm, spatial resolution = 2.3 × 2.3 mm^2 (104 × 104 pixels), flip angle = 35 °. Imaging was performed in the mid-sagittal

Table 2
Participant information and SSI-4 scores. Scores for stuttering frequency, duration, physical concomitants, and total SSI-4 score from the first rater (R1) are shown for each participant. For the second rater (R2), only frequency and duration were scored.

Participant ID	Age	Gender	Frequency (R1)	Duration (R1)	Physical concomitants (R1)	Total SSI-4 (R1)	Frequency (R2)	Duration (R2)
AWS1	47	M	13	8	7	28 (moderate)	12	8
AWS2	30	M	13	10	9	32 (severe)	14	10
AWS3	32	M	12	10	5	27 (moderate)	13	10
AWS4	26	M	17	10	8	35 (severe)	16	10
AWS5	19	M	9	6	0	15 (very mild)	10	6
AWS6	53	M	12	8	3	23 (mild)	12	12
AWS7	56	W	18	14	12	44 (very severe)	18	14

plane, prescribed using a real-time interactive imaging platform (RT-Hawk, Heart Vista, Los Altos, CA). The acquired images were reconstructed with a temporal resolution of 10.06 ms per frame (99.4 frames/s).[1]

2.2. Data analysis

2.2.1. Speech transcription and disfluency annotation

Audio recordings were transcribed using Whisper run locally, then manually checked and corrected. Transcription-audio alignment was initially performed using Montreal Forced Alignment (McAuliffe et al., 2017) and also manually reviewed and corrected. Stuttering-like disfluencies were identified and categorized as repetition, prolongation, block, or a combination of any of these three types, primarily based on auditory-perceptual judgment, by the first author (rater 1) and the third author (rater 2). The operational definitions for the perceptual categories of repetition, prolongation, and block were as follows. Repetitions were audible repetitions of a sound or syllable. Prolongations were audible lengthening of a sound. Blocks were cessation of sound; when the rater was not sure whether it was a block or a normal pause, the RT-MRI video was consulted to see if there was a clear speech gesture being held for an upcoming segment, which should not be the case in normal pauses. Inter-rater reliability for the perceptual ratings of disfluencies was assessed using Cohen's kappa (κ), with 95% confidence intervals.

Because the aim of our study was to characterize articulatory patterns associated with the canonical disfluency categories, the analysis focused on tokens in which a single disfluency type could be identified clearly. More complex or overlapping events were excluded because they combine multiple phenomena and would make it difficult to determine which articulatory pattern corresponds to which component of the event.

Phonemes were annotated as stuttered if they were perceived as repeated or prolonged. In the case of blocks, the stuttered phoneme was marked as the one following the block, subject to visual confirmation that the requisite supralaryngeal articulatory gestures specific to that phoneme were present during the block. The present analysis focused on tokens in which the stuttered phoneme was a consonant, since consonants accounted for the large majority of stuttering events in the dataset. Disfluency annotations and word-level segmentation were performed using Praat (Boersma & Weenink, 2023). The resulting time stamps were used to trim RT-MRI videos into short clips that contain the stuttered word and 200 ms before and after.

2.2.2. Region-of-interest (ROI) analysis

Region-of-interest (ROI) analysis was implemented using VocalTract ROI Toolbox (Blaylock, 2021) to extract articulatory constriction information from RT-MRI videos. This study is focused on three types of articulatory gestures that have distinct constriction locations in the oral cavity: labial gestures (involved in the production of bilabial or labiodental consonants /p, b, m, w, f, v/), tongue tip gestures (involved in the production of alveolar consonants /t, d, n, s, z/), and tongue dorsum gestures (involved in the production of velar consonants /k, g/). Movements of the velum, which control the oral-nasal process, were also examined. Hence, four pseudo-circular regions of interest (ROI) were respectively placed at the locations of labial, alveolar, and velar constrictions, and at the location of velopharyngeal closure, as determined by the procedure described below. The ROIs were named after the active articulators involved: LIP, TT, TD, VEL.

For each participant, standard deviation (SD) images were derived from selected subsets of RT-MRI videos to guide ROI placement. For each ROI, SD images were computed from a small set of videos (typically around 10) drawn from the corresponding consonant category: videos containing stuttered labial consonants were used to define the LIP ROI, videos containing stuttered alveolar consonants were used to define the TT ROI, and videos containing stuttered velar consonants were used to define the TD ROI. The VEL ROI was determined from SD images of videos showing clear velum raising and lowering, including roughly equal numbers of oral and nasal consonants with the same place of articulation. Specifically, the center of each ROI was placed inside the brightest area in the SD image within the anatomically appropriate constriction region, so that the ROI captured the location showing the greatest movement during constriction formation and release: the LIP ROI was positioned between the lips, the TT ROI just below the alveolar ridge, the TD ROI below the velum, and the VEL ROI at the velopharyngeal port. The LIP, TT, and TD ROIs were defined with a radius of 4 pixels and covered 49 pixels each, to ensure they were large enough to capture both the formation and release of constrictions, and to

[1] The data of AWS7 was acquired with a slightly lower temporal resolution (99.1 frames/s).

accommodate occasional head movement or slight variation in the exact constriction location across tokens. The VEL ROI was defined with a radius of 3 pixels (29 pixels total) because the velopharyngeal port is a smaller region.

Fig. 1 shows four ROIs defined on the SD images of one participant, respectively capturing the formation and release of labial, alveolar, velar constrictions, as well as the closing and opening of the velopharyngeal port. Time series of mean pixel intensity in the LIP, TT, TD ROIs index changes in labial/alveolar/velar constriction degree. When a constriction is formed, more articulator tissue enters the corresponding ROI, resulting in an increase in mean pixel intensity; when a constriction is released, the articulator leaves the ROI, resulting in a decrease. Similarly, mean pixel intensity within the VEL ROI indexes the degree of velic constriction, with higher values corresponding to velopharyngeal closure and lower values corresponding to opening.

Fig. 2 shows an example of time series of mean pixel intensity within the LIP and VEL ROIs during repetitions of /m(ə)/ in the word "Manchester," excerpted from the utterance "to Manchester" [tuː ˈmanʧɛstə], beginning at the vowel /uː/ in "to" and ending just after the affricate /ʧ/ in "Manchester." The drop in the VEL time series from time point 1 to time point 2 corresponds to velum lowering. The first rise in the LIP time series from time point 2 to time point 3 reflects the closing movement of lips. From time point 3 onward, the fluctuations in the LIP trajectory are due to the oscillation of the lower lip and jaw, but the lips only separated slightly during this period (as shown at time points 4 and 6). The full opening of the lips occurs at time point 7. Meanwhile, the velum stayed lowered from the silence after /uː/, throughout the repetitions of /m(ə)/ and the following /man/, until the onset of /ʧ/.

Given that the total scanning time for this study was rather long (~1 h), there was a chance that participants moved their heads between task scans. To mitigate potential confounds in ROI analysis arising from head movements, the mean image of each video was computed and compared to the others. The videos were then organized into batches such that in each batch the head position in the mean images is maintained with relative consistency. Then, ROIs were defined within each batch using the method detailed above. Additionally, if occasional fluctuation in the overall image brightness (rather than articulator movements) caused changes in the time series, these time series were excluded from further analysis. The VocalTract ROI Toolbox also normalizes the pixel intensity in each frame, further reducing the impact of potential confounds on ROI analysis.

