TonguePlay: Non-invasive Tongue Pressure Input for Gaming

POORIA GHAVAMIAN, KTH Royal Institute of Technology, Sweden
NATALIA SEMPERE, KTH Royal Institute of Technology, Sweden
ASREEN ROSTAMI, RISE Research Institutes of Sweden, Sweden and Stockholm University, Sweden
MARTIN HEDLUND, KTH Royal Institute of Technology, Sweden
XIAOYAN ZHOU, KTH Royal Institute of Technology, Sweden
OSKAR HOKKANEN ERIKSSON, KTH Royal Institute of Technology, Sweden
FLORIAN MÜLLER, TU Darmstadt, Germany
ANDRII MATVIIENKO, KTH Royal Institute of Technology, Sweden

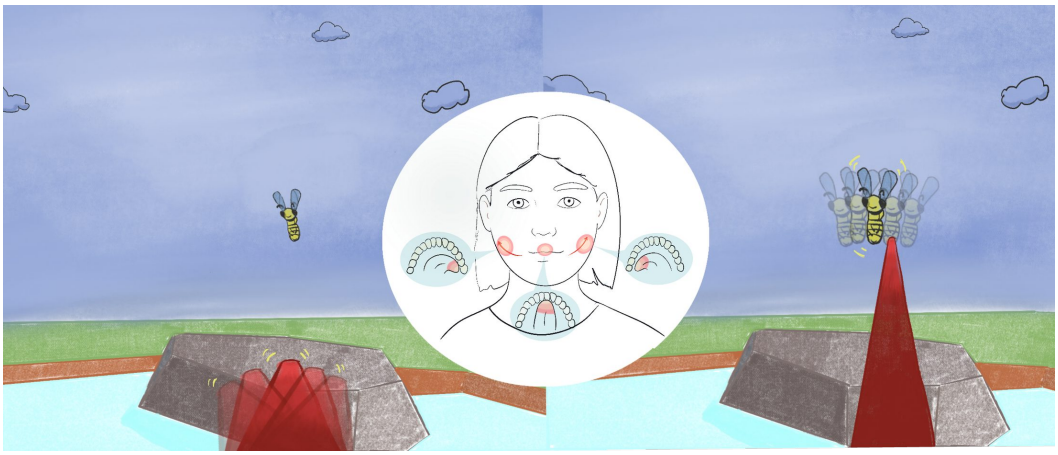


Fig. 1. (Center) A face with highlighted tongue contact areas: left (upper-left molar area), front (front palate), and right (upper-right molar area). Concentric rings indicate increasing pressure levels (low, medium, high) produced by pressing the tongue against the corresponding palate area with increasing force. Any of these three locations can be used for the same in-game action, giving the user flexibility. (Left) Study 2's frog game, egocentric condition: residual variability applied to the frog's tongue controlled by the player. (Right) Study 2's frog game, exocentric condition: residual variability applied to the target wasp.

Physiological input enables hands-free, embodied interaction in games, but its inherent signal variability complicates precise control. Most systems suppress this variability to achieve stability, leaving fine-grained

Authors' Contact Information: Pooria Ghavamian, KTH Royal Institute of Technology, Stockholm, Sweden, pooriag@kth.se; Natalia Sempere, KTH Royal Institute of Technology, Stockholm, Sweden, sempere@kth.se; Asreen Rostami, RISE Research Institutes of Sweden, Stockholm, Sweden and Stockholm University, Stockholm, Sweden, asreen.rostami@ri.se; Martin Hedlund, MID, KTH Royal Institute of Technology, Stockholm, Sweden, marthed@kth.se; Xiaoyan Zhou, DIVISION OF MEDIA TECHNOLOGY AND INTERACTION DESIGN, KTH Royal Institute of Technology, Stockholm, Sweden, xizhou@kth.se; Oskar Hokkanen Eriksson, KTH Royal Institute of Technology, Stockholm, Sweden, oskhe@kth.se; Florian Müller, TU Darmstadt, Darmstadt, Germany, florian.mueller@tu-darmstadt.de; Andrii Matviienko, KTH Royal Institute of Technology, Stockholm, Sweden, matviienko.andrii@gmail.com.

This work is licensed under a Creative Commons Attribution 4.0 International License.
© 2026 Copyright held by the owner/author(s).
ACM 2573-0142/2026/11-ARTGAMES073
https://doi.org/10.1145/3831314

variability of user effort underutilized. In this paper, we propose a dual-stream architecture that uses physiological signals for action control and residual effort-coupled variability as perceptual feedback to shape game feel. We investigate this approach using tongue-based input sensed via non-invasive surface electromyography (sEMG) in two studies. In Study 1 (n = 24), we evaluated the feasibility of tongue pressure in primary and modifier roles. Integrating head-gaze for selection with tongue pressure for confirmation significantly outperformed tongue-only input on performance measures, suggesting that separating spatial intent from physical effort can yield more stable control. In Study 2 (n = 12), we examined the experiential impact of residual variability in tongue pressure control as feedback. Rendering this residual jitter as a perceptual response led participants to misattribute their own effort to external resistance or animacy, a phenomenon we describe as physiological ventriloquism, where one's effort feels externally owned. We contribute a design perspective that reframes physiological variability not as error to be eliminated, but as expressive material when architecturally separated from control.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**.

