JUNE 18 2026

Perturbing postural stability affects acoustic properties but not fluency of adult speech production during face-to-face interaction[a)]

Mingtong Li ; Jiahao Eric Yang ; Susanne Fuchs ; Suzanne Aussems

J. Acoust. Soc. Am. 159, 5491–5503 (2026)
https://doi.org/10.1121/10.0043950

JASA ARTICLE

Perturbing postural stability affects acoustic properties but not fluency of adult speech production during face-to-face interaction[a)]

Mingtong Li,[1,2] Jiahao Eric Yang,[1,3] Susanne Fuchs,[4] and Suzanne Aussems[1,b)]
[1]*Department of Psychology, University of Warwick, Coventry, United Kingdom*
[2]*Department of Language and Linguistic Science, University of York, York, United Kingdom*
[3]*Institute for Digital Security and Behaviour, University of Bath, Bath, United Kingdom*
[4]*Leibniz-Zentrum Allgemeine Sprachwissenschaft, Berlin, Germany*

ABSTRACT:
Physical constraints, particularly those that challenge postural stability and control, may measurably influence speech production. This experimental study investigated how increased postural stability demands affect adults' speech production during an interactive word-guessing game. A total of 118 adults (59 pairs) played a game of *Taboo* while standing either on a stable surface (the ground) or an unstable surface (a wobble board), with the latter increasing demands on postural stability. Participants alternated between the roles of clue-giver and guesser; analyses focused exclusively on speech produced by clue-givers, who were instructed to describe target words (e.g., *coffee*) while avoiding five conceptually related Taboo words (e.g., *latte*, *drink*, *morning*, *beans*, *caffeine*). Acoustic analyses revealed that when clue-givers described target words to guessers while standing on an unstable surface, they produced speech with higher fundamental frequency and higher amplitude envelope, compared to the stable surface. In contrast, analyses of fluency measures, which included speech rate, pause duration, pause frequency, and filler particle frequency, showed no notable differences between conditions. Perturbing postural stability thus affects the acoustics but not the fluency of speech production.
 https://doi.org/10.1121/10.0043950

(Received 27 February 2026; revised 8 May 2026; accepted 14 May 2026; published online 18 June 2026)
[Editor: Ccile Fougeron] Pages: 5491–5503

I. INTRODUCTION

Speech production is a complex motor process that may be influenced by physical, anatomical, and postural constraints (e.g., Maassen *et al.*, 2007; Perrier, 2006; Shamei, 2024). Humans often communicate while standing (Stone *et al.*, 2007), a posture that requires continuous postural control to maintain the body's center of mass over the small base of support provided by the feet (Winter *et al.*, 2003). Because of this, the biomechanics of speaking are not limited to the articulators; rather, they involve coordinated activity across the whole body, including postural and limb movements, which influence vocal modulation and acoustic output (Pouw *et al.*, 2025). Understanding how such whole-body constraints alter speech acoustics can provide insight into the flexibility and limits of the human speech production system. Therefore, this study perturbed the postural control system in an experiment to examine how manipulating postural stability influenced adults' speech production processes in a face-to-face communication game.

Manipulating postural stability allows us to test how speech production operates when the body's balance system is placed under increased demand. Maintaining balance places cognitive and sensorimotor demands on the nervous system and may draw on resources that are also involved in speech production (Kerr *et al.*, 1985; Massion, 1994; Woollacott and Shumway-Cook, 2002). We therefore first examined global acoustic properties of speech, as speech output depends on coordinated respiratory and articulatory control. The relationship between posture and voice production is well established in clinical speech-language therapy (Cardoso *et al.*, 2019) and has also begun to receive attention in experimental work examining biomechanical influences on speech (Pouw *et al.*, 2025). Changes in acoustic measures such as fundamental frequency (f0) and amplitude envelope would suggest that postural perturbation directly influences the motor execution of speech. We then assessed speech fluency, including pauses and disfluencies, to test whether increased postural demands also affect, at least in part, the temporal organization of speech. Examining these two sets of variables enables us to determine whether destabilizing posture selectively influences the acoustic realization of speech, its fluency, or both.

[a)]This article is part of a Special Issue on Data and Models in Speech Production: A Situated Perspective.
[b)]Email: s.aussems.1@warwick.ac.uk

A. Body posture as a physical constraint on speech production

Humans experience physical constraints from their bodies when speaking. For example, magnetic resonance imaging (MRI) studies have investigated the influence of body posture on speech, because participants usually lie supine in the scanner, which is a position quite different from typical upright speaking. To investigate potential effects of this posture, several studies examined vowel production with different body positions (e.g., Stone *et al.*, 2007; Vorperian *et al.*, 2015). Although Stone *et al.* (2007) reported weak effects and individual differences, Vorperian *et al.* (2015) reported a higher f0 in the supine position compared to the upright position. Furthermore, comparing the body postures of speakers in upright, supine and a -10° inverted position, Engström *et al.* (2024) reported small, participant-specific effects on acoustic voice quality measures, with changes localized to specific sound pressure level and f0 ranges. Thus, body posture can alter the acoustic properties of adult speech, although the effects tend to be small and variable across individuals.

Dynamic actions like gesturing while speaking also act as internal postural perturbations that require anticipatory bodily adjustments. Pouw *et al.* (2025) investigated this by having adult participants perform arm movements, including trials with a 1 kg wrist weight on the arms, during sustained vowel vocalizations. They reported that the onset of postural muscle activity often precedes or coincides with the changes in the amplitude envelope and that the overall body could also be tensed to counteract the weight perturbations. This suggests that upper limb movements generate physical impulses that perturb postural stability, invoking whole-body muscle movements that physically constrain the respiratory-vocal system to counteract the perturbation. Because the lower body serves as a fundamental base of support for human upright posture, adjusting to balance perturbations in the lower limbs should similarly require widespread postural stabilization. Hence, our study extends this previous work by investigating how continuous perturbations of postural stability in the lower body affect concurrent speech production.

B. Cognitive load and speech planning

Postural control may influence how people speak indirectly by increasing cognitive load because maintaining balance under physical constraint may draw on shared attentional resources. Dual-task studies report measurable interference between balance and cognition in adults: as balancing difficulty increases, reaction times on concurrent mental tasks slow, and accuracy on high-load tasks declines, consistent with general capacity limitations (Yardley *et al.*, 2001). Related work in healthy adults also indicates that adding a concurrent cognitive task modifies postural sway, particularly when the postural task requires less well-practiced control strategies (Dault *et al.*, 2001), supporting the view that postural control is not fully automatic (Woollacott and Shumway-Cook, 2002). Importantly, the reverse is also true because articulation itself can also influence postural sway. For example, healthy adults who perform tasks requiring verbal responses produce more pronounced changes in sway frequency and sway path than when they perform equivalent tasks silently (Dault *et al.*, 2003), highlighting that "talking while balancing" combines attentional and speech motor/respiratory demands. We see the real-world consequences of this interference in aging populations; older adults, for instance, face a higher risk of falls when performing manual or cognitive tasks while standing or walking (Menant *et al.*, 2014). Taken together, these findings suggest that a relationship exists between postural stability and speaking, mediated by cognitive load; however, the directionality of the relationship is not always clear.

Cognitive load, in turn, shapes speech by constraining the resources available for planning and monitoring. Disfluencies and hesitations have become an important window into language processing in real-world communication (Corley and Stewart, 2008; Ferreira and Bailey, 2004). When planning demands increase, speakers produce slower speech with more and longer silent pauses (Mitchell *et al.*, 1996) and divided-attention dual tasks increase filled pauses and repetitions during narrative production (Oomen and Postma, 2001). Cognitive workload can also affect acoustic output, for example, by increasing vocal intensity and altering pitch-related measures such as f0 variability during simultaneous task performance (Lively *et al.*, 1993). Recent multimodal research provides objective physiological evidence for this resource drain. Using pupillometry as a direct measure of cognitive effort, Mangalam *et al.* (2025) demonstrated that healthy young adults exert significantly greater cognitive effort simply to maintain their balance on an unstable surface. Importantly, they found that the mechanical challenge of balancing on a mediolateral wobble board directly drives this pupil dilation, independent of physical fatigue. Taken together, this evidence supports the prediction that when postural control becomes attentionally demanding (e.g., when standing on an unstable surface), fewer resources remain for speech planning and monitoring, increasing the likelihood of slower, more disrupted speech.