2.2.3. *Pixel intensity time series preprocessing*

Wavelet decomposition and soft thresholding ('db4' wavelet, decomposition level = 5, threshold = 0.1) were used to smooth pixel intensity time series from each ROI and reduce noise in the original data. We were interested in the entire time course of primary oral articulator movement during stuttering, from the very first attempt to form a constriction to its final release. Hence, on the smoothed pixel intensity time series from the LIP/TT/TD ROI, we first manually labeled the approximate start and end of the constriction gesture corresponding to the stuttered phoneme. These labels, which served as anchors for subsequent gestural landmark delimitation, were placed at time points before and after the acoustics of the stuttered consonant where the relevant constriction was relatively open, and the corresponding active articulator showed relatively little movement. In this step, if the oral gesture of the stuttered phoneme could

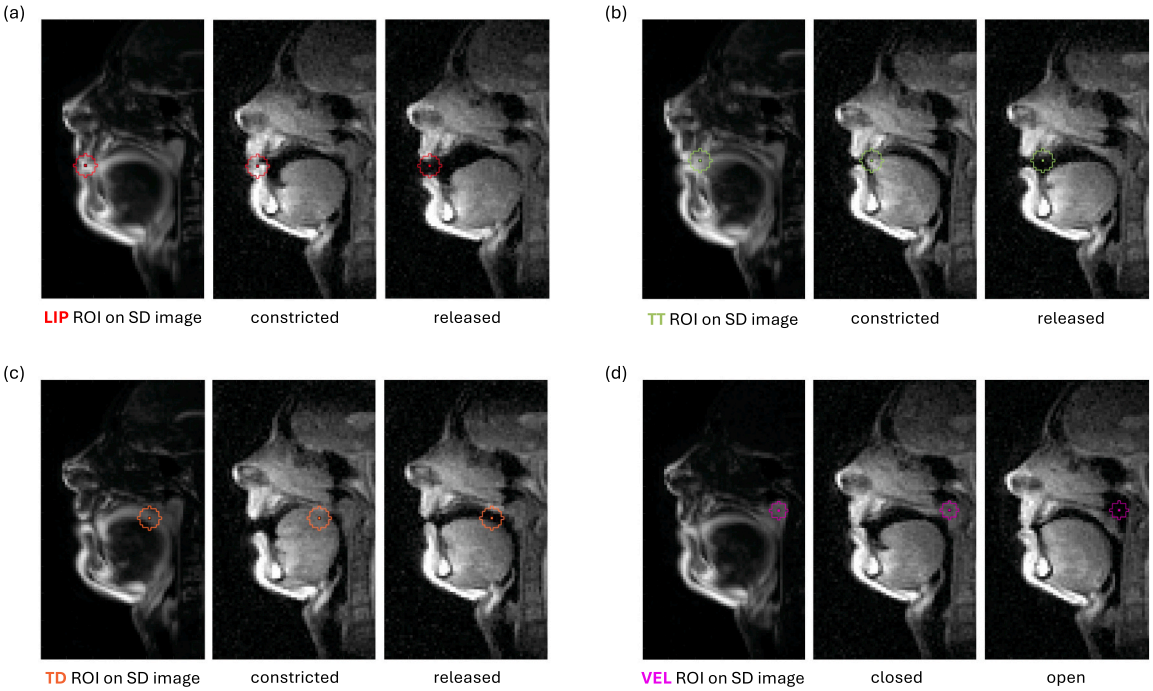

Fig. 1. Cropped images from AWS1 showing ROIs for (a) LIP, (b) TT, (c) TD, and (d) VEL. In each panel, from left to right: ROI on the standard deviation image, an example with an active constriction (mostly bright pixels within the ROI), and an example with a released/open state (mostly dark pixels within the ROI).

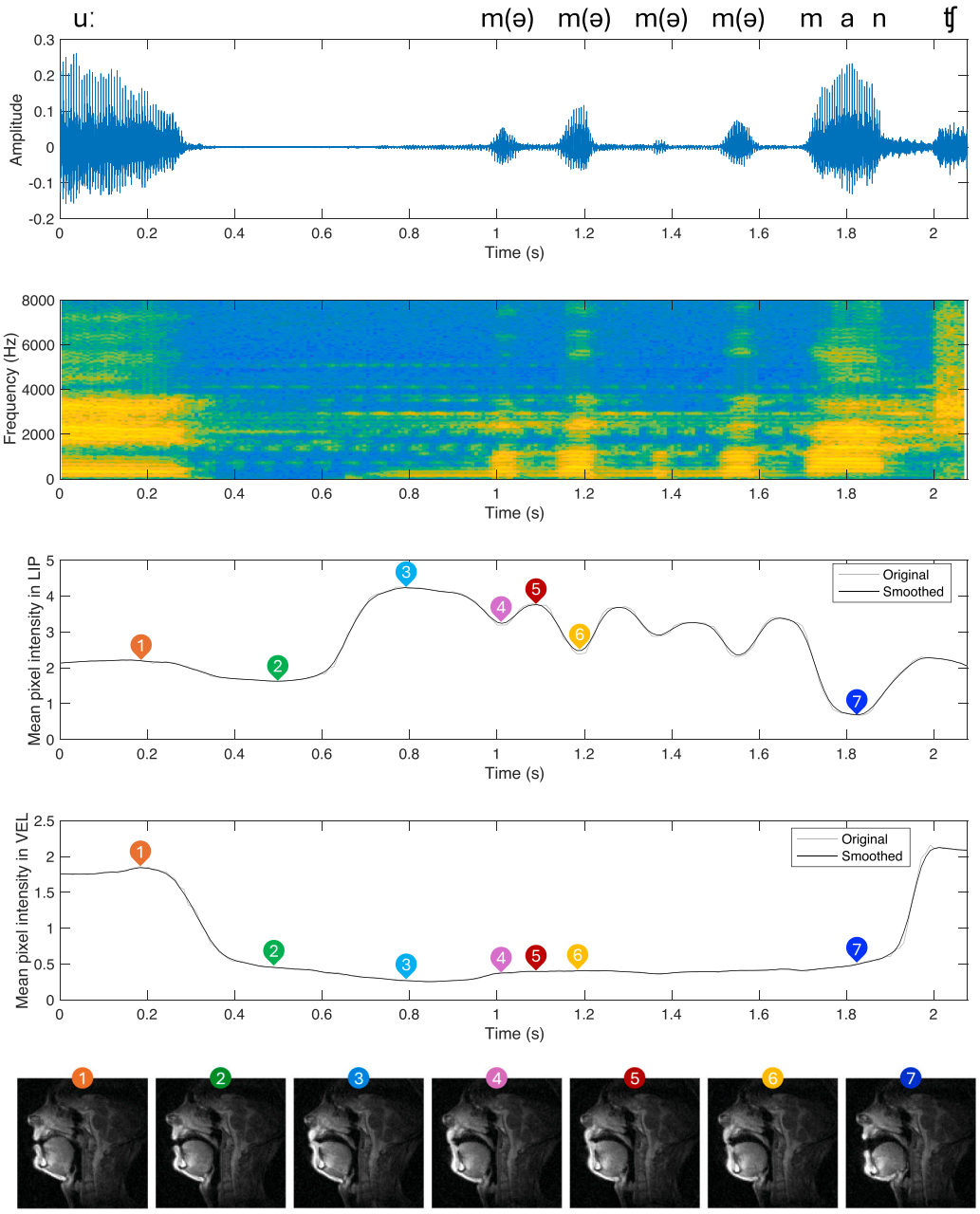

Fig. 2. Waveform, spectrogram, and time series of mean pixel intensity within the LIP and VEL ROIs during repetitions of /m(ə)/ in the utterance "to Manchester" produced by AWS1. The numbered MRI frames in the bottom row correspond to the colored time points on the time series.

not be visually distinguished from that of the previous phoneme (e.g., "is sssitting", "is n-n-n-n-no one", "of fffish", "to wwwalk"), the time series was excluded from further analysis. Then, within the search range bounded by the start and end labels, a gestural delimiting algorithm was used to automatically delimit four landmarks in the temporal structure of a gesture based on the velocity profile, which was approximated by the absolute change of pixel intensity between adjacent samples. Specifically, the four gestural temporal landmarks are: gesture onset (GONS), gesture offset (GOFFS), the onset of constriction target (TONS), and the offset of constriction target (TOFFS). From prior research (Lu et al., 2022), gestures during stuttered speech may show either temporally extended constriction intervals or repeated partial constrictions and releases (Fig. 3). The delimiting procedure was designed to accommodate both types of temporal pattern when identifying gestural landmarks. The algorithm identified the first and last constriction maxima within the search interval and used the associated velocity profile before and after these maxima to delimit GONS, TONS, TOFFS, and GOFFS automatically. Landmark locations were determined by applying 20% threshold criteria to the relevant portions of the velocity profile. The full step-by-step algorithm is provided in the supplementary material (Text S1). Fig. 3 shows the gestural temporal