Additional Key Words and Phrases: tongue-based interaction, surface electromyography (sEMG), physiological variability, game feel, extended reality

ACM Reference Format:
Pooria Ghavamian, Natalia Sempere, Asreen Rostami, Martin Hedlund, Xiaoyan Zhou, Oskar Hokkanen Eriksson, Florian Müller, and Andrii Matviienko. 2026. TonguePlay: Non-invasive Tongue Pressure Input for Gaming. *Proc. ACM Hum.-Comput. Interact.* 10, 7, Article GAMES073 (November 2026), 29 pages. https://doi.org/10.1145/3831314

1 INTRODUCTION

As gaming expands into Extended Reality (XR) and mobile contexts, conventional manual input is often unavailable or socially inappropriate [36, 51, 72]. Physiological signals, such as muscle activity, offer a promising hands-free alternative [17]. More importantly, these body-centric inputs provide not just a binary trigger but also a rich stream of data on tremor, exertion, and muscle-activation variability that reflects the player's physical state [50]. However, current biofeedback-driven games typically view this input through a dichotomy between direct voluntary control and indirect involuntary control, often discarding this nuance. Signals are either heavily smoothed to serve as a reliable input [17] for game action or averaged over time to passively modulate game context, such as difficulty [62]. In both cases, the continuous variability of the living body is treated as noise to be suppressed.

Yet at the same time, games routinely simulate this very "noise" to add physicality to in-game actions and enhance game feel [34, 60, 71]. Currently, this "juice" is simulated procedurally [60]: charge-up animations, camera shake, and reactive screen effects create a sense of effort and responsiveness [71]. We propose an alternative way to generate these effects using the player's own physiological variability. However, bringing this physical expressivity to mobile contexts requires an input modality that is socially discreet yet capable of expressive force. While eye gaze is fast for targeting, it lacks an expressive confirmation mechanism; similarly, voice is expressive but socially intrusive. The tongue, however, offers a unique solution: it is fatigue-resistant, highly dexterous, and capable of applying variable pressure while remaining entirely hidden.

In this paper, we explore how players' own physiological variability could drive game feel. We investigate non-invasive tongue pressure as a novel input channel. Instead of eliminating the pressure signal's residual variability, defined as the physiological variability remaining after filtering electrical interference and noise, we ask if this data can be architecturally separated. By splitting the signal, we aim to provide stabilized control of the input while exposing the raw, effort-coupled variability for feedback, rendering exertion not as a control state but as sensation (e.g., the tremor of a strained muscle visualized as a shaking game object rather than cursor jitter). This paper therefore

asks two interrelated questions: (1) Can non-invasive tongue pressure serve as a viable hands-free input in XR, and which roles assignments and modality pairings best support stable, usable control? (2) When residual physiological variability is architecturally separated from that control signal and rendered as game feedback, how do its temporal context and spatial locus shape player experience and agency attribution? To explore this, we conducted a two-study investigation using a novel dual-stream architecture. Study 1 ($n = 24$) establishes the first viable configuration for tongue pressure interaction in XR. Results confirm that while combining tongue position and pressure for simultaneous control is fatiguing, pressure functions well as a discrete modifier for head-gaze selection. Building on this validated configuration, Study 2 ($n = 12$) then examines the experiential impact of residual signal across two games. We show that when residual variability is decoupled from control and rendered as game feel, it reshapes the player's relationship to the virtual world. This feedback makes otherwise invisible effort perceptible and reveals graded controllability during novel motor acquisition, shifting interaction from whole-body effort to more isolated modulation of tongue pressure. It also gives rise to a phenomenon we term *physiological ventriloquism* (defined in Section 4), in which the player's own muscular exertion is misattributed to the external animacy or resistance of the game environment.

Our contributions are:

- A novel method for stabilizing physiological input during action control while using residual effort-coupled variability to shape perceptual feedback in real time.
- An empirical evaluation of tongue input performance and user experience, clarifying the complementary roles of tongue position and pressure for hands-free interaction in VR.
- Qualitative insights into how the locus of physiological feedback shifts player interpretation and perception.

2 RELATED WORK

This section situates our work at the intersection of embodied game input, physiological interaction, and tongue-based interfaces. We first examine how XR and mobile games using embodied input negotiate a trade-off between expressive, skill-based interaction and the physical, social, or attentional costs of overt movement. We then review physiological input as a potential means to address this tension, noting that its inherent variability is typically suppressed to enable stable control and interpretability. Finally, we survey tongue-based interfaces, showing that prior work largely treats the tongue as a discrete positional effector, while non-invasive tongue pressure (and its use as a semantically meaningful input dimension) remains underexplored. Together, this literature frames our investigation of tongue pressure as an effort-sensitive input whose variability can be leveraged rather than eliminated.