C. The current study

This experimental study investigated whether perturbing postural stability influences the acoustics and fluency of speech production during face-to-face communication. Participants completed a Taboo-style word-guessing task in pairs, alternating between two roles: clue-giver and guesser. We manipulated postural stability only when participants served as the clue-giver, asking them to stand either on an unstable surface (a wobble board) or a stable surface (the ground). Guessers always stood on the ground. In each trial, the clue-giver received a target word (e.g., *coffee*) together with a set of five conceptually related "taboo" words (e.g., *latte*, *drink*, *beans*, *morning*, *caffeine*). The clue-giver then had one minute to describe the target word without producing the taboo words. The guesser attempted to identify the target word within the time limit. Across rounds, participants switched roles so that each person served as clue-giver in both the unstable (wobble board) and stable (ground) conditions. All speech was recorded, but only the speech produced by clue-givers was analyzed as postural stability was manipulated only in this role.

If postural stability influences speech production, then adult participants should show acoustic and fluency differences between the unstable and stable conditions. Specifically, we examined six confirmatory hypotheses. For the acoustic measures, we predicted that clue-givers would produce speech with a higher f0 (H1) and a higher amplitude envelope (H2) in the unstable condition relative to the stable condition. For the fluency measures, we predicted that clue-givers would produce speech at a faster rate (H3), produce longer pauses (H4), pause more frequently (H5), and produce more filler particles (H6) in the unstable condition relative to the stable condition.

Finally, we conducted two manipulation checks. First, we confirmed that the wobble board increased postural stability demands by measuring postural sway with a gyroscope. Next, we compared gyroscope metrics from the unstable condition between speaking and non-speaking segments. If the wobble board indeed increases postural stability demands, we would expect significant differences in postural sway between speaking and non-speaking segments in this condition. This expectation is consistent with prior research showing that adults exhibit greater postural sway when speaking compared with performing equivalent tasks silently (Dault *et al.*, 2003).

II. METHODS

The study materials were documented on the Open Science Framework (OSF), including demographic information about participants, stimulus materials, lab setup and walkthrough video of the experiment procedure, PRAAT coding manual, and R analysis scripts and data files (Li *et al.*, 2026, https://osf.io/68uxv/).

A. Design

The experiment had a within-subject design. There was one independent variable, namely, postural stability, which had two levels: participants stood on the ground in the stable condition and on a wobble board in the unstable condition. The dependent variables were f0, amplitude envelope, speech rate, pause duration, pause frequency, and filler particle frequency. All participants completed the word-guessing game as both clue-giver and guesser, switching roles across 4 rounds of 10 trials each (40 in total). Because the postural stability manipulation applied only to clue-givers, our analysis focused on the variables of interest in this role, where each clue-giver completed 10 trials in the stable and 10 in the unstable conditions (20 trials in total).

B. Participants

The study involved a final sample of 118 adult participants (19 males, 95 females, 2 non-binaries, 2 prefer not to disclose), ranging between 18 and 27 years old [mean (M) = 18.93 years, standard deviation (SD) = 1.36 years]. All participants were university students who enrolled in an English-taught degree and self-reported high English language proficiency. Of the 118 participants, 98 participants reported English as their first language, including 35 participants who also reported learning at least 1 additional language from birth. Furthermore, none of the participants reported any health or medical conditions (e.g., hearing problems, mobility issues, balance disorders, sports injuries in the past six months, and pregnancy). An additional 14 participants (7 pairs) were tested, but their data were excluded from the analysis because of (1) continuous microphone popping or rustling seriously affecting the audio quality of the recordings ($n = 8$), (2) low English language proficiency and speaking a foreign language throughout the task ($n = 2$), (3) not adhering to the balance manipulation ($n = 2$), and (4) for several of the aforementioned reasons ($n = 2$). We also recorded participants' hand and leg dominance as well as their height and weight for later kinematic analysis, but we do not report those measures here as they are outside of the scope of this study. The study received full ethical approval from the Humanities & Social Sciences Research Ethics Committee at University of Warwick (reference: HSSREC 28/24–25) and informed written consent was obtained for all participants.

C. Trial exclusions

We excluded 25 trials of a total of 2360 trials. Of the included pairs, 39 had no trial exclusions, 15 had 1 trial exclusion, and 5 had 2 trial exclusions. Trials were excluded for the following reasons: (1) persistent microphone popping or rustling noise ($n = 6$), (2) persistent external noise (e.g., lawn mower or car engine outside, $n = 2$), (3) experimenter accidentally skipping a trial via button press ($n = 8$), (4) participant skipping a trial ($n = 1$), participants asking the experimenter questions mid-trial ($n = 4$), (5) participants accidentally saying the target word ($n = 3$), and (6) participants speaking a language other than English ($n = 1$). The total number of valid trials was 2335.

D. Materials

Stimulus words were created using a data-driven approach drawing on existing datasets of sensory experience and semantic relatedness for English words. First, a set of target words was selected based on sensory experience ratings from Juhasz and Yap (2013). Words with high sensory ratings were chosen to maximize perceptual richness and to encourage rich communicative behaviors. For each target word, a set of five conceptually related taboo words was generated using ConceptNet, a large-scale semantic network of general knowledge (Speer *et al.*, 2017). Conceptual relatedness was operationalized using association strength data derived from Verbosity, an association-based online game embedded within ConceptNet that collects common-sense semantic relations (von Ahn *et al.*, 2006). For each target, the five taboo words with the strongest conceptual associations were selected (see full set of stimulus words in the Appendix, Table I).

E. Procedure

A schematic overview of the testing area is depicted in Fig. 1. Participants attended the session in naturally occurring pairs, as determined by the undergraduate research pool's sign-up system from the Department of Psychology at the University of Warwick. Upon arrival, the experimenter guided participants individually to separate quiet cubicles. In each cubicle, participants read the Participant Information Leaflet, provided informed consent, and completed a Qualtrics survey that included demographic questions (e.g., age, gender, language). Participants also completed a series of health-related questions (e.g., history of balance difficulties, hearing problems, mobility issues, recent sports injuries, pregnancy). The health-related questions were included solely to assess participants' eligibility to take part in the study and to ensure participants' safety while standing on the wobble board. Participants also indicated their dominant hand and leg and whether they were paired with a friend or stranger in the survey.

After completing the survey, participants were fitted with the study equipment: a hair net, a protective helmet, and a lanyard holding a wireless microphone sender approximately 15 cm from the participant's mouth. The experimenter verified that all equipment was correctly positioned and functioning before proceeding. The experimenter then introduced the Taboo game and delivered the following instructions:

"You're going to play a game called Taboo. There are two roles in this game: the clue-giver and the guesser. The clue-giver will see a target word and five taboo words on a screen which is not visible to the guesser. The clue-giver's goal is to make the guesser say the target word within 60 seconds. The clue-giver cannot say the target word or any of the taboo words. However, the clue-giver may use other spoken words, hand gestures, and body language. If the clue-giver accidentally says a taboo word, just continue playing—I will make a note of it. You will play four rounds in total, and the roles are assigned before each round, so you will each have chances to be both the clue-giver and the guesser. Please always stay behind the green line in your designated area. The clue-giver will always stand on this side (pointing to the clue-giver side of the room), and the guesser on that side (pointing to the guesser side of the room)."

After providing the instructions, the experimenter answered any questions from participants before proceeding.

1. Wobble board practice session

Following the instructions, participants completed a wobble board practice session. The experimenter switched on a side-view camera and activated a gyroscope attached to the wobble board, then demonstrated safe stepping on and off the board. Participants practiced stepping and balancing with the wobble board in turns until they felt comfortable and consented to proceed. Note that participants wore shoes when standing on the wobble board and were advised before