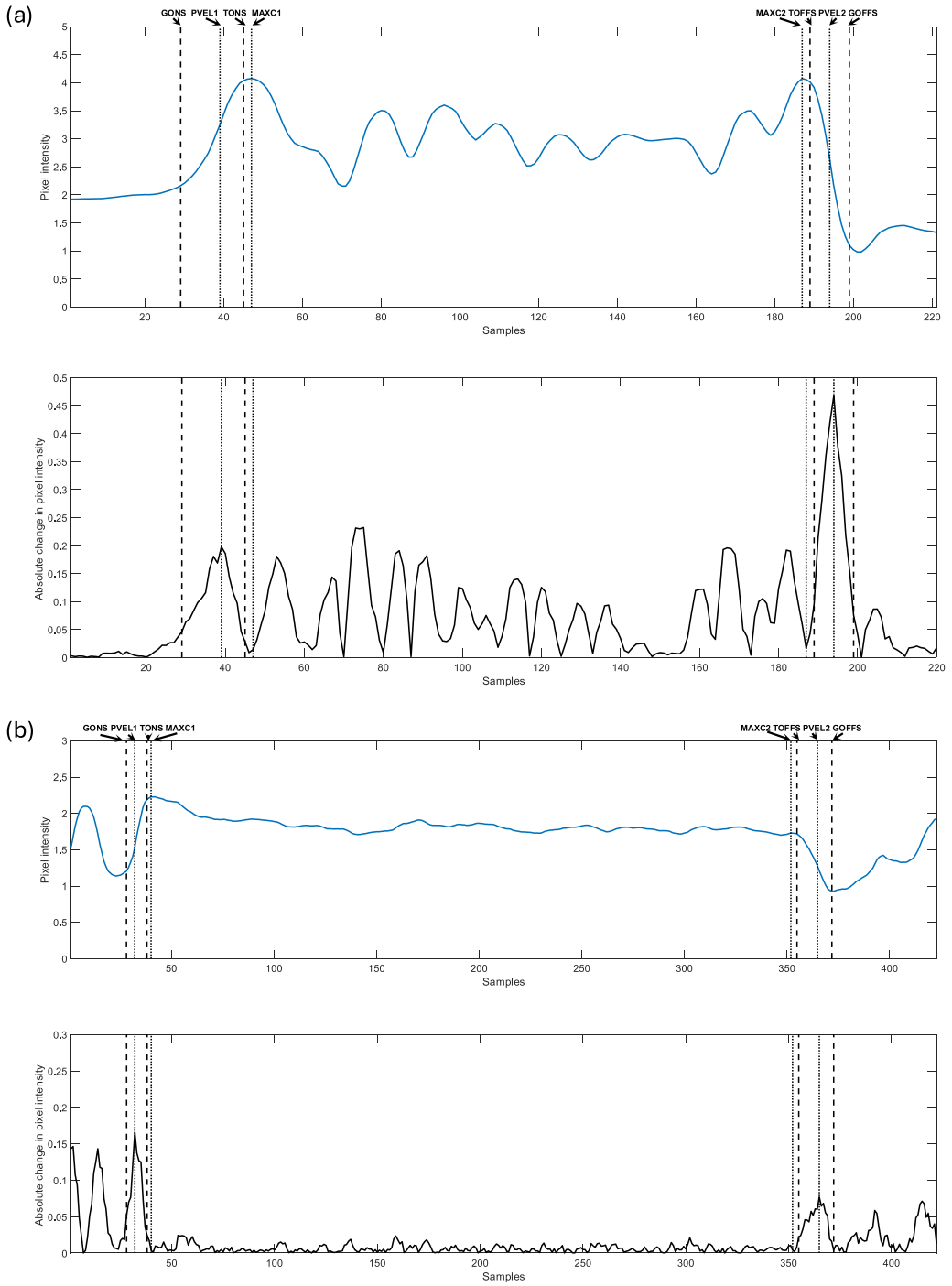

Fig. 3. (a) The upper panel is the pixel intensity time series showing the labial gesture during the repetition of /b/ in the word "because" produced by AWS1; the lower panel is the time series of the absolute change in pixel intensity; four dashed lines are GONS, TONS, TOFFS, GOFFS in sequence; four dotted lines are PVEL1, MAXC1, MAXC2, PVEL2 in sequence. (b) Labial gesture during a block of /b/ in the word "beyond" produced by AWS4, following the same format as (a).

landmarks defined on two instances of stutters, together with four intermediate reference points used by the delimiting procedure: the peak velocities associated with the formation- and release-side portions of the gesture (PVEL1 and PVEL2) and the first and last constriction maxima within the search interval (MAXC1 and MAXC2).

With the gestural temporal landmarks defined, pixel intensity time series were trimmed from GONS to GOFFS. Subsequently, each time series was normalized using min-max normalization, making the minimum value 0 and the maximum value 1. In this scale, 1

corresponds to maximum constriction and 0 to maximum opening within the given interval.

For time series from the VEL ROI, we did not perform separate gestural delimitation, since velum raising and lowering gestures are not determined solely by the stuttered phoneme but can span across sequences of oral or nasal segments, or even silences (as shown in Fig. 2). We simply trimmed the VEL time series according to the GONS and GOFFS of the stuttered gestures identified in the LIP, TT, or TD time series. Since these trimmed signals were not bounded by their own local extrema, min-max normalization could not be performed on individual tokens. Instead, min-max normalization was carried out across all VEL time series from the same participant, with 1 corresponding to maximum velic closure and 0 to maximum velic opening for that participant.

2.2.4. Clustering analysis

To identify recurrent patterns of articulatory gestures during stuttered speech without imposing *a priori* categories, unsupervised clustering analysis was performed. Given that the task was to cluster time series data, using Euclidean distance as the similarity metric could lead to unintuitive results, as Euclidean distance is not sensitive to the temporal relation between samples inside time series. Dynamic time warping (DTW) distance, by contrast, offers a more appropriate assessment of similarity by accounting for temporal alignment. Pairwise DTW distances were computed between all time series to form a distance matrix, which was used as the input to the clustering algorithm.

A further consideration is that our time series varied in length due to the different durations of each stuttering instance. Conventional K-means clustering operates on fixed-length feature vectors and relies on Euclidean distance, and therefore does not naturally accommodate variable-length time series or non-Euclidean distance measures such as DTW. For these reasons, we instead applied k-medoids clustering using Partitioning Around Medoids (PAM) style algorithm (Kaufman & Rousseeuw, 1990). In this approach, each cluster is represented by one observed time series from the dataset (the medoid). Each time series was assigned to the nearest medoid based on the precomputed DTW distance matrix, and medoid selections were iteratively updated to reduce the total within-cluster DTW distance.

The optimal number of clusters (k) was determined using the silhouette method (Rousseeuw, 1987). This method measures how similar an observation is to its own cluster compared to other clusters, expressed as its silhouette width. Higher silhouette widths indicate better-separated clusters. Silhouette widths were computed from the same DTW distance matrix. To ensure stability of the solution, for each candidate k, the clustering procedure was repeated 20 times. Within each run, the PAM procedure was initialized 5 times with different random initial medoid selections, and the solution with the smallest total within-cluster DTW distance was retained. Silhouette widths were then computed for the retained solution, and the mean and standard deviation of the average silhouette width across runs were calculated. The value of k that maximized the mean silhouette width was selected as the optimal number of clusters.

For each instance of stutter, we extracted the time series of the primary oral articulator, which are either LIP, TT, or TD time series. These time series were included in the same clustering experiment because they were trimmed using the same gestural delimitation process, which yields signals that consistently rise during constriction formation and fall during release. We also extracted the VEL time series for each stutter, which were clustered in a separate analysis because these trajectories vary in direction: they decrease when the velum lowers for nasal sounds and increase when the velum raises for oral sounds. This polarity difference motivated clustering the VEL signals separately.

To assess the extent to which clustering results are aligned with visually identifiable patterns, we conducted a manual evaluation on a random subset of 200 oral articulator time series. The representative shape of each resulting cluster was first visually inspected. Based on these visual patterns, each sampled time series was manually assigned to the cluster it most closely resembled by the first author. These manual assignments were then compared with the algorithmic cluster labels to assess agreement between clustering output and human judgment.

2.2.5. Statistical analysis

To examine whether and how the constriction patterns of stuttering relate to the perceptual types of disfluency and the phonological features of the stuttered phonemes, logistic mixed-effects modeling was used. The phonological features considered were place (labial, alveolar, velar) and manner. Manner was treated as a two-way distinction between stops (plosives and nasals, which involve a complete oral constriction) and continuants (fricatives and approximants, which involve a partially constricted oral gesture). These two features were chosen because they reflect the primary dimensions along which constriction gestures differ: constriction location and constriction degree.