2.1 Embodied Input and its limitations in Games

Games have long leveraged the body as an expressive resource, with tight coupling between bodily action and in-game response supporting agency, engagement, and skill expression [57, 60, 71]. Extended Reality (XR) systems further intensify this coupling through immersion, mobility, and multimodal interaction [4, 35, 58]. However, common inputs often expose only a narrow slice of motor capability. Spatially tracked controllers abstract movement into handheld proxies, leading to fatigue or tracking instability [58, 73]. While vision-based tracking removes hardware, it introduces "ghost appendages" (virtual limbs that lag or detach from physical movement) and "gorilla arm" fatigue (muscle strain from holding arms unsupported in mid-air) [51, 76]. Hands-free alternatives address some of these physical costs but introduce unique trade-offs. Gaze enables rapid targeting yet, on its own, suffers from the Midas-touch problem (the unintended activation of targets by

looking at them) and limited confirmation mechanisms [2, 29, 49]. Speech affords rich expression but is socially intrusive in shared or public play contexts [12, 72]. Facial and oral inputs offer accessibility and hands-free benefits but can be visually conspicuous, cognitively demanding, or prone to false activation [33, 40, 69, 72, 78]. A recurring tension exists: expressive modalities demand overt effort, while discreet modalities often lack reliable confirmation mechanisms [18, 21]. This motivates exploring channels that support expressive, skill-based play without relying on overt movement or sustained visual attention.

2.2 Physiological Input in Games: Control Versus Expressivity

Physiological signals enable hands-free interaction without overt motion [54]. In games, these signals are conventionally divided into those that players can intentionally control (direct input), typically mapped to game actions, and those that they cannot (indirect input), such as heart rate, which are instead mapped to game context like difficulty [52, 62]. In both cases, fine-grained variability is typically suppressed or averaged away. Signals used for action prioritize stability and predictability, and so their continuous variability is suppressed through smoothing, windowing, thresholding, and conservative activation criteria for reliable actuation [17]. Conversely, in indirect modulation, the signal is averaged away to support interpretability. As a result, a design space remains underexplored in which physiological input is responsive enough to support moment-to-moment play while preserving bounded variability that may meaningfully shape player experience.

Importantly, games already employ variability to communicate game mechanics and enhance game feel. Within game feel design, designers routinely employ bounded variability to communicate effort, anticipation, and impact without altering the underlying game state [60, 71]. Prior to action resolution, anticipatory feedback such as charge-up animations or vibration in "spring-locked mode" (where sustained input builds tension that resolves when released, similar to drawing a bow) conveys preparation and tension [60, 67]. During or after an action, reactive feedback amplifies consequence through exaggerated audiovisual effects such as hit-stop, screen shake, or sound modulation [34, 60, 67]. These effects rely on procedural variability to shape perception rather than control outcomes. The spatial application of this feedback also shapes players' attribution and interpretation. While "Game Feel" as described by Swink [71] focuses on a positive feeling of control, we view game feel in a broader sense. Feedback applied to the environment can change the context of an action [62]. We leverage this distinction to explore how the spatial assignment of muscle tremor, whether applied to the player's body or the game world, modulates the interpretation of physiological effort.

By combining inherent physiological variability with procedural game feel techniques, we propose a seamful game design approach [10, 19] where the "seams" of a system (signal variability in this study) are made visible and exploitable rather than hidden away. Instead of suppressing variability of effort as noise, we preserve both the intentional, filtered signal and its residual variability, using each for distinct purposes. In this framing, physiological variability functions analogously to game feel effects. They do not interfere with control, but instead manifest as perceptible tension, instability, or anticipation. This grounds game feel in the player's physical state, transforming organic variability from control noise into an experiential resource for play.

2.3 Tongue Input and the Pressure Design Space

The tongue is a compelling modality due to its high neural density, fatigue resistance, and ability to be used discreetly with a closed mouth [11, 12, 20, 53, 65]. Despite this, most interfaces treat the tongue as a discrete positional effector for symbolic gestures and taps to commands [15, 21, 28, 46, 55, 65, 68, 77]. This emphasis collapses the tongue's rich control space into categorical input, limiting its expressive capacity in play contexts. To clarify how prior work constrains expressivity,

we analyze tongue interfaces along three dimensions: Sensorial Dimension (Position vs. Pressure), Resolution (Discrete vs. Continuous), and Input Role (Primary vs. Modifier). Table 1 summarizes this design space.

Sensorial Dimension. *Position* captures the spatial location of tongue movement (e.g., left, right, front) and is associated with "where" to act. *Pressure* captures muscular force exerted by the tongue and is associated with "how much" to act. In motor control theory, these map to orthogonal control streams: isotonic devices support positional control (0^{th}-order), while isometric devices support force or rate modulation (1^{st}-order) [61, 75].