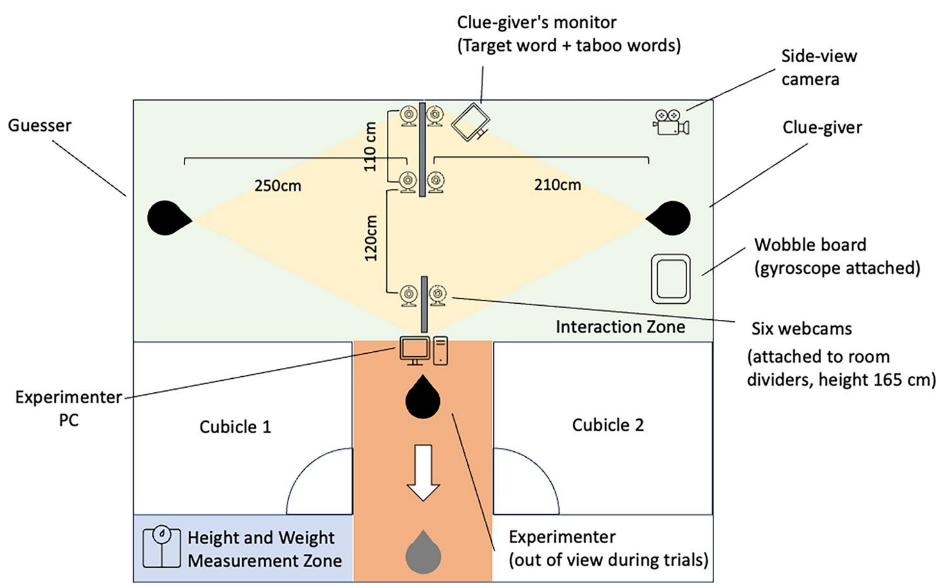

FIG. 1. Experimental setup and room layout. The green area shows the participants' interaction zone, the yellow area shows the participants' viewing zone, the orange area shows the experimenter's movement zone, and the blue area shows the height and weight measure zone. Two cubicles were used for participants to read the information sheet and complete the consent form as well as a survey. Six webcams were mounted on the room dividers (gray bars), and an additional side-view camera recorded the clue-giver's movements. A monitor on the clue-giver's side displayed the target word and the taboo words, visible only to the clue-giver. The experimenter initiated each trial from the experimenter PC and then moved out of view of participants until the target word was correctly guessed or the trial time elapsed.

attending to wear clothing and footwear suitable for light physical activity.

2. Camera calibration

Once the participants agreed to proceed, the experimenter asked the participants to move to the two designated areas. These areas were marked with green tape on opposite sides of the room, ensuring full-body visibility across six cameras. Each participant briefly adopted a T-pose (standing upright with both arms extended horizontally) for camera calibration.

3. Practice phase

Participants completed four practice trials (up to 60 s each). On each practice trial, the clue-giver viewed a screen displaying one target word (in white) and five taboo words (in black). For each participant, the first practice trial as clue-giver took place on the ground (stable condition), and the second took place on the wobble board (unstable condition). After two practice trials, the participants switched roles by walking to the opposite sides of the room and repeated the two-trial practice sequence with new words.

To minimize experimenter influence in the dyadic interaction (e.g., making eye contact with one of the participants), the experimenter remained out of the participants' line of sight during all trials and only approached the experimenter personal computer (PC) to advance to the next trial in between trials.

4. Experimental phase

The main experiment consisted of four rounds in total. Each round consisted of ten trials. In each trial, the clue-giver viewed a target word and 5 taboo words and attempted to depict the target word to the guesser within 60 s. The experimenter advanced the trial via button press after the guesser produced the correct target word. If the target word was not identified within the time limit, the participants were allowed to finish their interaction before proceeding.

After each round, participants switched roles. For each participant's clue-giver turns, the order of stable (ground) vs unstable (board) conditions was counterbalanced. For half the pairs the clue-giver started in the stable (ground) condition and half the pairs in the unstable (board) condition. Trials were randomized live during the experiment.

5. Body measurement

After completing all rounds of the word-guessing game, participants removed the equipment as well as their shoes and were guided individually to a private measurement zone where their height and weight were recorded, for future kinematic analyses (not reported here).

F. Apparatus

Participants stood on a wooden wobble board (44.5 cm × 34.3 cm × 7.6 cm; StrongTek Professional Wobbleboard, StrongTek, Sugar Land, TX) with a slip-resistant surface (Fig. 2). The board featured two cylindrical supports that restricted movement to the mediolateral axis (similar to those in Mangalam *et al.*, 2025), specifically challenging lateral balance while maintaining stability in the anterior-posterior direction. This design was chosen over a 360° omnidirectional board to ensure participant safety and to prevent rotations that would have disrupted the face-to-face communicative interaction between partners.

Wobble board motion was recorded via a WitMotion gyroscope (BWT901CL) attached to the underside of the board, which transmitted three-axis acceleration and orientation data to an iPad via Bluetooth. To accommodate battery life across sessions, two identical units were used interchangeably, ensuring the same mounting position and recording procedure were maintained throughout.

Sessions were recorded using six Elgato Facecams (Corsair Gaming, Inc., Milpitas, CA) (1920 × 1080 pixels, 30 frames per second), with three cameras capturing each participant. An additional side-view camera (Canon Legria HF R66, Canon Inc., Tokyo (Ota-ku), Tokyo, Japan) monitored the clue-giver to ensure compliance with the balance task, verifying that the board did not touch the ground and that participants did not lean against the wall. This recording served as a crucial *post hoc* check as the experimenter moved out of view during the trials. Audio was captured via two DJI Mic Mini (SZ DJI Technology Co., Ltd., Shenzhen, Guangdong, China) wireless lavalier microphones, each with an omnidirectional pickup pattern and a 20 Hz–20 kHz frequency response. Absolute amplitude calibration was not performed as the study utilized a within-subjects design focused on relative changes between conditions. Because the microphone distance and gain settings remained constant for each participant across both stable and unstable trials, these relative measures provided a reliable basis for comparing the effects of postural instability.

Finally, all participants wore a hairnet (for hygiene reasons) and a helmet (for safety reasons) for the entire duration of the session. The helmet was adjustable to the size of

FIG. 2. Wooden wobble board. Top, the wobble board from frontal view, bottom, the arch underneath the wobble board.

their head and attached with an adjustable clip under their chin. Furthermore, for safety reasons, participants were surrounded by large crashmats in case of a fall. None of the participants experienced any injury or harm during the study.

G. Speech transcription

Each trial was recorded as a single stereo WAV file containing both speakers. Recordings were transcribed with WhisperX (Bain *et al.*, 2023; based on OpenAI Whisper; Radford *et al.*, 2023) using the large-v3 model. To preserve verbatim speech, we prompted the system to retain all spoken elements, including disfluencies, such as filler particles, false starts, and repetitions. Speakers were identified by comparing the root mean square (RMS) amplitude per word and mapped to clue-giver vs guesser based on the counterbalanced trial schedule. Outputs were saved as JavaScript Object Notation (JSON) files and converted to PRAAT TextGrids (Boersma and Weenink, 2024) with separate tiers per role. To prepare the annotations for acoustic analysis, adjacent word intervals separated by less than 200 ms were merged, whereas longer gaps were labeled as pauses. Three student assistants, blind to the study's hypotheses, manually corrected the automated transcripts using PRAAT (Boersma and Weenink, 2024) in two passes. Each assistant was assigned one-third of the dataset (∼20 files per coder). In the first pass, coders verified speaker identification (clue-giver vs guesser), corrected transcription errors, and annotated noise (e.g., overlapping speech, laughter, microphone artifacts) on a dedicated noise tier. In the second pass, coders refined the temporal boundaries of speech and pause segments to ensure high accuracy. To facilitate boundary correction, coders were instructed to display three acoustic parameters simultaneously: the spectrogram (dynamic range: 50 dB), pitch contour (f0 range: 80–350 Hz, filtered autocorrelation method), and intensity trace (range: 0–90 dB; averaging method: median; subtract mean pressure enabled). Coders typically zoomed to 1–3 s windows to optimally visualize acoustic features. Filler particles (e.g., "uh," "uhm," and "um") were transcribed using consistent orthography. Pauses were annotated as <pause> only when silent intervals exceeded 200 ms, excluding intervals at the beginning or end of recordings. Breathing sounds without clear speech acoustics (e.g., sighs) were coded as pauses rather than speech.

H. Data processing

For the acoustic variables, all data processing of the audio recordings was performed in PRAAT (Boersma and Weenink, 2024), R, and RSTUDIO (R Core Team, 2025). Given that the stereo audio files incorporated a single speaker for each channel, the stereo signal was divided into two distinct files, with each file containing one speaker from the interacting pair. To ensure the integrity of the data and prevent octave jumps, in a first step we determined the minimal and maximal f0 boundaries from all data, allowing an extreme range between 30 and 800 Hz. We determined the minimum and maximum boundary as the 0.05% and 99.5% quantiles from all f0 median data. These corresponded to a minimal value of 80 Hz and a maximal value of 335 Hz, which were then used as the minimum and maximum f0 range using the ksvF0 algorithm (Drugman *et al.*, 2012) implemented using the wrassp package (Winkelmann and Jänsch, 2015). The analysis window shift was set to match the target temporal resolution (5 ms, corresponding to 200 Hz). The resulting f0 contour was exported as a time series in milliseconds in a comma-separated values (CSV) file. In a subsequent analysis, f0 values were extracted from the respective speech segments from the clue-giver that did not overlap with segments on the guesser tier or noise tier to avoid acoustically disturbed signals. We averaged all f0 values within each clue-giver speech segment. For trials containing multiple speech segments, these segment-level means were further averaged to yield a single mean f0 per trial.