Because place and manner are not fully crossed in our dataset (there were no velar continuants), including both in a single additive model would make their effects difficult to separate and interpret; therefore, they were examined in two separate models. In the first model, the constriction pattern obtained from clustering was predicted from perceptual disfluency type (repetition, prolongation, block) and place of articulation, including their interaction. In the second model, constriction pattern was predicted from perceptual disfluency type and manner of articulation, again including their interaction. Subject was included as a random intercept in both models to account for between-speaker variability. Reference levels for categorical predictors were left at their default settings in the model (perceptual disfluency type: blocks; place: alveolars; manner: continuants). Statistical significance was evaluated at $\alpha = .05$.

3. Results

3.1. Articulatory constriction patterns during stuttering

3.1.1. Lip, tongue tip, tongue dorsum movement

A total of 726 stuttering tokens (31.96% from passage reading, 21.63% from interview; 46.42% from picture description) were analyzed in our study, yielding 726 constriction time series of the primary oral articulator for the first clustering analysis. Specific information is shown in Table 3.

Fig. 4 shows the result of clustering the time series of the primary oral articulator (i.e., LIP, TT, or TD time series). The silhouette analysis indicated that the optimal number of clusters was k = 2 (Fig. 4a). At k = 2, the mean silhouette width is 0.485, suggesting a reasonable separation between the two clusters. The mean time series within each cluster (Fig. 4b) revealed two fundamental patterns in the movement of the primary oral articulator during stuttering. Cluster 1 (n = 552) contained trajectories characterized by a relatively flat plateau at high pixel intensity, which indicates a period of sustained constriction. We refer to this pattern as "fixation." Cluster 2 (n = 174) consisted of trajectories with multiple alternating increases and decreases in pixel intensity, reflecting more than one constriction-release cycle driven by the oscillatory movement of the primary oral articulator during stuttering. We refer to this pattern as "oscillation." Both patterns were consistently observed in all participants, but their proportions vary (Fig. 4c). The relative proportions of the two patterns do not appear to be related to stuttering severity. Most participants showed a higher percentage of fixation. This is why fixation was the dominant pattern in the data.

Overall, clustering results aligned well with human judgments (Table 4). Out of 200 manually labeled time series, 165 (82.5%) time series are predicted correctly by the clustering algorithm. Most discrepancies arose from oscillatory cases that were misassigned to cluster 1 ("fixation").

We also repeated the clustering separately for labial, alveolar, and velar constriction trajectories. In each case, the optimal solution remained k = 2, and the resulting clusters again corresponded to fixation and oscillation (Supplementary Figure S1). Clustering solutions with k = 3 and k = 4 were also examined (Supplementary Figure S2). These solutions primarily subdivided the oscillation pattern into several variants, without revealing additional qualitatively distinct patterns.

3.1.2. Velum movement

A total of 726 VEL time series were clustered in a separate experiment, with results shown in Fig. 5. The silhouette analysis indicated that the optimal k was 2 (Fig. 5a), with a mean silhouette width of 0.61, suggesting a clear separation between clusters. The mean time series of each cluster (Fig. 5b) exhibited velum movements in opposite directions. Cluster 1 (n = 555) contained time series characterized by relatively high pixel intensity values that remained stable throughout the stuttering moment, reflecting a raised velum and a closed velopharyngeal port. Cluster 2 (n = 171) contained time series showing an overall decrease in pixel intensity followed by a rise, corresponding to the opening and subsequent closing of the velopharyngeal port.

Since Cluster 1 corresponded to velum gestures for oral sounds and Cluster 2 for nasal sounds, we further examined whether cluster assignment aligned with the oral-nasal distinction of the stuttered phoneme. The comparison showed strong alignment (Table 5). 89.39% of tokens were assigned to the cluster consistent with the oral/nasal category of the stuttered phoneme (oral → Cluster 1, nasal → Cluster 2). Errors were primarily due to a small number of oral tokens assigned to Cluster 2. Visual inspection of these tokens revealed considerable raising or lowering movement of the velum at the beginning or ending of these intervals (instead of stabilized at a raised position as in the mean trajectory of Cluster 1), but for most of the time within each token, the velum was in a raised position, and the velopharyngeal port was closed. Thus, the velum constriction trajectories during stuttering simply reflected whether the stuttered phoneme was nasal or oral, without exhibiting stuttering-related patterns.

3.2. Relationships between articulatory constriction patterns, perceptual types of disfluency, and phonological features of stuttered phonemes

Given that clusters in the VEL signal largely reflected the oral–nasal distinction rather than stuttering behavior, VEL clusters were not included in the statistical analyses. Only clusters derived from the primary oral articulators were used as the response variable in the models. Inter-rater reliability for the perceptual disfluency labels (repetition/prolongation/block) was Cohen's $\kappa = 0.86$ (95% CI [0.83, 0.89], N = 726), suggesting "almost perfect" agreement (Landis & Koch, 1977). The statistical results reported below are based on rater 1 labels; using rater 2 labels yielded the same pattern of significance/non-significance.

Table 3
Counts of time series by participant and by phonetic category.

	AWS1	AWS2	AWS3	AWS4	AWS5	AWS6	AWS7	Total
Labial stop	89	21	52	49	13	31	14	269
Labial continuant	79	31	23	44	8	14	11	210
Alveolar stop	25	20	18	38	1	17	6	125
Alveolar continuant	8	11	5	16	3	4	3	50
Velar stop	24	5	13	12	1	14	3	72
Total	225	88	111	159	26	80	37	726

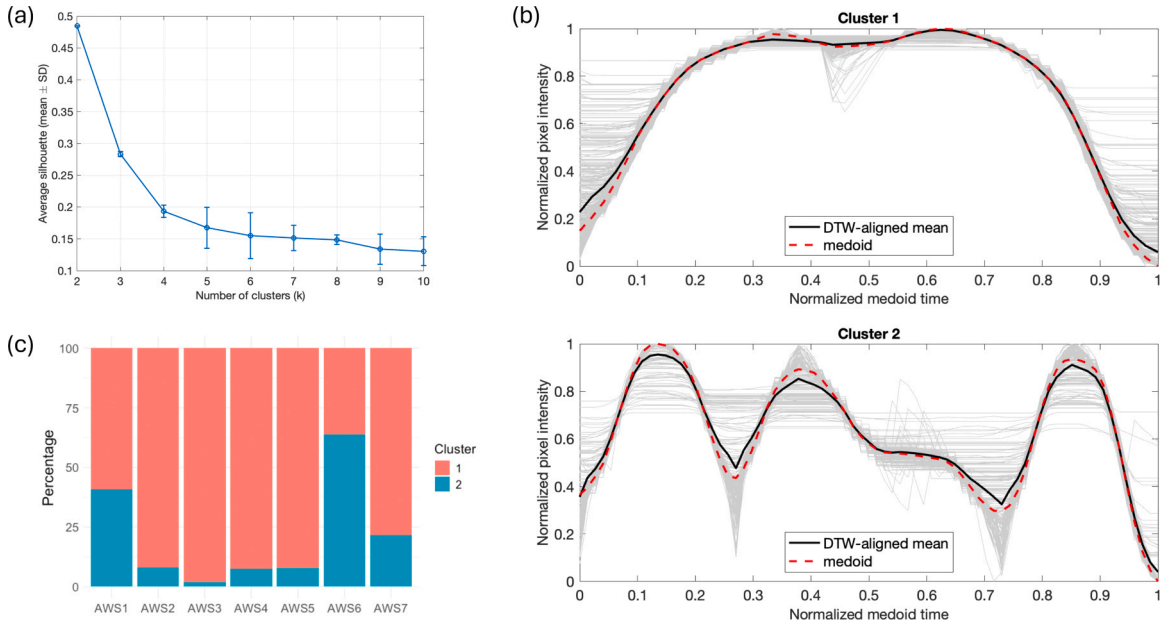

Fig. 4. Results of k-medoids clustering over time series of primary oral articulator. (a) Average silhouette width (mean ± SD) across k values, based on 20 replicates per k. (b) Clusters obtained for the optimal k, with time series aligned to the medoid time axis using dynamic time warping (black = cluster mean, red dashed = cluster medoid, gray = aligned individual signals). (c) The distribution of cluster 1 and cluster 2 in each participant.

Table 4
Confusion matrix comparing cluster assignment of primary oral articulator time series with human labeling of articulatory patterns (fixation vs. oscillation).