Resolution. Input can be *discrete* (mapped to a fixed set of values, e.g., three tongue positions or pressure levels) or *continuous* (varying smoothly along a range, e.g., analog force or trajectory). Resolution determines interaction expressivity and control fidelity.

Input Role. A *primary* stream controls the main action (e.g., selection or pointing). A *modifier* stream adjusts or confirms another input (e.g., a tongue press to confirm a target). For example, head-gaze (head-directed movement) acts as the primary channel for spatial targeting while tongue pressure acts as a modifier for confirmation. A combination that directly addresses the Midas-touch problem described above, where gaze alone cannot distinguish intent from look.

Table 1. Tongue input design space by Position (rows) and Pressure (columns). **Cells show counts (k) of systems, and one exemplar.** † intra-oral, * extra-oral. **Full enumeration in Table 6 (Appendix).**

Pressure → ***Position ↓***	**None**	**Discrete**	**Continuous**
None	—	—	k=1 (TonguePoint† [63])
Discrete	k=17 (On the Tip of My Tongue* [12])	**This Paper***	—
Continuous	k=3 (TDS*† [38])	—	k=2 (MouthMouse† [46])

Across prior systems, tongue position is overwhelmingly favored, while pressure is often ignored, reduced to velocity control, or used only to infer position. To distinguish sensing capability from interaction design, we adopt a **Sensor–Semantics Criterion**: a system is categorized as involving Position or Pressure only if it (1) senses magnitude directly or via robust inference, and (2) treats variations in that magnitude as semantically meaningful rather than symbolic proxies. To illustrate how this criterion works in practice, consider the two contrasting systems from Table 1. *On the Tip of My Tongue* [12] uses resistive textile sensors on the outer cheek to detect tongue contact patterns. Although the sensors capture continuous pressure, the pressure signal is used purely as a proxy for spatial gesture shape (swipe left, right, etc.) and the pressure magnitude itself carries no interaction-level meaning. It therefore satisfies the *position* criterion (discrete) but not the *pressure* criterion. In contrast, *MouthMouse* [46] uses force-sensitive resistors that measure both the direction and continuous magnitude of tongue force. Cursor speed scales directly with applied force, making pressure a semantically meaningful control stream. It satisfies both criteria in continuous form, with position as primary and pressure as modifier.

Applying this criterion reveals a neglected quadrant: Discrete Position + Discrete Pressure. Prior work on pressure widgets suggests users can reliably distinguish several discrete pressure levels [61], making discrete pressure a promising channel for restoring expressivity without the

cognitive load of fully continuous control. By capturing the magnitude of the continuous signal and discretizing it for action while preserving residual variability for feedback, this quadrant offers a path to expressive, effort-coupled game input.

2.4 Positioning Our Work

Applying our framework to existing systems reveals concentrated design patterns and absences. Most systems fall within the *Discrete Position, No Pressure* quadrant, emphasizing symbolic gesture recognition (directional swipes, contact triggers) over expressive or continuous control. Among the few systems that satisfy the pressure criterion, all require intraoral hardware. The *Discrete Position + Discrete Pressure* quadrant therefore remains unexplored, and no non-invasive system treats pressure as a semantically meaningful channel in its own right.

Among hands-free modalities reviewed above, tongue pressure occupies a unique ecological niche. Gaze offers rapid targeting but lacks expressive confirmation; voice is rich but socially intrusive; facial gestures and breath can be conspicuous or prone to false activation. The tongue combines properties that none of these alternatives share simultanesouly: high dexterity and force expressivity, social discretion, and fatigue resistance suited for extended sessions [20, 53]. The system is therefore well-suited for two primary user groups: (1) XR and mobile users in social or public contexts, and (2) individuals with upper-limb motor impairment for whom the tongue offers a reliable alternative effector, remaining functional even in high-level spinal cord injury [30].

Building on the body of work outlined above, this paper explores non-invasive tongue pressure input for play and how its effort-coupled variability is interpreted. In Study 1, we evaluate the underexplored *Discrete Position + Discrete Pressure* (Table 1) design space and investigate how tongue pressure functions for selection, confirmation, and drag-and-drop tasks. In Study 2, we use the most viable configuration found in Study 1 across two game contexts to investigate how mapping physiological variability to game feedback, either on the player's avatar or the environment shapes player experience and agency attribution. This framing allows us to examine tongue pressure not only as an input mechanism, but as a vehicle for grounding game feel in the player's own physical effort. We describe both studies in detail in the following.

3 STUDY 1: TONGUE INPUT FEASIBILITY

Study 1 establishes whether non-invasive tongue sEMG can function as a reliable interactive input in XR, and which role allocations and modality pairings are viable for interaction.

3.1 Participants

Twenty-four participants (15 female, 9 male) aged 22–59 years ($M = 29.7$, $SD = 9.46$) completed the study. One additional participant was excluded due to corrupted data. No participant had prior experience with tongue-based input or surface EMG interfaces. The study followed institutional ethical guidelines. Because the protocol involved non-invasive surface EMG and posed minimal risk, formal ethics-board approval was not required. All participants provided informed consent, received a cinema ticket as compensation, and all data were pseudonymized.