The amplitude envelope was computed using the analytic signal obtained via a Hilbert transformation of the waveform. The instantaneous amplitude (i.e., the magnitude of the signal) was then smoothed using a Hanning window with a bandwidth of 5 Hz (cf. MacIntyre *et al.*, 2022). Missing values introduced by smoothing were replaced with zeros. The resulting envelope was downsampled to 200 Hz by linear interpolation and uniform resampling across the full duration of the sound. Similarly to f0, we extracted all amplitude envelope data for a given speech segment of the clue-giver overlap from the guesser or noise and averaged them. If there were several speech segments, the amplitude envelope data were averaged to create a mean amplitude envelope.

For the fluency variables, all data processing of the TextGrids was performed in R and RSTUDIO (R Core Team, 2025) and PYTHON (Python Software Foundation, 2026). There are several ways to compute the speech rate. We decided on a data-driven approach that counts the peaks of the amplitude envelope, which are supposed to correspond to syllable nuclei, per duration (see Weston *et al.*, 2024 and references therein). The amplitude envelope was extracted for each trial. It was bandpass filtered in the 2–8 Hz range and smoothed using a Kolmogorov–Zurbenko filter with a window of 100 ms. Filtering and smoothing were necessary to focus on syllabic rhythm and reduce sub-syllabic fluctuations. Peaks in the smoothed envelope were detected using a threshold at the 20th percentile of the initial peak amplitudes. Speech rate was finally computed as the number of detected peaks divided by the duration of the respective speech interval. These values were aggregated across trials for each clue-giver and experimental condition. For pause duration and pause frequency, the <pause> mention in the TextGrids was extracted and summed per trial. We made sure to count any <pause> with guesser speech as a <turn> rather than a silent pause, and we made sure to exclude any <pause> with overlapping annotations on the noise tier as these could influence the pause duration. The same holds for pause duration, where all the durations of silent pauses were summed, while excluding turns and noise. A PYTHON script was used to extract occurrences of uhm, uh, and um from the transcribed clue-giver interval tier, using regular expressions to ensure that only filler particles were captured and not for example, parts of a longer word containing the same string of letters (e.g., "trumpet," "instrument"). These were then collapsed into a single column with the total filler particle frequency per trial.

For the two manipulation checks, wobble board motion was recorded using a gyroscope (WitMotion BWT901CL) attached to the underside of the wobble board. The gyroscope provided two metrics of balance in the unstable condition, namely angular velocity (AsX–AsZ) and tilt (AngleX–AngleZ).

I. Operationalization

Our variables of interest were operationalized in the following way. The f0 was averaged across each speech segment and across speech segments per trial, where a higher value indicated a higher mean f0. The data from male and female participants were pooled because previous research has shown that changes in f0 follow a similar pattern in both groups (Vorperian *et al.*, 2015). Thus, although male and female speakers differ in their absolute f0, the direction and magnitude of change are comparable. There were 2295 trials with a measure of mean f0.

Similarly, the amplitude envelope was averaged across each speech segment and across speech segments per trial, where a higher value indicated a higher mean amplitude envelope. There were 2305 trials with a measure of mean amplitude envelope.

Speech rate was calculated as the number of amplitude envelope peaks across each speech segment divided by the duration of the respective segment (in seconds), which was then averaged across speech segments per trial, with a higher value indicating a higher speech rate (faster speech). There were 2303 trials with a measure of mean speech rate.

Pause duration was calculated as the number of seconds of silent pauses, which was then averaged across pauses per trial, with a higher value indicating longer pausing (silence). A total of 2165 trials contained a measure of mean pause duration and were included in the model analysis.

Pause frequency was calculated as the number of silent pauses, summed into totals per trial, which a higher number indicating more frequent pausing. All 2335 trials contained a pause frequency.

Filler particle frequency was calculated as the number of times clue-givers said uh or uhm, and um summed into totals per trial, with a higher number indicating more frequent filler particles. All 2335 trials contained a filler particle frequency.

Two balance metrics for our manipulation checks were derived from the gyroscope recordings. Angular velocity RMS (°/s) was the RMS of AsX per trial; tilt SD (°) was the SD of unwrapped AngleX per trial. A total of 1153 usable unstable (board) trials contained these balance metrics.

J. Analysis approach

For all statistical analyses we used R and RSTUDIO (R Core Team, 2025). Visualizations were created with ggplot2 (Wickham, 2016). Models were formulated using lme4 (Bates *et al.*, 2015) for continuous outcomes (e.g., f0, amplitude envelope, speech rate, pause duration) and the glmmTMB package (Brooks *et al.*, 2017) for frequency outcomes (e.g., pause frequency, filler particle frequency), which were modelled with a negative binomial distribution. Each dependent variable (DV) was analyzed using a mixed-effects model with postural stability as a fixed effect and random intercepts for pair and target word, as well as by-participant (clue-giver) random intercepts and slopes for postural stability [DV $\sim$ postural_stability + (1 | pair) + (1 + postural_stability | clue_giver) + (1 | target_word)]. Following recommendations for a maximal random effects structure (Barr *et al.*, 2013), this structure accounts for the non-independence of dyadic interactions, varying difficulty levels across target words, and individual variation in the effect of postural stability. All models converged with the maximal random effects structure without including an optimizer.

Angular velocity and tilt were analyzed using a mixed-effects model with segment type (speaking vs non-speaking) as a fixed effect and random intercepts for pair and target word, as well as by-participant (clue-giver) random intercepts and slopes for speaking segment [DV $\sim$ segment_type + (1 | pair) + (1 + segment_type | clue_giver) + (1 | target_word)]. Non-speaking was restricted to pause intervals (silence). When the maximal random-effects structure did not converge, we reduced the random effects structure (cf. Barr *et al.*, 2013), and the final models that converged included random intercepts for clue-giver and target word.

III. RESULTS

A. Acoustic measures

1. f0

A linear mixed-effects model revealed a significant effect of postural stability on mean f0, $\beta = 3.34$, standard error (SE) = 0.67, $t\ (117) = 5.01$, $p < 0.001$ [Fig. 3(A)]. Mean f0 was higher in the unstable (board) condition (M = 198 Hz, SE = 3.86, 95% confidence interval (CI) [190, 206]) than in the stable (ground) condition (M = 195 Hz, SE = 3.85, 95% confidence interval (CI) [187, 202]). The estimated mean difference between conditions was small (3.34 Hz, 95% CI [2.03, 4.64]).

2. Amplitude envelope

A linear mixed-effects model revealed a significant effect of postural stability on mean amplitude envelope, $\beta = 0.000294$, SE = 0.000133, $t\ (117) = 2.20$, $p = 0.030$ [Fig. 3(B)]. Mean amplitude envelope was higher in the unstable (board) condition (M = 0.00468, SE = 0.00037, 95% CI [0.00395, 0.00541]) than in the stable (ground) condition (M = 0.00439, SE = 0.00032, 95% CI [0.00376, 0.00502]).