	Cluster 1	Cluster 2
Human_fixation	128	4
Human_oscillation	31	37

A logistic mixed-effects model tested whether the likelihood of exhibiting an oscillatory pattern (vs. fixation-like pattern) differed across perceptual disfluency types and places of articulation. There was a significant main effect of perceptual disfluency type. Relative to blocks, repetitions were significantly more likely to involve oscillatory patterns ($\beta = 2.27$, SE = 0.63, $p < .001$), whereas prolongations were marginally less likely to do so ($\beta = -1.91$, SE = 1.08, $p = .08$) (Fig. 6a). Place of articulation did not show a significant main effect: labial and velar consonants did not differ reliably from alveolars in the likelihood of involving oscillatory patterns (both $p > .50$). Because prolongations of velar consonants were rare in this dataset (n = 1), the interaction term involving the prolongation–velar combination was not interpretable. Among alveolars and labials, we did not observe evidence for a disfluency type × place interaction. This is also confirmed by a follow-up model fit to alveolar and labial tokens only: neither the prolongation–block contrast nor the repetition–block contrast differed between alveolars and labials ($p = .87$ and $p = .31$, respectively).

A second logistic mixed-effects model tested whether oscillatory patterns varied with perceptual disfluency type and manner class. For continuants (the baseline manner), prolongations were significantly less likely than blocks to involve oscillatory patterns ($\beta = -2.16$, SE = 0.98, $p = .027$). In contrast, repetitions did not differ significantly from blocks for continuants ($\beta = 0.81$, SE = 0.86, $p = .35$). Although repetitions of continuants showed a numerically higher proportion of oscillation than blocks (Fig. 6b), the estimated effect was small relative to its uncertainty ($|z| = 0.943$), hence not significant. Manner itself was not a significant main effect ($\beta = -0.70$, SE = 0.85, $p = .41$). However, the interaction between disfluency type and manner was significant for repetitions ($\beta = 2.40$, SE = 0.90, $p = .008$). This interaction indicates that the articulatory behavior of repetitions depends strongly on whether the stuttered segment is a stop or a continuant (Fig. 6b). Among continuants, repetitions behaved similarly to blocks. However, repetitions of stop consonants showed a sharply increased likelihood of oscillatory articulation, exceeding both blocks and prolongations of stops. The interaction involving prolongations was not significant, suggesting that the reduced likelihood of oscillation for prolongations relative to blocks was consistent across both stops and continuants.

Across both models, perceptual disfluency type was the strongest predictor of articulatory constriction pattern. In particular, prolongations consistently showed a reduced likelihood of articulator oscillation compared to blocks, regardless of whether place or manner was considered. Blocks tended to involve articulator fixation, and prolongations were even more so (Fig. 6a). The two models appeared to diverge in how repetitions behaved, but the findings are actually compatible. In the place model, repetitions were substantially more likely than blocks to involve articulator oscillation. In the manner model, this main effect disappeared because

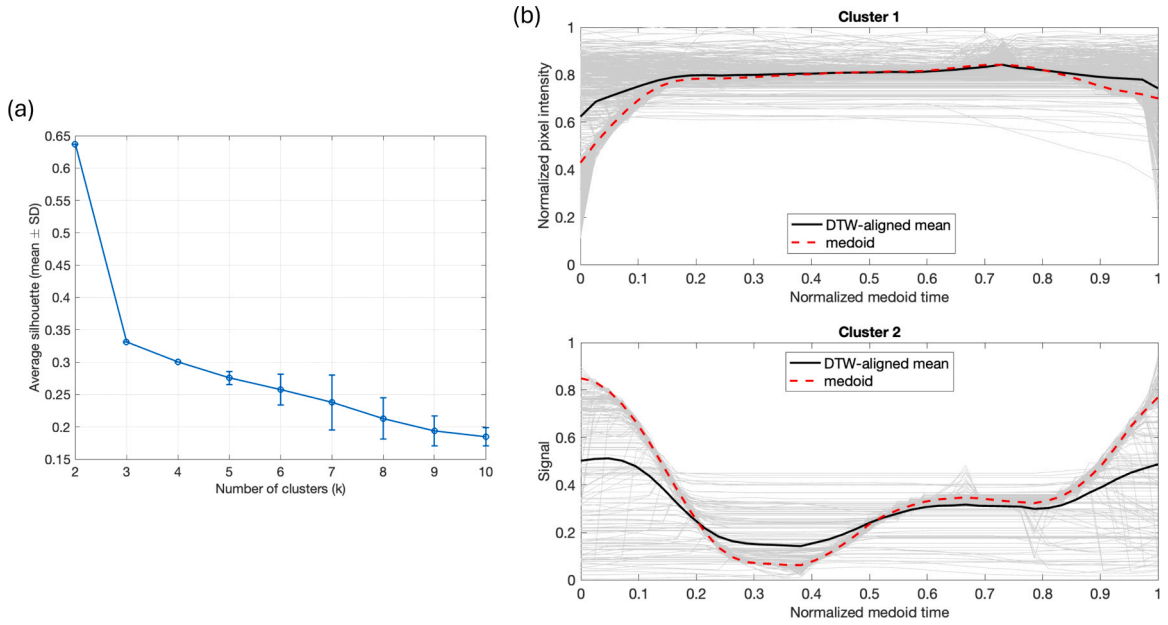

Fig. 5. Results of k-medoids clustering over time series of velum movement. (a) Average silhouette width (mean ± SD) across k values, based on 20 replicates per k. (b) Clusters obtained for the optimal k, with time series aligned to the medoid time axis using dynamic time warping (black = cluster mean, red dashed = cluster medoid, gray = aligned individual signals).

Table 5
Confusion matrix comparing cluster assignment of velum time series with the oral-nasal classification of the stuttered phoneme.

	Cluster 1	Cluster 2
True_oral	552	74
True_nasal	3	97

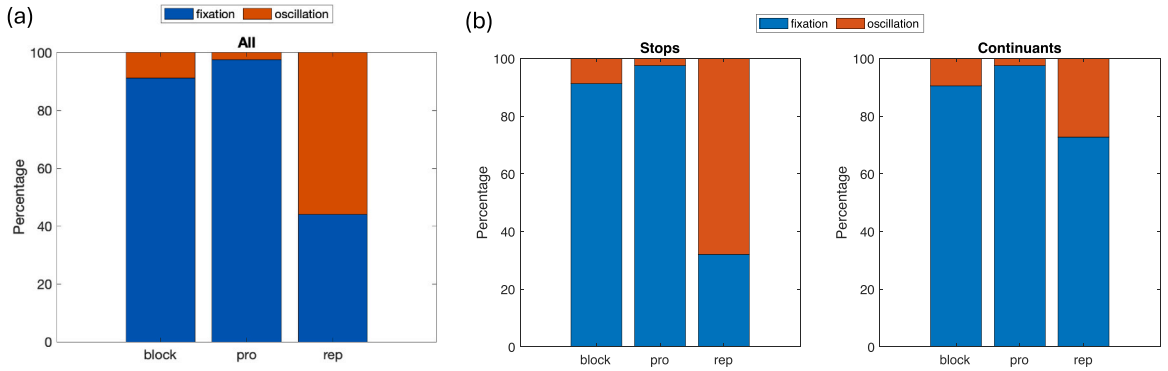

Fig. 6. Distribution of articulatory constriction patterns (fixation vs. oscillation) across perceptual disfluency types. (a) Percent composition of fixation and oscillation for blocks, prolongations, and repetitions when all data are pooled. (b) The same distribution shown separately for stops (left) and continuants (right).

repetitions did not behave uniformly across manner classes: repetitions of stops were highly oscillatory, whereas repetitions of continuants were not. Thus, the manner model revealed an underlying dependency that the place model did not capture, showing that the oscillatory behavior of repetitions is specific to stops rather than a general property of repetitions across all segment types.