3.2 Study Design

Study 1 examines the feasibility of using non-invasive tongue sEMG for real-time interaction in XR. Each session followed a fixed structure: a model training phase (pre-study, approx. 10-15 min), followed by three progressive experimental blocks (Selection, Integration, and Drag-and-Drop), each containing two VR tasks, and a final questionnaire completed on a laptop at the end of the session. The study evaluates tongue input across two functional roles and modality pairings. The experimental design follows a structure of increasing complexity: moving from isolated control of

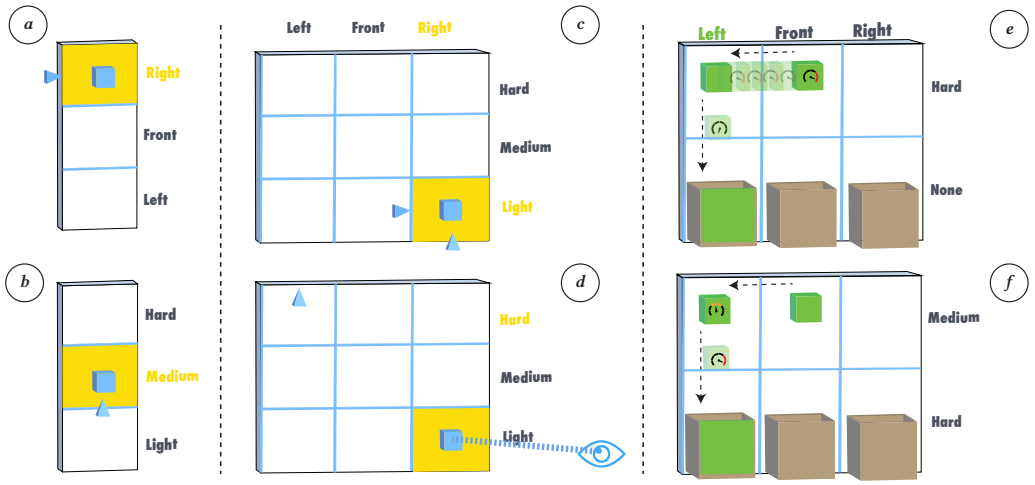


Fig. 2. Schematic representations of the six VR tasks used in Study 1 (see supplementary video for live VR footage). (a) shows Position-Only and (b) Pressure-Only control. The middle panel depicts integration tasks where (c) Position+Pressure and (d) Head-Gaze+Pressure combinations (assuming left-side pressure application) were evaluated in a 3×3 matrix for selection and confirmation. The right panel illustrates our Drag-and-Drop techniques: in both (e) and (f), the horizontal arrows represent lateral tongue position moving the object left or right toward the target bin; the downward arrow represents the object dropping into the bin, triggered by a pressure action rather than a tongue position change. (e) Hold-to-Hold, requiring sustained pressure while moving to the target location before releasing, and (f) Press-to-Drop, using medium pressure for positioning and hard pressure for dropping at the same location.

each channel (Selection), to multimodal combinations (Integration), and culminating in complex action (Drag-and-Drop), which is a form of "Spring-locked Mode" used in creating anticipation in games [60].

Participants completed six interactive VR tasks (Figure 2), divided into three blocks designed to progressively test the limits of physical coordination and cognitive load. The session (approx. 60–90 minutes) began with a model training phase, where participants provided data for 9 distinct position-pressure combinations (3 positions × 3 pressure levels) to train the user-dependent regressor. Block order was fixed to support skill acquisition, progressing from baseline to combined control to sustained manipulation, while task order within blocks was counterbalanced across eight predefined sequences: (1) **Selection Block** establishes baseline performance for position and pressure in isolation, (2) **Integration Block** examines how channels combine with each other and with head-gaze for 2D selection, and (3) **Drag-and-Drop Block** evaluates sustained control and coordination through two distinct pressure-based manipulation techniques. Head-gaze was chosen as the paired modality for Integration Block because it is natively supported by XR headsets without additional hardware, is the most common hands-free targeting mechanism in XR systems, and unlike voice or facial gestures does not interfere with mouth-based sensing. We prioritized head-gaze over eye-tracking because it requires no per-session calibration and avoid microsaccadic jitter that can cause unintended target drift during sustained use [37]. This pairing therefore represents the most ecologically valid combination for the target deployment context. To have consistent sampling density within fatigue constraints, each task consisted of two repetitions per target: 6 trials per task for the 3-target 1D tasks (Selection and Drag-and-Drop blocks) and 18 trials per task for the 9-target 2D grid tasks (Integration block). Physically, tongue position was produced by moving the