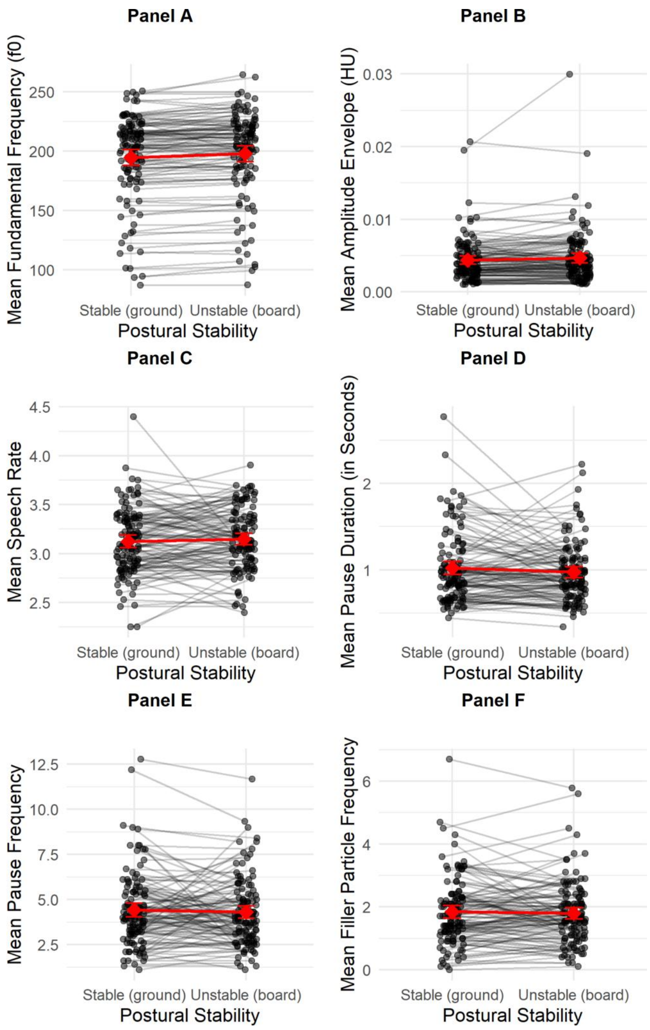

FIG. 3. Panel plot of two acoustic measures and four fluency measures ($N = 118$). Means for (A) f0, (B) amplitude envelope in Hilbert units, (C) speech rate, (D) pause duration, (E) pause frequency, and (F) filler particle frequency are on the y axis by postural stability on the x axis. Red diamonds represent means with red lines connecting means across postural stability conditions and error bars representing 95% confidence intervals. Grey dots represent means of individual participants with lines connecting means across conditions.

The estimated mean difference between conditions was small (0.000294, 95% CI [0.000032, 0.000555]). Mean amplitude envelope was approximately 0.56 dB higher in the unstable than in the stable condition (95% CI [0.06, 1.04] dB).

B. Fluency measures

1. Speech rate

A linear mixed-effects model indicated no significant effect of postural stability on mean speech rate, $\beta = 0.025$, SE = 0.025, $t\ (117) = 1.00$, $p = 0.320$ [Fig. 3(C)]. Mean speech rate was only slightly higher in the unstable (board) condition (M = 3.15, SE = 0.034, 95% CI [3.08, 3.22]) compared with the stable (ground) condition (M = 3.12, SE = 0.037, 95% CI [3.05, 3.20]). The estimated mean difference between conditions was small (0.025 syllables/s, 95% CI [–0.024, 0.075]).

2. Pause duration

A linear mixed-effects model revealed no significant effect of postural stability on mean pause duration, $\beta = -0.05$, SE = 0.03, $t\ (116) = -1.63$, $p = 0.106$ [Fig. 3(D)]. Mean pause duration (in seconds) was only slightly shorter in the unstable (board) condition (M = 0.97, SE = 0.038, 95% CI [0.89, 1.05]) than in the stable (ground) condition (M = 1.02, SE = 0.04, 95% CI [0.94, 1.11]). The estimated mean difference between conditions was small (0.052 s, 95% CI [−0.11, 0.01]).

3. Pause frequency

A negative binomial mixed-effects model revealed no significant effect of postural stability on pause frequency ($b = -0.03$, SE = 0.04, $z = -0.73$, $p = 0.468$) [Fig. 3(E)]. The effect size was small, with pause frequency in the unstable (board) condition estimated to be 2.6% lower than in the stable (ground) condition (rate ratio = 0.97, 95% CI [0.91, 1.05]). Estimated marginal means indicated that clue-givers' pause frequency in the unstable (board) condition (M = 3.78, SE = 0.28, 95% CI [3.26, 4.37]) was only slightly lower than in the stable (ground) condition (M = 3.88, SE = 0.29, 95% CI [3.35, 4.49]).

4. Filler particle frequency

A negative binomial mixed-effects model revealed no significant effect of postural stability on filler particle frequency ($b = -0.01$, SE = 0.04, $z = -0.14$, $p = 0.889$) [see Fig. 3(F)]. The effect size was negligible, with filler particle frequency in the unstable (board) condition estimated to be 0.5% lower than in the stable (ground) condition (rate ratio = 0.99, 95% CI [0.92, 1.07]). Estimated marginal means indicated that clue-givers' filler particle frequency in the unstable (board) condition (M = 1.52, SE = 0.12, 95% CI [1.30, 1.77]) was virtually identical to that of the stable (ground) condition (M = 1.53, SE = 0.12, 95% CI [1.30, 1.78]).

C. Manipulation checks

1. Angular velocity and tilt distributions across the unstable condition

Finally, we performed two manipulation checks using two balance metrics derived from our gyroscope recordings. First, we visualized the angular velocity and tilt of the wobble board (see histograms in Fig. 4). Two balance metrics were computed for each participant and averaged across the unstable (board) condition. Angular velocity RMS indexed the speed of mediolateral sway (M = 12.84°/s, SD = 3.73). Tilt SD captured the dispersion of postural sway (M = 4.20°, SD = 1.23). Both balance metrics were clearly distributed above the zero baselines, with angular velocity indicating there was continuous speed of movement and tilt, indicating the board was not in a flat or horizontal position. Thus, the first manipulation check confirmed that the wobble

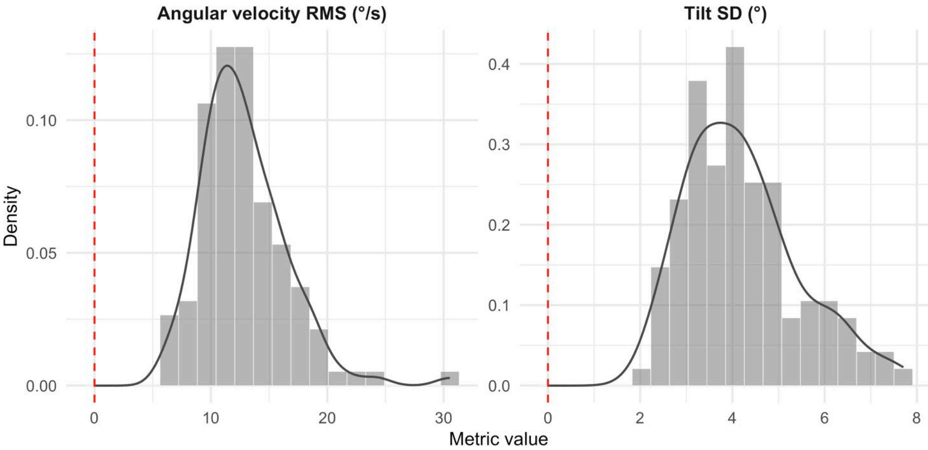

FIG. 4. Distribution of angular velocity (left panel) and tilt (right panel) metrics across clue-givers in the unstable (board) condition ($N = 117$). Each data point represents an individual's mean across all unstable trials. Histograms show the density distribution with a kernel density overlay (solid black line). The dashed red dashed vertical lines mark the zero-baselines representing the board in a perfectly still (left panel) and horizontal (right panel) position.

board increased postural instability demands across participants in the unstable condition.

2. Angular velocity and tilt by segment type in the unstable (board) condition

Second, we compared both balance metrics between speaking and non-speaking segments within the trials of the unstable (board) condition. This within-trial contrast isolates segments when clue-givers were speaking vs silent pauses, while controlling the same participant, target word, and wobble board context.

Two linear mixed-effects models revealed significant effects of segment type (speaking vs non-speaking) on both balance metrics (see Fig. 5): angular velocity RMS, $\beta = 0.90$, SE = 0.32, t (2070) = 2.82, $p = 0.005$; tilt SD, $\beta = 0.55$, SE = 0.09, t (2063) = 6.31, $p < 0.001$. In both cases, speaking intervals were associated with higher balance variability than non-speaking intervals. Estimated marginal means showed greater mediolateral sway during speaking than during non-speaking segments for angular velocity (speaking: M = 12.49, SE = 0.42, 95% CI [11.67, 13.31]; non-speaking: M = 11.59, SE = 0.42, 95% CI [10.76, 12.42]) and tilt (speaking: M = 3.87, SE = 0.13, 95% CI [3.61, 4.12]; non-speaking: M = 3.31, SE = 0.13, 95% CI [3.06, 3.57]). Thus, the second manipulation check confirmed that adults produced more pronounced changes in postural sway when speaking (compared to non-speaking) in our unstable (board) condition.