To further illustrate how perceptual disfluency types arise from underlying articulatory behavior, we additionally examined the perceptual outcomes associated with each constriction pattern using a finer four-way manner division: plosive and nasal (both stops) and fricative and approximant (both continuants) (Fig. 7). Several patterns can be observed. Oscillations predominantly resulted in repetitions across all manner classes. In particular, oscillations in nasals yielded exclusively repetitions. This links to an earlier finding

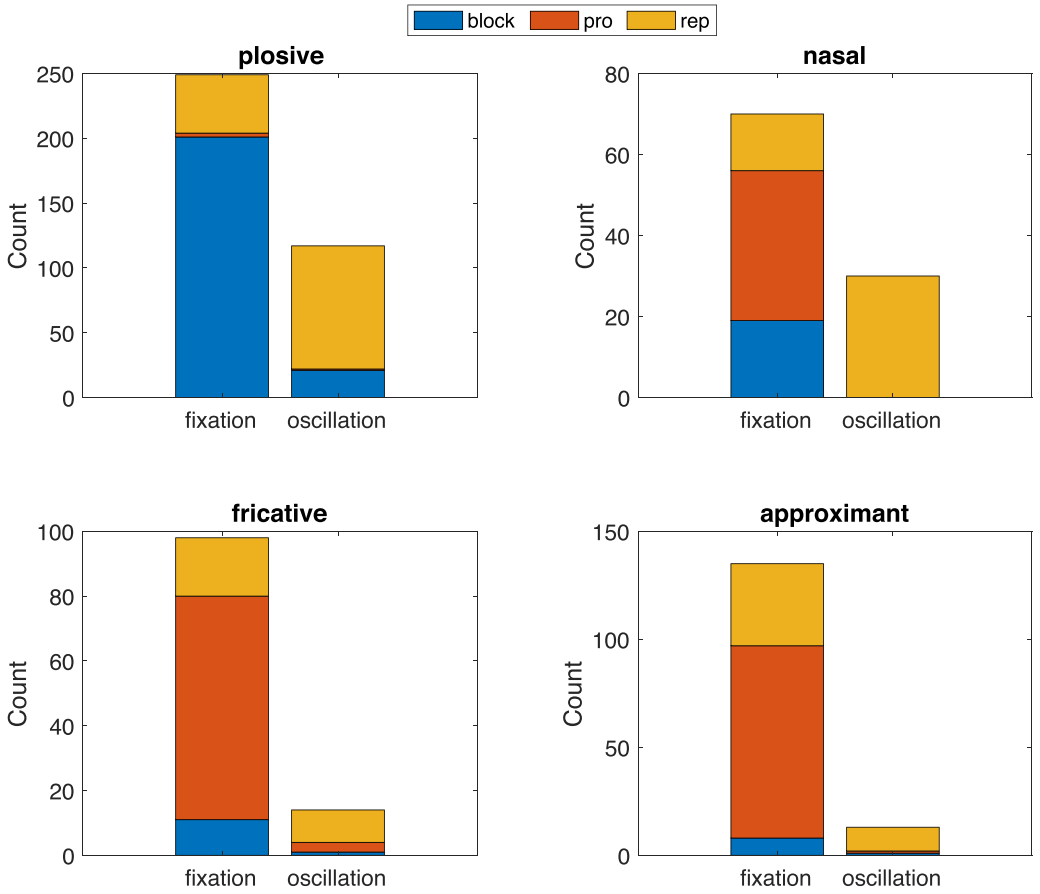

Fig. 7. Distribution of perceptual disfluency types (block, prolongation, repetition) within each articulatory constriction pattern (fixation, oscillation) across four manner classes.

from clustering analysis: the velum was always lowered when the stuttered phoneme was nasal. Because the nasal tract was open, oscillatory movement of the primary oral articulator was audible and perceived as repetitions. Fixations showed a different pattern: fixations in plosives predominantly resulted in blocks, whereas fixations in nasals, fricatives, and approximants mostly resulted in prolongations.

4. Discussion

Using clustering analysis, this study reveals two fundamental patterns in the constriction trajectories of primary oral articulators during repetitions, prolongations, and blocks. One pattern, *fixation*, features an overly extended constriction period. The other pattern, *oscillation*, features repeated cycles of constriction and release. Both patterns exist in all participants, with individual variability in their relative proportions. Velum constriction trajectories during stuttering moments, however, were not stuttering-related, but rather reflected whether the stuttered phoneme was nasal or oral. Thus, articulatory patterns associated with stuttering were expressed primarily through the oral articulators. Unsurprisingly, fixation and oscillation of the primary oral articulator were related to the perceptual types of disfluency, but the phonological properties of the stuttered phoneme played a role in shaping this relationship. Prolongations, regardless of phoneme type, were less likely to involve oscillation compared to blocks. Repetitions, on the other hand, were more likely to involve articulator oscillation than blocks and prolongations, but only when the stuttered segment was a stop, not when it was a continuant.

4.1. Two recurrent articulatory patterns underlie stuttering disfluencies

Repetitions, prolongations, and blocks are commonly treated as basic descriptive categories of stuttering. In practice, they are typically assigned by human raters based on auditory-perceptual characteristics of speech, which inevitably involve some degree of ambiguity and subjectivity. The present results, derived from a quantitative analysis of articulatory behavior, suggest that these perceptual categories are not the most basic level at which state stuttering is organized. Rather, at the level of articulatory constriction tasks, different perceptual disfluency types appear to arise from a smaller set of recurrent motor patterns. This suggests that the

articulatory basis of state stuttering is more structured and lower-dimensional than perceptual classification alone would imply. In this sense, the present findings shift the focus from perceptual labels to the articulatory organization underlying them.

These two patterns align with impressionistic descriptions that have surfaced repeatedly in the literature (Forster & Hardcastle, 1998; Teesson et al., 2003; Zimmermann, 1980b), as well as the classic taxonomy of clonic and tonic stuttering (Sheehan, 1974). More recent studies also found similar results. Lu et al. (2022) reported patterns of fixation and oscillation, although their third pattern, which resembled the spatiotemporal unfolding of a typical gesture but with a higher constriction degree, was not replicated here. Given that the present study examined a much larger number of stuttering tokens, the third pattern reported in Lu et al. (2022) may not represent a robust feature of stuttering. Fixation and oscillation are also consistent with two of the four patterns described in Didirková et al. (2019): maintained postures and reiterations of movements. The additional two patterns they identified, anarchical movements and combinations of all three types, however, may stem from the fact that they considered the movement of multiple articulators and identified patterns through manual analysis. The present study used unsupervised clustering to let recurrent patterns emerge from the data rather than assigning them impressionistically.

Depending on the theoretical framework adopted, the nature of fixation and oscillation may be interpreted in different ways. From a physiological perspective, fixation might reflect a spasmodic disruption, and oscillation might be interpreted as attempts to exit such disruption, or the result of tremor (Alm et al., 2025). From a feedback/feedforward control perspective, fixation may reflect the inability to recognize the right context for terminating the current gesture (Guenther, 2016); oscillation may reflect the commands to correct sensorimotor errors (Civier et al., 2010). Within the APTD framework, fixation and oscillation can be interpreted as two "unusual" modes in the temporal unfolding of an articulatory task, resulting from transitions between dynamical regimes (Lu, Iskarous, et al., 2025). The present study does not by itself determine which interpretation is more plausible. Further investigation is needed.

4.2. Articulatory patterns and perceptual disfluency types

Prolongations were less likely to involve oscillation than blocks and showed the highest proportion of fixation (Fig. 6a). This is expected since a percept of prolongation requires sustained constriction. Blocks also mostly corresponded to fixations, though with a slightly lower proportion (Fig. 6a). Manual inspection of RT-MRI videos revealed that a small number of blocks did involve articulator oscillation without clearly audible consequences, mostly in plosives (Fig. 7). In the other direction, whether a fixation is perceived as a prolongation or a block depended on the manner of the stuttered phoneme. Fixations in plosives mostly were heard as blocks, whereas fixations in nasals, fricatives, and approximants mostly were heard as prolongations (Fig. 7). This is expected because only when the vocal tract is not completely closed can air go through and generate sustained sounds. Repetitions were more likely to involve articulator oscillation than prolongations and blocks, but this held only for stops (plosives and nasals). For continuants (fricatives and approximants), repetitions did not reliably correspond to articulator oscillation. One possible reason is that, in continuants, intermittent airflow can create a percept of repetition even when the articulator remains relatively fixed.

Beyond the main questions addressed here, the data also suggest that stuttering on stops is more likely than stuttering on continuants to involve articulator oscillation (Fig. 7). This raises the possibility that the phonological properties of the intended segment shape not only how articulatory patterns map onto perceptual disfluency types, but also how stuttering itself is manifested articulatorily. One possible explanation is that different constriction goals involve different relations between auditory and somatosensory feedback during target attainment: for stops, the acoustic target is achieved before the somatosensory target (i.e., not just closure, but tissue compression), whereas in continuants the two are more closely aligned. If people who stutter generally have less efficient feedback control (Archibald & De Nil, 1999; Daliri et al., 2018; Loucks & De Nil, 2006; Max & Daliri, 2019), this difference may make repeated re-evaluation of the target attainment more likely in stops and sustained target holding more likely in continuants.