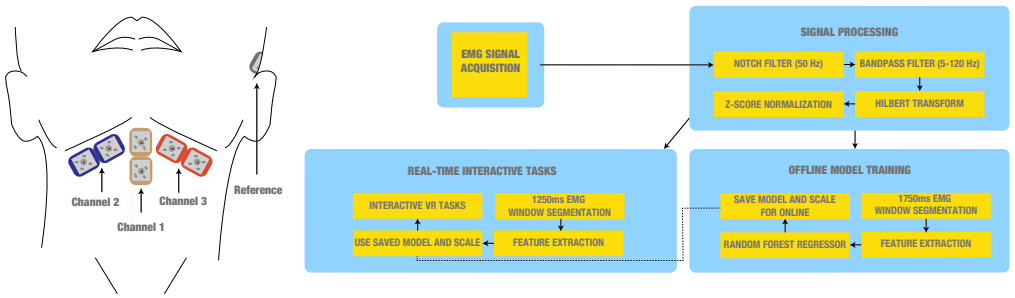


Fig. 3. Overview of pipeline used in our EMG-based tongue-input system, illustrating the steps from signal acquisition via submental electrodes to real-time classification for VR interactive tasks.

tongue to the palate: left corresponded to the upper-left molar area, right to the upper-right molar area, and front to the front of the palate. Pressure levels were calibrated per participant: during the training phase, each participant produced their maximum effort at each location, and the regressor output was normalized relative to that maximum. In pressure-only or modifier conditions, a light press at any of the three palate positions registers as low, a firmer press as medium, and a hard press as high. The following is a breakdown of each task in detail:

(1) Position-Only. Participants controlled a horizontal cursor via tongue position and moved it to one of three targets. A confirmation occurred automatically after holding the cursor over the target for 3 s.

(2) Pressure-Only. Tongue pressure acted as the sole input. Participants modulated continuous pressure to fall within one of three discrete buckets (low, medium, high). This task isolated pressure as a primary input channel.

(3) Position+Pressure. Participants selected targets on a 3×3 grid by combining horizontal position (column selection) with discrete pressure levels (row selection). Successful selection required meeting both criteria during the 3s dwell period.

(4) Head-Gaze+Pressure. Participants fixated on one of nine grid items using head-gaze and then performed a hard tongue press to confirm selection. This pairing decouples attention (aiming) from intention (commitment), offering a possible resolution to the Midas-touch problem.

(5) Hold-to-Hold. Participants applied sustained pressure to "grab" an object, used tongue position to move it to a target bin, and then released pressure to drop it.

(6) Press-to-Drop. Participants "pinned" the object with a medium press, moved it freely without sustained pressure, and dropped it with a hard press. This contrasts discrete latching against the sustained mechanic of Hold-to-Hold.

3.3 Apparatus

sEMG was recorded using an OpenBCI Cyton board at 250 Hz with six Ambu WSP30 electrodes placed along the submental triangle and bilaterally along the jawline (see Fig. 4 and Fig. 3). Electrodes were arranged in differential pairs (30 mm spacing) and aligned with muscle fiber orientation to reduce crosstalk and improve signal quality. Participants wore a Meta Quest 3 headset; VR tasks were implemented in Unity.

In brief: the raw electrical signal is cleared of noise and converted into a a smooth amplitude envelope representing moment-to-moment muscle effort. Raw sEMG was processed through a 50 Hz notch filter [14, 47, 59] and a fourth-order zero-phase Butterworth band-pass filter (5–120 Hz).

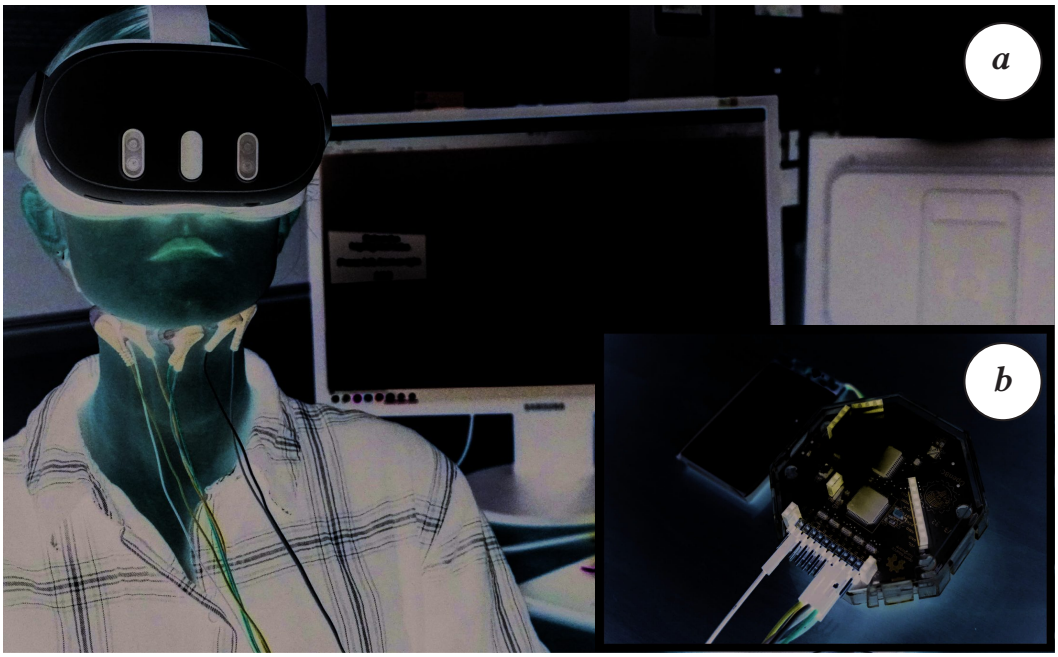


Fig. 4. Experimental setup. (a) Meta Quest 3 with submental sEMG; Monitor displays the VR task. (b) Inset: OpenBCI Cyton board with three differential EMG channels and a bias connection.