IV. GENERAL DISCUSSION

Perturbing postural stability selectively influenced the dynamics of the speech production system. Two key findings emerged. First, in line with our predictions, the acoustic properties of adult speech changed when standing on an unstable surface (i.e., a wobble board). Specifically, adults spoke with higher f0 (H1) and higher amplitude envelope (H2) compared to standing on a stable surface. Second, contrary to our predictions, the same increased postural stability demands did not disrupt the fluency of adult speech. Specifically, adults maintained consistent speech rate (H3), pause duration (H4), pause frequency (H5), and filler particle frequency (H6) regardless of whether they were standing on the ground or the wobble board. This pattern of results suggests that increased postural stability demands influenced the acoustic properties of speech, whereas the fluency of speech remained unaffected.

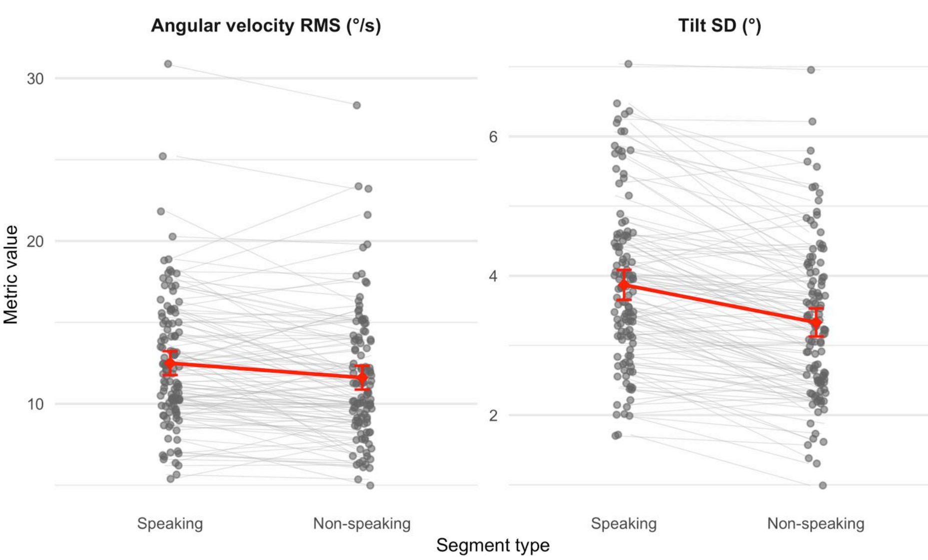

FIG. 5. Angular velocity (left panel) and tilt (right panel) metrics by segment type (speaking vs non-speaking) on the x axis ($N = 117$). Red diamonds represent means with red lines connecting means across segment type and error bars representing 95% confidence intervals. Grey dots represent means of individual participants with lines connecting means across segment types. Note that one participant is excluded from the manipulation checks due to missing gyroscope data.

The key finding that perturbing postural stability leads to acoustic differences in speech production is consistent with evidence that changes in body posture (e.g., upright vs supine) produce subtle, individualized changes in vocal output (Engström *et al.*, 2024; Stone *et al.*, 2007; Vorperian *et al.*, 2015) and that dynamic upper-limb movements co-vary with f0 and amplitude envelope, suggesting that gesture-related muscle activity may act as internal postural perturbations affecting the respiratory and vocal systems (Pouw *et al.*, 2025). In this light, the subtle increase in f0 and amplitude envelope observed during lower-limb balance perturbations in our study may reflect a broader postural adjustment to imbalance while standing on the wobble board, potentially including increased stiffening of the body. Such whole-body adjustments may involve subtle adjustments within the speech motor system, including changes in muscular effort and respiratory support, which could in turn influence speech acoustics. Although the magnitude of these mean acoustic shifts was small (mean increase of 0.56 dB, 95% CI [0.06, 1.04] dB, in the unstable relative to the stable condition) and likely below perceptual thresholds, they are consistent with the biomechanical modulation of vocal output under increased postural stability demands. Additionally, this explanation is not mutually exclusive with increased arousal, task engagement, or stress, when balancing on the wobble board, which may all jointly contribute to acoustic changes to speech production.

In contrast, the key finding that perturbing postural stability did not disrupt speech fluency is somewhat unexpected. We did not observe differences between balance conditions in speech rate, pause duration, pause frequency, or filler particle frequency, despite previous evidence that increased attentional or cognitive demands can affect speech fluency (e.g., Mitchell *et al.*, 1996; Oomen and Postma, 2001). Direct comparisons with previous studies are difficult because prior work has primarily examined static changes in body posture and gravitational context (Engström *et al.*, 2024; Stone *et al.*, 2007; Vorperian *et al.*, 2015) or experimentally imposed cognitive load effects on breathing and disfluency measures (Mitchell *et al.*, 1996), rather than dynamic postural instability during interactive speech production. One possibility is that speech planning processes are relatively robust to moderate postural perturbation, allowing speakers to maintain fluency despite increased postural stability demands. Alternatively, the wobble board manipulation may not have been sufficiently demanding to disrupt higher-level speech planning processes. In addition, our sample consisted of young, healthy adults who may have been less susceptible to fluency disruption under postural instability, potentially because speaking while moving is a common feature of everyday communication.

Two manipulation checks confirmed that our experimental manipulation indeed increased postural stability demands, even for a sample of young adults. First, gyroscope recordings confirmed that the wobble board was moving and tilting continuously for all participants. This was important to check because even though adult participants stood on the wobble board continuously during a trial, participants could reach brief periods of relative balance, suggesting that postural stability varied moment by moment rather than remaining uniformly unstable. When we subsequently compared speaking and non-speaking segments within the unstable (board) condition, adults showed increased postural sway during active speech than during silent pauses. This finding supports the view that speaking and maintaining upright posture draw on shared motor and attentional resources, consistent with prior work in young adults (Dault *et al.*, 2003) and older adults (Menant *et al.*, 2014).

Additionally, we manipulated postural stability via a wobble board, which primarily engaged the lower limbs while requiring whole-body stabilization. Participants controlled the board with their feet and legs during spontaneous interactive speech, so that perturbations originated in the lower body and were not temporally constrained by explicit movement instructions. This contrasts with upper-limb paradigms where participants produced sustained vowels while performing instructed arm movements. Pouw *et al.* (2025) found that instructed upper-limb movements produced force impulses that co-varied with vocal changes in the amplitude envelope, and that these vocal changes could be predicted from postural muscle activity. In our wobble board manipulation, peaks in the amplitude envelope tended to precede peaks in lower-limb gyroscope data. It is possible that during spontaneous speech, initiating speech-related movements (involving head, jaw, and respiratory motions) may perturb balance and elicit compensatory adjustments in the lower body. In contrast, in the instructed arm-movement paradigm, the perturbation is voluntarily initiated in the limbs, which may lead to anticipatory postural adjustments preceding vocal changes. Together, these differences suggest that the source of perturbation (upper vs lower body) and the nature of speaking task (sustained vowel vs spontaneous speech) may influence the coordination between posture and the vocal system in different ways.

A. Strengths and limitations

The current study has several strengths. First, the sample size was large relative to previous studies on posture influences on speech production (e.g., Pouw *et al.*, 2025; Vorperian *et al.*, 2015), and on suprapostural balance tasks (Mangalam *et al.*, 2025) yielding high statistical power. Second, our study used a face-to-face interaction task, where two participants collaborated in a word-guessing game and communicated freely throughout the task. This ecological validity of our approach offers insights into postural stability influences on speech production in an interactive face-to-face context compared to previous studies, which used tasks such as isolated vowel production, reading sentences, or operational speech recordings (e.g., Engström *et al.*, 2024; Pouw *et al.*, 2025; Vorperian *et al.*, 2015).

At the same time, the spontaneous nature of this interactive task necessitated a focus on global acoustic metrics, such as f0 and amplitude, rather than more detailed measures like F2 slopes or changes in the formant space of vowels. Unlike controlled reading tasks, the Taboo paradigm does not guarantee the production of specific phonemes across participants. This lack of phonetic control makes it difficult to reliably compare formant transitions or vowel spaces, which are highly sensitive to specific coarticulatory environments. Although our global acoustic measures captured significant effects, we acknowledge they may not have been sensitive to more subtle speech adjustments. Future studies using standardized speech stimuli—such as reading identical sentences across conditions—could further complement these findings by employing second formant (F2) slopes or formant space metrics to isolate these more granular articulatory changes.