Taken together, perceptually defined repetition, prolongation, and block, although clinically useful, do not map transparently onto underlying articulatory behavior. From this perspective, articulatory labels such as fixation and oscillation may provide a more direct characterization of the motor basis of stuttering events.

4.3. Scope, limitations, and future directions

Although clustering allows intrinsic patterns in the data to be identified in an unsupervised, data-driven way, it was not error-free at the level of individual tokens. The most common misassignments involved oscillatory tokens being assigned as fixation, especially when releases were small or when the constriction signal did not reflect articulator oscillation (this happened in some repetitions of /k/ and /g/, where tongue oscillation can be clearly seen from the videos, but the dorsum constriction remained largely unreleased). The opposite error occurred less often and usually reflected image noise or minor movement during an otherwise sustained constriction. These cases do not undermine the identification of two recurrent articulatory patterns, but they do indicate that token-level classification is sensitive to signal quality and to the limits of the measurement method, which results in an underestimation of oscillatory patterns.

Another limitation is that the present analysis excluded tokens in which disfluency types were complex and overlapping, and therefore did not directly address the full range of multi-component stuttering events. At the same time, such cases need not fall outside the present framework. Fixation and oscillation may best be understood as basic components of more complex stuttering behaviors. In our data, mixtures of the two within a single stuttering moment were not uncommon, paralleling the fact that repetitions, prolongations, and blocks can co-occur within a single event. A key direction for future work is to systematically assess whether fixation and oscillation function as fundamental combinatorial units from which more complex stuttering behaviors are composed.

The present study did not code secondary stuttering behaviors. This issue is somewhat mitigated by the RT-MRI setting, which

constrained many prominent secondary behaviors, especially large head and body movements, because participants were required to remain relatively still in the scanner. In addition, the analysis focused on the constriction trajectory of the primary articulator responsible for the intended segment rather than on all visible movement during the stuttering event. Subtle secondary behaviors might still be present in the videos, but the recurrent patterns identified here are unlikely to be reducible to secondary behaviors alone. Fixation and oscillation were observed in all speakers, suggesting that they reflect shared aspects of state stuttering rather than purely idiosyncratic accessory behaviors.

The present study focused on the proximal articulators for the stuttered segment and on supralaryngeal constriction dynamics. A fuller account of the production basis of repetitions, prolongations, and blocks will require integrating articulatory evidence with phonatory and aerodynamic measures. It will also be important to examine how gestures coordinate during stuttering. For example, when the lips undergo fixation or oscillation during a disfluent labial consonant, initial evidence suggests that the tongue may nevertheless anticipate the following vowel (Harrington, 1987; Lu et al., 2022), but this warrants further validation. Addressing these questions will help clarify the articulatory manifestation of stuttered speech in a broader sense. Finally, although the present sample is larger than in most previous articulatory studies of stuttered speech, broader sampling will still be needed to determine how fully the patterns identified here generalize across speakers.

5. Conclusion

The overarching goal of this study was to uncover the articulatory basis of stuttering disfluencies. By grounding the analysis in the constriction tasks for the intended segment and focusing on the dynamics of the primary oral articulator, two recurrent patterns emerged in the present data: fixation and oscillation. These patterns were shared across speakers and cut across the traditional perceptual categories of repetition, prolongation, and block. The mapping between articulatory pattern and perceptual disfluency type was not one-to-one, and depended in part on the phonological properties of the stuttered segment. Taken together, these findings suggest that the articulatory organization of state stuttering is more structured and more basic than perceptual classification alone would imply. Articulatory characterization therefore complements auditory and acoustic description and provides a more direct window into the speech production processes involved in stuttering.

CRediT authorship contribution statement

Christina Hagedorn: Investigation, Methodology, Validation, Writing – review & editing. **Louis Goldstein:** Writing – review & editing, Supervision, Resources, Methodology, Investigation, Funding acquisition, Conceptualization. **Shrikanth Narayanan:** Writing – review & editing, Supervision, Resources, Methodology, Funding acquisition, Data curation. **Yijing Lu:** Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study is part of the doctoral dissertation of Yijing Lu at the University of Southern California (USC) and was supported by USC Morkovin Endowed Fellowship, USC Gold Family Fellowship, and U.S. National Science Foundation Grants #2311676, #2106930. Magnetic resonance imaging (MRI) data collection was conducted at and supported by the Dynamic Imaging Science Center at USC. We thank Prakash Kumar for his help with MRI data collection and Mary Yung for her help with the administrative coordination. We thank Dani Byrd, Khalil Iskarous, Charlotte E. E. Wiltshire and Krishna Nayak for useful discussions. We thank the National Stuttering Association, Christian Ting, Kanish Arhant-Sudhir, Max Gattie, Loryn McGill, Barbara Moore, and Mary Byram Washburn for their help with participant recruitment. Lastly, we thank all the participants who took part in this study.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jfludis.2026.106224.

Data availability

Data will be made available on request.

References

Alm, P. A., Brösemyr, T., Grinde, S., Johansson, S., Karlsson, J., Nordgren, G., Olsson, R., Rocksten, F., Sandsten, M., Sör, I., & White, D. (2025). Stuttering: The nature of the speech disruptions—A multimodal study of articulation and phonation. *Frontiers in Human Neuroscience, 19*, Article 1623308. https://doi.org/10.3389/fnhum.2025.1623308

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*(1), 25–42. https://doi.org/10.1016/S0094-730X(98)00023-0

Baker, R., & Hazan, V. (2011). DiapixUK: Task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods, 43*(3), 761–770. https://doi.org/10.3758/s13428-011-0075-y

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. https://doi.org/10.1111/ejn.12765

Blaylock, R. (2021). VocalTract ROI Toolbox [Computer software]. GitHub. ⟨https://github.com/reedblaylock/VocalTract-ROI-Toolbox⟩.

Boersma, P., & Weenink, D. (2023). Praat: Doing phonetics by computer (Version 6.3.14) [Computer software]. ⟨http://www.praat.org/⟩.

Browman, C. P., & Goldstein, L. (1990). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics, 18*(3), 299–320. https://doi.org/10.1016/S0095-4470(19)30376-6

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica, 49*(3–4), 155–180. https://doi.org/10.1159/000261913

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. https://doi.org/10.1093/brain/111.2.439

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. https://doi.org/10.1016/j.jfludis.2010.05.002

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. https://doi.org/10.1002/hbm.24063

Daliri, A., Wieland, E. A., Cai, S., Guenther, F. H., & Chang, S. (2018). Auditory-motor adaptation is reduced in adults who stutter but not in children who stutter. *Developmental Science, 21*(2), Article e12521. https://doi.org/10.1111/desc.12521

Darley, F.L., Aronson, A.E., & Brown, J.R. (1975). *Motor speech disorders* (3rd ed.). W.B. Saunders.

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*(2), 127–144. https://doi.org/10.1016/0094-730X(94)00024-N

Didirková, I., Maguer, S. L., Hirsch, F., & Gbedahou, D. (2019). Articulatory behaviour during disfluencies in stuttered speech. *Proceedings of ICPhS2019* (pp. 2991–2995).

Fairbanks, G. (1960). *Voice and articulation drillbook* (2nd ed).). Harper & Row.

Forster, C., & Hardcastle, W. (1998). An electropalatographic (EPG) study of the speech of two stuttering subjects. *International Journal of Language & Communication Disorders, 33*(S1), 358–363. https://doi.org/10.3109/13682829809179451

Frisch, S. A., Maxfield, N., & Belmont, A. (2016). Anticipatory coarticulation and stability of speech in typically fluent speakers and people who stutter. *Clinical Linguistics & Phonetics, 30*(3–5), 277–291. https://doi.org/10.3109/02699206.2015.1137632

Guenther, F. H. (2016). *Neural control of speech*. The MIT Press. https://doi.org/10.7551/mitpress/10471.001.0001

Harrington, J. (1987). Coarticulation and stuttering: An acoustic and electropalatographic study. In H. F. M. Peters, & W. Hulstijn (Eds.), *Speech Motor Dynamics in Stuttering* (pp. 381–392). Springer. https://doi.org/10.1007/978-3-7091-6969-8_30.

International Phonetic Association. (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge University Press.