To extract an amplitude envelope under noisy, long-window conditions, the Hilbert transform was used [9] (Figure 3). The envelope is segmented into time windows, from which statistical descriptors are extracted to characterize the shape of the input for the model input. Feature extraction occurred on 1750 ms non-overlapping windows for offline model training. Seven time-domain features (RMS, ZC, WL, MAV, IAV, VAR, STD) and one frequency-domain feature (mean frequency) were computed per channel. Following the train–test split, features were standardized using participant-specific z-scoring [59]. Rather than classifying gestures as discrete commands, the model estimates continuous pressure magnitude and position, preserving effort information. Unlike classification-based tongue interfaces that dominate prior work, we employ a regression-based model to capture the continuous magnitude of tongue muscle activity. This allows us to discretize pressure only where required for reliable action, while preserving residual variability for feedback use. A multi-output Random Forest regressor was trained to estimate continuous pressure magnitude and tongue position, with hyperparameters tuned via grid search [39]. Signal processing and inference were implemented in Python (using SciPy and scikit-learn) and streamed via UDP to the VR application. For Study 1, incoming data were buffered into conservative 1250 ms non-overlapping windows. This was a deliberate methodological choice, as non-overlapping windows guarantee that each prediction is drawn from a statistically independent segment of signal, enabling reliable attribution of errors to specific interaction events. In a real-time application, standard overlapping (sliding) windows would reduce end-to-end latency to 200 ms, as demonstrated in Study 2 (see Section 4.2).

3.4 Measurements

We extracted four primary performance metrics from the interactive tasks:

- **Normalized error.** Spatial accuracy was quantified as the *normalized error*, computed from the Manhattan distance between the highlighted target cell and where participants dwelled with the cursor for 3 seconds to make their selection. Because tasks varied in grid dimensionality, the Manhattan distance functions as a layout-agnostic metric:

$$E_{\text{norm}} = \frac{|x_{\text{selected}} - x_{\text{target}}| + |y_{\text{selected}} - y_{\text{target}}|}{E_{\max}}$$

  where $E_{\max}$ is the maximum possible error for that specific block's layout (2 for Selection/Drag-and-Drop, 4 for Integration).
- **Number of crossings.** The count of target re-entries after the initial entry, measuring control stability:

$$N_{\text{crossings}} = \max\Big(0,\ \sum_{i=2}^{n} \mathbb{I}(\text{cell}_{i-1} \neq \text{target} \ \wedge \ \text{cell}_i = \text{target}) - 1\Big)$$

- **Time to target.** Task completion time minus the fixed 3-second dwell. This dwell requirement was a methodological choice to minimize false positives and ensure that sustained pressure control was feasible. However, for our analysis and for real-world applications, such long dwell times are unnecessary. In Study 2, this dwell constraint is removed to enable immediate interaction.
- **Task success.** A binary variable indicating whether the final confirmed selection matched the intended target.

Participants completed a questionnaire assessing perceived ease of control, responsiveness, mental/physical demand, intuitiveness, and social comfort. Items were informed by NASA-TLX [26], SUS [7], and themes in hands-free interaction research [18, 51, 72, 74], rather than used as full validated scales (full item wording reported in Fig. 6). We selected items targeting the perceptual dimensions most relevant to a novel physiological input in XR, rather than administering full instruments whose higher-order constructs (overall usability, workload index) are less informative at this exploratory stage. Participants also answered preference questions for each block (e.g., Position-only vs. Pressure-only, Position vs. Head-gaze, Hold-to-Hold vs. Press-to-Drop). Open-ended questions gathered participants' reasoning and qualitative feedback.

3.5 Data Analysis

Performance was analyzed using generalized linear mixed-effects models (GLMMs) [3] with *Task* as a fixed effect and *Participant* as a random intercept. To strictly adhere to data characteristics, we modeled: (1) **Normalized Error & Success:** Binomial distribution, (2) **Time-to-Target:** Gamma distribution with a log link (to address skew), (3) **Crossings:** Negative binomial distribution (to handle overdispersion). Estimated marginal means (EMMs) were computed using the Kenward–Roger method [44], with pairwise contrasts Tukey-adjusted for multiple comparisons. Model diagnostics (posterior predictive checks) confirmed no assumption violations.