Another limitation is that we have not yet analyzed gesture data, which was beyond the scope of this study. Co-speech gestures and other upper-body movements may also perturb postural stability in addition to the wobble board or, alternatively, help to stabilize or counteract imbalance during speaking. That is, when standing on an unstable surface, people may instinctively extend their arms to redistribute their weight and help maintain balance. Moreover, rhythmic upper-limb movements could modulate f0 and the amplitude envelope in monologue speech production within the gesture-speech model (Pouw *et al.*, 2021; Rusiewicz *et al.*, 2014). Our findings of acoustic changes thus likely reflect a combined influence of increased postural stability demands and spontaneous co-speech gestures, and currently we are unable to attribute acoustic differences uniquely to the manipulation of the lower limbs as we did not constrain gesture with the upper limbs.

Finally, the sample may not be fully representative of the general population. First, it was gender-imbalanced (95 females, 19 males, 2 non-binaries, 2 prefer not to disclose), which should be considered when interpreting the generalizability of the findings. However, given that all analyses focused on within-subject changes in f0, this imbalance is unlikely to have substantially affected the main results. Second, the sample consisted of healthy young adults, and the findings may therefore not generalize to older adults or to individuals with balance impairments or speech and motor disorders.

B. Directions for future research

Although we did not observe effects of postural instability on fluency measures, this does not rule out the influence of postural stability on the relationship between cognitive load and speaking. Maintaining balance under perturbation relies on domain-general resources such as attention and executive control, which are also required for language production. In the present analyses, however, we focused primarily on how adults spoke (e.g., acoustics and fluency), rather than on what they said. A more fine-grained linguistic analysis of the content of speech, for example, of lexical diversity (flexibility of word choice) and taboo word production (inhibitory control failures), may provide a more sensitive test of cognitive load effects. Such an approach would move beyond asking whether postural stability affects fluency *per se* and instead examines whether postural instability moderates the impact of cognitive load on the content and organization of spoken language, thereby increasing the demands of speaking.

Another potential avenue for future work could be looking into individual differences. It is well established in the literature on balance control that individuals adapt differently to postural perturbations (Massion, 1994; Taubert *et al.*, 2024). Although for some individuals it is not challenging to stand on a wobble board, for others it might be more demanding. Depending on how much individuals become unstable in their posture and how many resources they use in compensating for balance perturbations, may have an impact on cognitive load. When examining Figs. 3(D)–3(F), noticeable differences emerge between the stable and unstable conditions for several adult participants. In some cases, these differences even occur in opposite directions, which reduces the likelihood of a consistent effect at the population level. Rather than interpreting the disfluencies as reflecting an absence of an overall effect, it may be more appropriate to view them as revealing meaningful individual-level variation. Exploring how individuals adapt to perturbations and how these adaptation strategies relate to speech production and cognitive processes represents a promising direction for future research.

Furthermore, a notable feature of the behavioral data were that, in a small number of trials, participants produced little or no speech. Anecdotally, participants sometimes adopted a gesture-only strategy to convey the target word (i.e., charades-like)—for instance, miming taking a photograph when the target word was camera. These trials yielded almost no speech, sometimes only a brief sound mimicking a camera shutter, yet were often highly effective, with the guesser correctly identifying the word almost immediately. This suggests that participants flexibly adapted their communicative strategy to task demands. Because our measures focused on acoustic speech variables, we cannot determine the extent to which speech and gesture were co-produced in other trials; however, these gesture-only trials suggest that participants could, at least occasionally, fully substitute speech with gesture. This is consistent with the possibility that sufficient postural control on the wobble board allowed flexible allocation of attentional and motor resources to alternative communicative behaviors without disrupting higher-level planning processes. Future research could examine how postural perturbation affects gesture production and kinematics and how speakers communicate meanings across speech and gesture. Speech and gesture form a tightly integrated system (Goldin-Meadow, 2003; McNeill, 1992), but the relative contribution of each modality may vary depending on physical constraints. Under conditions of postural instability, speakers may therefore flexibly make use of speech and gesture to maintain efficient communication.

V. CONCLUSION

Postural stability is a fundamental aspect of human movement and balance, which has been largely overlooked in models of speech motor control (Liu *et al.*, 2024). Most existing models focus on gestures or motor commands for specific speech movements, without considering how maintaining an upright posture may influence speech. This study informs speech production models by showing that perturbing postural stability alters the acoustic properties of speech. Specifically, increased postural demands led to higher f0 and higher amplitude envelopes of adult speech during a word-guessing game. These differences likely reflect a combination of motor control, respiratory adjustments, and broader biomechanical processes, highlighting that speech production engages the whole body. Accordingly, models of speech production should incorporate postural stability, especially as some motor resources are shared between balancing and speaking.

ACKNOWLEDGMENTS

M.L. and J.E.Y. contributed equally to this work. This work was supported by The British Academy/Leverhulme Trust (Grant No. SRG2324\240765) awarded to S.A. We would like to thank all adult participants who took part in this study as part of their undergraduate research experience. We thank Ekam Gidda and Zoha Rehman for their assistance with piloting and creating the walkthrough video of our experiment procedure as part of their undergraduate placement year. Finally, we would like to thank Eliza Rafiq, Hannah Lowe, and Shreshta Mannava, who manually corrected all speech transcripts as part of their undergraduate research placement year.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Ethics Approval

The study received full ethical approval from the Humanities & Social Sciences Research Ethics Committee at the University of Warwick (HSSREC 28/24–25) in the United Kingdom. Informed consent was obtained from all participants prior to commencing the study.

DATA AVAILABILITY

The data that support the findings of this study are openly available via the Open Science Framework at https://osf.io/68uxv/. An audiovisual corpus with data from 60 participants (30 pairs), who consented to making their data publicly available online, can be accessed via Zenodo (Yang *et al.*, 2026, https://doi.org/10.5281/zenodo.19853626).

APPENDIX

List of target words and conceptually related taboo words (Table I).

TABLE I. List of target words and conceptually related taboo words.

Target word[a]	Taboo 1	Taboo 2	Taboo 3	Taboo 4	Taboo 5
walnut	fruit	shell	hard	brazil	eat
spinach	vegetable	eat	green	leafy	popeye
music	sound	dance	record	song	notes
traffic	road	sign	stop	ticket	rush
birthday	cake	celebration	age	born	party
sushi	rice	eat	roll	fish	raw
zipper	clothes	close	pants	teeth	metal
bacon	pork	pig	egg	side	ham
gravy	sauce	brown	potato	meat	mashed
christmas	holiday	december	santa	winter	jesus
pasta	macaroni	noodle	sauce	egg	twist
river	water	stream	flow	big	bridge
soda	drink	beverage	pop	carbonated	coke
exam	test	bar	fail	pass	mark
balloon	helium	dirigible	round	air	blow
hamster	pet	gerbil	rat	mouse	guinea pig
jungle	forest	jeep	gorilla	leopard	camp
puppy	dog	baby	small	young	four
leather	wallet	material	shoe	cow	suede
hunger	stomach	food	feeling	famished	need
cologne	perfume	male	smell	scent	stuff
chlorine	pool	chemical	disinfectant	gas	cleaning
sugar	sweet	white	powder	cane	salt
tiger	animal	cat	striped	lion	big
muscle	body	strength	tissue	strong	flesh
flower	plant	pretty	rose	petals	bloom
siren	noise	police	loud	cars	alarm
bitter	taste	sour	flavor	bad	acrid
flannel	fabric	shirt	material	soft	cotton
movie	film	cinema	entertainment	moving	actors
pillow	head	bed	cushion	sleeping	rest
trumpet	instrument	horn	brass	musical instrument	musical
beauty	pretty	beast	looks	prettiness	good
blanket	bed	sleeping	cover	warm	keeping
coffee	latte	drink	morning	beans	caffeine
doughnut	hole	bagel	round	food	pastry
butter	bread	margarine	yellow	dairy	spread
cherry	fruit	red	small	sundae	pit
picture	photo	image	photograph	frame	camera
yellow	color	sun	banana	primary	daffodils

[a]The first column lists the target words, whereas the other columns list five conceptually related Taboo words.

Bain, M., Huh, J., Han, T., and Zisserman, A. (**2023**). "WhisperX: Time-accurate speech transcription of long-form audio," arXiv:2303.00747.
Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (**2013**). "Random effects structure for confirmatory hypothesis testing: Keep it maximal," J. Mem. Lang. **68**, 255–278.
Bates, D., Mächler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4," J. Stat. Software **67**, 1–48.
Boersma, P., and Weenink, D. (**2024**). "Praat: Doing phonetics by computer (version 6.4.59) [computer program]," http://www.praat.org/ (Last viewed May 27, 2026).
Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Mächler, M., and Bolker, B. M. (**2017**). "glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling," R J. **9**, 378–400.