Jackson, E. S., Tiede, M., Beal, D., & Whalen, D. H. (2016). The impact of social–cognitive stress on speech variability, determinism, and stability in adults who do and do not stutter. *Journal of Speech, Language, and Hearing Research, 59*(6), 1295–1314. https://doi.org/10.1044/2016_JSLHR-S-16-0145

Kaufman, L., & Rousseeuw, P. (1990). *Finding groups in data: An introduction to cluster analysis. New York*: Wiley. https://doi.org/10.2307/2532178

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. https://doi.org/10.1044/jslhr.4302.548

Kumar, P., Tian, Y., Lim, Y., Cui, S. X., Hagedorn, C., Byrd, D., & Nayak, K. S. (2024). State-of-the-art speech production MRI protocol for new 0.55 Tesla scanners. *Proc. Interspeech 2024* (pp. 2590–2594).

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310

Lim, Y., Kumar, P., & Nayak, K. S. (2024). Speech production real-time MRI at 0.55. *T Magnetic Resonance in Medicine, 91*(1), 337–343. https://doi.org/10.1002/mrm.29843

Loucks, T. M. J., & De Nil, L. F. (2006). Oral kinesthetic deficit in adults who stutter: A target-accuracy study. *Journal of Motor Behavior, 38*(3), 238–247. https://doi.org/10.3200/JMBR.38.3.238-247

Lu, Y., Goldstein, L., Byrd, D., & Narayanan, S. (2025). Effective elicitation of stuttering in magnetic resonance imaging data collection using a suite of connected speech tasks. *Journal of Speech, Language, and Hearing Research, 68*(9), 4275–4289. https://doi.org/10.1044/2025_JSLHR-24-00764

Lu, Y., Iskarous, K., & Goldstein, L. (2025). Towards a dynamical model of transitions between fluent and stuttered speech. *Proc Interspeech 2025*, 310–314. https://doi.org/10.21437/Interspeech.2025-2320

Lu, Y., Wiltshire, C. E. E., Watkins, K. E., Chiew, M., & Goldstein, L. (2022). Characteristics of articulatory gestures in stuttered speech: A case study using real-time magnetic resonance imaging. *Journal of Communication Disorders, 97*, Article 106213. https://doi.org/10.1016/j.jcomdis.2022.106213

MacPherson, M. K., & Smith, A. (2013). Influences of sentence length and syntactic complexity on the speech motor control of children who stutter. *Journal of Speech, Language, and Hearing Research, 56*(1), 89–102. https://doi.org/10.1044/1092-4388(2012/11-0184)

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. https://doi.org/10.1044/1092-4388(2003/017)

Max, L., & Daliri, A. (2019). Limited pre-speech auditory modulation in individuals who stutter: data and hypotheses. *Journal of Speech, Language, and Hearing Research, 62*(8S), 3071–3084. https://doi.org/10.1044/2019_JSLHR-S-CSMC7-18-0358

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal forced aligner: trainable text-speech alignment using kaldi. *Proceedings of Interspeech 2017* (pp. 498–502). https://doi.org/10.21437/Interspeech.2017-1386

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*(3), 310–318. https://doi.org/10.1007/s00221-004-1952-8

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. https://doi.org/10.1044/1092-4388(2004/024)

Narayanan, S., Nayak, K., Lee, S., Sethy, A., & Byrd, D. (2004). An approach to real-time magnetic resonance imaging for speech production. *The Journal of the Acoustical Society of America, 115*(4), 6.

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. https://doi.org/10.1002/mrm.24276

Packman, A., & Attanasio, J. S. (2017). *Theoretical issues in stuttering* (2nd ed.).). Routledge. https://doi.org/10.4324/9781315636559

Ponssen, D., Joseph, A. A., Neef, N. E., Frahm, J., & Sommer, M. (2024). Visualising the dynamic morphology of stuttering using real-time MRI. *The Lancet, 403*(10438), 1789–1790. https://doi.org/10.1016/S0140-6736(24)00624-X

Riley, G.D. (2009). Stuttering severity instrument: SSI-4. Pro-Ed.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1*(4), 333–382. https://doi.org/10.1207/s15326969eco0104_2

Sasisekaran, J. (2013). Nonword repetition and nonword reading abilities in adults who do and do not stutter. *Journal of Fluency Disorders, 38*(3), 275–289. https://doi.org/10.1016/j.jfludis.2013.06.001

Sheehan, J. G. (1974). Stuttering behavior: A phonetic analysis. *Journal of Communication Disorders, 7*(3), 193–212. https://doi.org/10.1016/0021-9924(74)90031-8

Smith, A., Goffman, L., Sasisekaran, J., & Weber-Fox, C. (2012). Language and motor abilities of preschool children who stutter: Evidence from behavioral and kinematic indices of nonword repetition performance. Journal of Fluency Disorders, Special Issue: 9th Oxford Dysfluency Conference, 37(4), pp. 344–358. https://doi.org/10.1016/j.jfludis.2012.06.001.

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. https://doi.org/10.1044/jslhr.4302.521

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. https://doi.org/10.1016/j.jfludis.2009.12.001

Teesson, K., Packman, A., & Onslow, M. (2003). The Lidcombe behavioral data language of stuttering. *Journal of Speech, Language, and Hearing Research, 46*(4), 1009–1015. https://doi.org/10.1044/1092-4388(2003/078)

Uecker, M., Zhang, S., Voit, D., Karaus, A., Merboldt, K.-D., & Frahm, J. (2010). Real-time MRI at a resolution of 20 ms. *NMR in Biomedicine, 23*(8), 986–994. https://doi.org/10.1002/nbm.1585

Usler, E., Smith, A., & Weber, C. (2017). A lag in speech motor coordination during sentence production is associated with stuttering persistence in young children. *Journal of Speech, Language, and Hearing Research, 60*(1), 51–61. https://doi.org/10.1044/2016_JSLHR-S-15-0367

Wiltshire, C. E. E., Chiew, M., Chesters, J., Healy, M. P., & Watkins, K. E. (2021). Speech movement variability in people who stutter: A vocal tract magnetic resonance imaging study. *Journal of Speech, Language, and Hearing Research, 64*(7), 2438–2452. https://doi.org/10.1044/2021_JSLHR-20-00507

Zimmermann, G. (1980b). Articulatory behaviors associated with stuttering. *Journal of Speech, Language, and Hearing Research, 23*(1), 108–121. https://doi.org/10.1044/jshr.2301.108

Zimmermann, G. (1980a). Articulatory dynamics of fluent utterances of stutterers and nonstutterers. *Journal of Speech, Language, and Hearing Research, 23*(1), 95–107. https://doi.org/10.1044/jshr.2301.95

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), 35–42. https://doi.org/10.1044/jshr.2601.35

Yijing Lu received Ph.D. in Linguistics from University of Southern California, United States in 2024. She is currently a postdoctoral researcher in the department of Linguistics, University of Potsdam, Germany. Her research interests include phonetics, speech production, articulation, stuttering, and real-time speech MRI.

Louis Goldstein received Ph.D. in Linguistics from University of California, Los Angeles, United States in 1977. He is currently a full professor in the department of Linguistics, University of Southern California, United States. His research interests include phonetics, speech production, articulatory modeling, dynamical models.

Christina Hagedorn received Ph.D. in Linguistics from University of Southern California, United States in 2015. She is currently an Associate Professor at the College of Staten Island, City University of New York, United States, and a faculty member in the doctoral programs in Linguistics and Speech-Language-Hearing Sciences at the CUNY Graduate Center. She is a certified speech-language pathologist. Her research program focuses on speech motor control and motor speech disorders, including speech production under altered anatomical and biomechanical constraints.

Shrikanth Narayanan received Ph.D. in Electrical Engineering from University of California, Los Angeles, United States in 1995. He is currently University Professor and holder of the Niki and Max Nikias Chair in Engineering at the University of Southern California (USC) and serves as the inaugural Vice President for Presidential Initiatives on the Senior Leadership Team of USC's President. He is a Professor in the Signal and Image Processing Institute of USC's Ming Hsieh Electrical & Computer Engineering department with joint appointments as Professor in Computer Science, Linguistics, Psychology, Neuroscience, Pediatrics and Otolaryngology-Head and Neck Surgery. He is also the inaugural director of the Ming Hsieh Institute, a Research Director for the Information Sciences Institute at USC and a Visiting Faculty Researcher at Google. His research interests include human communication and information processing, theory & applications of signal processing, information/estimation theory, statistical systems modeling, speech and language processing & automatic speech recognition, speech production modeling, articulatory acoustics, speech synthesis, biomedical signal processing and modeling, imaging & instrumentation for speech research, human machine interaction, multimodal-multimedia interfaces, devices, and systems.