Cardoso, R., Lumini-Oliveira, J., and Meneses, R. F. (**2019**). "Associations between posture, voice, and dysphonia: A systematic review," J. Voice **33**, 124.e1–124.e12.
Corley, M., and Stewart, O. W. (**2008**). "Hesitation disfluencies in spontaneous speech: The meaning of um," Lang. Linguist. Compass **2**, 589–602.
Dault, M. C., Geurts, A. C. H., Mulder, T. W., and Duysens, J. (**2001**). "Postural control and cognitive task performance in healthy participants while balancing on different support-surface configurations," Gait Posture **14**, 248–255.
Dault, M. C., Yardley, L., and Frank, J. S. (**2003**). "Does articulation contribute to modifications of postural control during dual-task paradigms?," Cognit. Brain Res. **16**, 434–440.
Drugman, T., Bozkurt, B., and Dutoit, T. (**2012**). "A comparative study of pitch extraction algorithms on a large variety of speech data," https://arxiv.org/pdf/1912.12609 (Last viewed May 27, 2026).
Engström, H., Włodarczak, M., and Ternström, S. (**2024**). "Mapping the effect of body position: Voice quality differences in connected speech," in *Proceedings of FONETIK 2024*, Stockholm, Sweden (June 3–5), pp. 21–26.
Ferreira, F., and Bailey, K. G. D. (**2004**). "Disfluencies and human language comprehension," Trends Cognit. Sci. **8**, 231–237.
Goldin-Meadow, S. (**2003**). *Hearing Gesture: How Our Hands Help Us Think* (Harvard University Press, Cambridge, MA).
Juhasz, B. J., and Yap, M. J. (**2013**). "Sensory experience ratings for over 5,000 mono- and disyllabic words," Behav. Res. Methods **45**, 160–168.
Kerr, B., Condon, S. M., and McDonald, L. A. (**1985**). "Cognitive spatial processing and the regulation of posture," J. Exp. Psychol. Hum. Percept. Perform. **11**, 617–622.
Li, M., Yang, J., Fuchs, S., and Aussems, S. (**2026**). "Files to support: Perturbing postural stability affects acoustic properties but not fluency of adult speech production during face-to-face interaction," [dataset]. Open Science Framework. https://osf.io/68uxv.
Liu, Y., Fels, S., Shamei, A., Khan, N., and Gick, B. (**2024**). "Integrating posture control in speech motor models: A parallel-structured simulation approach," arXiv:2407.18516.
Lively, S. E., Pisoni, D. B., Van Summers, W., and Bernacki, R. H. (**1993**). "Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences," J. Acoust. Soc. Am. **93**, 2962–2973.
Maassen, B., Kent, R., Peters, H., van Lieshout, P., and Hulstijn, W. (**2007**). *Speech Motor Control: In Normal and Disordered Speech* (Oxford University Press, Oxford, UK).
MacIntyre, A. D., Cai, C. Q., and Scott, S. K. (**2022**). "Pushing the envelope: Evaluating speech rhythm with different envelope extraction techniques," J. Acoust. Soc. Am. **151**(3), 2002–2026.
Mangalam, M., Barfi, M., Deligiannis, T., and Schlattmann, B. (**2025**). "A multimodal biomechanical and eye-tracking dataset of suprapostural coordination in health young adults," Sci. Data **12**, 1311.
Massion, J. (**1994**). "Postural control system," Curr. Opin. Neurobiol. **4**, 877–887.
McNeill, D. (**1992**). *Hand and Mind: What Gestures Reveal about Thought* (University of Chicago Press, Chicago, IL).
Menant, J. C., Schoene, D., Sarofim, M., and Lord, S. R. (**2014**). "Single and dual task tests of gait speed are equivalent in the prediction of falls in older people: A systematic review and meta-analysis," Ageing Res. Rev. **16**, 83–104.
Mitchell, H. L., Hoit, J. D., and Watson, P. J. (**1996**). "Cognitive-linguistic demands and speech breathing," J. Speech Lang. Hear. Res. **39**, 93–104.
Oomen, C. C. E., and Postma, A. (**2001**). "Effects of divided attention on the production of filled pauses and repetitions," J. Speech Lang. Hear. Res. **44**, 997–1004.
Perrier, P. (**2006**). "About speech motor control complexity," in *Speech Production: Models, Phonetic Processes, and Techniques*, edited by J. Harrington and M. Tabain (Psychology Press, London, UK), pp. 13–26.
Pouw, W., de Jonge-Hoekstra, L., Harrison, S. J., Paxton, A., and Dixon, J. A. (**2021**). "Gesture-speech physics in fluent speech and rhythmic upper limb movements," Ann. N.Y. Acad. Sci **1491**(1), 89–105.
Pouw, W., Werner, R., Burchardt, L. S., and Selen, L. P. J. (**2025**). "The human voice aligns with whole-body kinetics," Proc. Biol. Sci. **292** (2047), 20250160.
Python Software Foundation (**2026**). "Python (version 3.13) [computer program]," https://www.python.org/ (Last viewed May 27, 2026).
Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., and Sutskever, I. (**2023**). "Robust speech recognition via large-scale weak supervision," in *Proceedings of the 40th International Conference on Machine Learning (ICML 2023)*, Honolulu, HI (July 23–29) (PMLR), Vol. 202, pp. 28492–28518.
R Core Team (**2025**). "r: A language and environment for statistical computing (version 4.5.2) [computer program]," https://www.R-project.org (Last viewed May 27, 2026).
Rusiewicz, H. L., Shaiman, S., Iverson, J. M., and Szuminsky, N. (**2014**). "Effects of perturbation and prosody on the coordination of speech and gesture," Speech Commun. **57**, 283–300.
Shamei, A. (**2024**). "Speech postures are postures: Towards a unified approach to postural control in gross and fine motor skills," Ph.D. dissertation, University of British Columbia, Vancouver, Canada.
Speer, R., Chin, J., and Havasi, C. (**2017**). "ConceptNet 5.5: An open multilingual graph of general knowledge," in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 31, pp. 4444–4451.
Stone, M., Stock, G., Bunin, K., Kumar, K., Epstein, M., Kambhamettu, C., Li, M., Parthasarathy, V., and Prince, J. (**2007**). "Comparison of speech production in upright and supine position," J. Acoust. Soc. Am. **122**, 532–541.
Taubert, M., Ziegler, G., and Lehmann, N. (**2024**). "Higher surface folding of the human premotor cortex is associated with better long-term learning capability," Commun. Biol. **7**, 635.
von Ahn, L., Kedia, M., and Blum, M. (**2006**). "Verbosity: A game for collecting common-sense facts," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Montréal, Québec, Canada (April 22–27) (ACM Digital Library, New York), pp. 75–78.
Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., Gentry, L. R., and Yandell, B. S. (**2015**). "Effect of body position on vocal tract acoustics: Acoustic pharyngometry and vowel formants," J. Acoust. Soc. Am. **138**, 833–845.
Weston, H., Pouw, W., and Fuchs, S. (**2024**). "On the relation between leg motion rate and speech tempo during submaximal cycling exercise," J. Speech Lang. Hear. Res. **67**, 3931–3946.
Wickham, H. (**2016**). "ggplot2: Create elegant data visualisations using the grammar of graphics (version 3.4.0) [computer program]," https://CRAN.R-project.org/package=ggplot2 (Last viewed May 27, 2026).
Winkelmann, R., and Jänsch, K. (**2015**). "wrassp: Interface to the ASSP library (version 1.0.6) [computer program]," https://CRAN.R-project.org/package=wrassp (Last viewed May 27, 2026).
Winter, D. A., Patla, A. E., Ishac, M., and Gage, W. H. (**2003**). "Motor mechanisms of balance during quiet standing," J. Electromyogr. Kinesiol. **13**, 49–56.
Woollacott, M., and Shumway-Cook, A. (**2002**). "Attention and the control of posture and gait: A review of an emerging area of research," Gait Posture **16**, 1–14.
Yang, J. E., Li, M., Fuchs, S., and Aussems, S. (**2026**). "An audiovisual corpus of adult-adult interactions while playing a Taboo word-guessing game under postural stability constraints," [dataset]. Zenodo. https://doi.org/10.5281/zenodo.19853626
Yardley, L., Gardner, M., Bronstein, A., Davies, R., Buckwell, D., and Luxon, L. (**2001**). "Interference between postural control and mental task performance in patients with vestibular disorder and healthy controls," J. Neurol. Neurosurg. Psychiatry **71**, 48–52.