3.6 Results

Performance and Evaluative Experience. We found that integration of Head-gaze with pressure input significantly outperformed Position+Pressure. As shown in Figure 5, Table 2 and Table 3, Head-Gaze+Pressure demonstrated drastic reductions in normalized error ($\Delta = -0.223$, $p < .001$) and time to target ($\Delta = -16.15s$, $p < .001$). The latter is largely explained by control stability: regression analysis indicated that each crossing event added approximately 4.7 s to completion time ($p < .001$), and Head-Gaze reduced these events significantly ($\Delta = -1.17$, $p < .001$). Finally, Head-Gaze+Pressure showed significantly higher success probability ($OR = 18.23$).

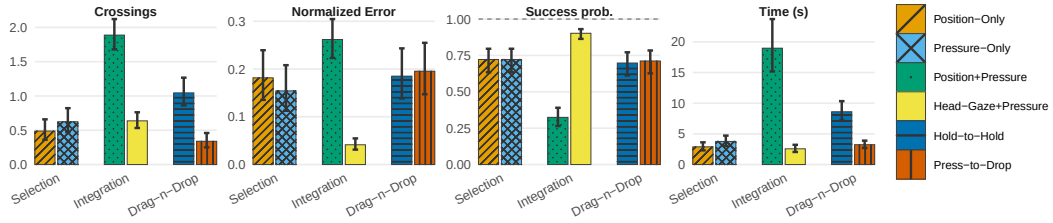


Fig. 5. Objective performance across tasks (estimated marginal means ± 95% CI).

	Selection		Integration		Drag-and-Drop	
Measure	Position-Only	Pressure-Only	Position+Pressure	Head-Gaze+Pressure	Hold-to-Hold	Press-to-Drop
Error	0.186 (0.026)	0.165 (0.025)	0.266 (0.020)	**0.043 (0.006)**	0.186 (0.026)	0.207 (0.028)
95% CI	*[0.139, 0.243]*	*[0.122, 0.220]*	*[0.228, 0.308]*	*[0.033, 0.057]*	*[0.139, 0.243]*	*[0.157, 0.267]*
Time (s)	2.83 (0.33)	3.77 (0.44)	18.78 (2.14)	**2.64 (0.30)**	8.61 (0.83)	**3.44 (0.34)**
95% CI	*[2.25, 3.55]*	*[3.00, 4.74]*	*[15.02, 23.49]*	*[2.10, 3.31]*	*[7.13, 10.40]*	*[2.84, 4.16]*
Crossings	0.493 (0.076)	0.604 (0.086)	1.833 (0.109)	**0.666 (0.060)**	1.058 (0.104)	**0.318 (0.051)**
95% CI	*[0.365, 0.665]*	*[0.457, 0.799]*	*[1.632, 2.059]*	*[0.559, 0.794]*	*[0.872, 1.283]*	*[0.233, 0.435]*
Success	0.716 (0.043)	0.709 (0.043)	0.323 (0.031)	**0.897 (0.017)**	0.698 (0.041)	0.698 (0.041)
95% CI	*[0.626, 0.792]*	*[0.619, 0.786]*	*[0.265, 0.387]*	*[0.858, 0.926]*	*[0.612, 0.772]*	*[0.612, 0.772]*

Table 2. Performance measures: estimated marginal means (standard errors), with 95% confidence intervals in brackets. Bold values indicate significance within the block.

Measure	**Selection** (Pos-Only − Pres-Only)	**Integration** (Pos+Pres − Head-Gaze+Pres)	**Drag-and-Drop** (Hold-to-Hold − Press-to-Drop)
Error	Δ = -0.021	Δ = **-0.223**	Δ = 0.021
	[-0.082, 0.041], $p = .513$	*[-0.258, -0.188]*, $p < .001$	*[-0.044, 0.085]*, $p = .529$
Time (s)	Δ = 0.945	Δ = **-16.147**	Δ = **-5.174**
	[-0.133, 2.022], $p = .086$	*[-20.056, -12.238]*, $p < .001$	*[-6.653, -3.694]*, $p < .001$
Crossings	Δ = 0.111	Δ = **-1.167**	Δ = **-0.740**
	[-0.087, 0.310], $p = .272$	*[-1.389, -0.945]*, $p < .001$	*[-0.948, -0.532]*, $p < .001$
Success	OR = 0.97	OR = **18.23**	OR = 1.00
	[0.58, 1.62], $p = .895$	*[12.45, 26.70]*, $p < .001$	*[0.60, 1.66]*, $p = 1.000$

Table 3. Estimated differences (Δ) and odds ratios (OR), with 95% confidence intervals in brackets. Bold values indicate statistical significance ($p < .05$).

Head-Gaze+Pressure was also the only interaction technique to exhibit a significant, albeit slight, learning effect ($\beta = -0.004$, $p = .029$), indicating gradual user adaptation. In the Selection Block, error rates between Position-Only and Pressure-Only did not differ significantly ($p = .513$), nor did success probability ($p = .895$). For Drag-and-Drop, Press-to-Drop delivered significantly

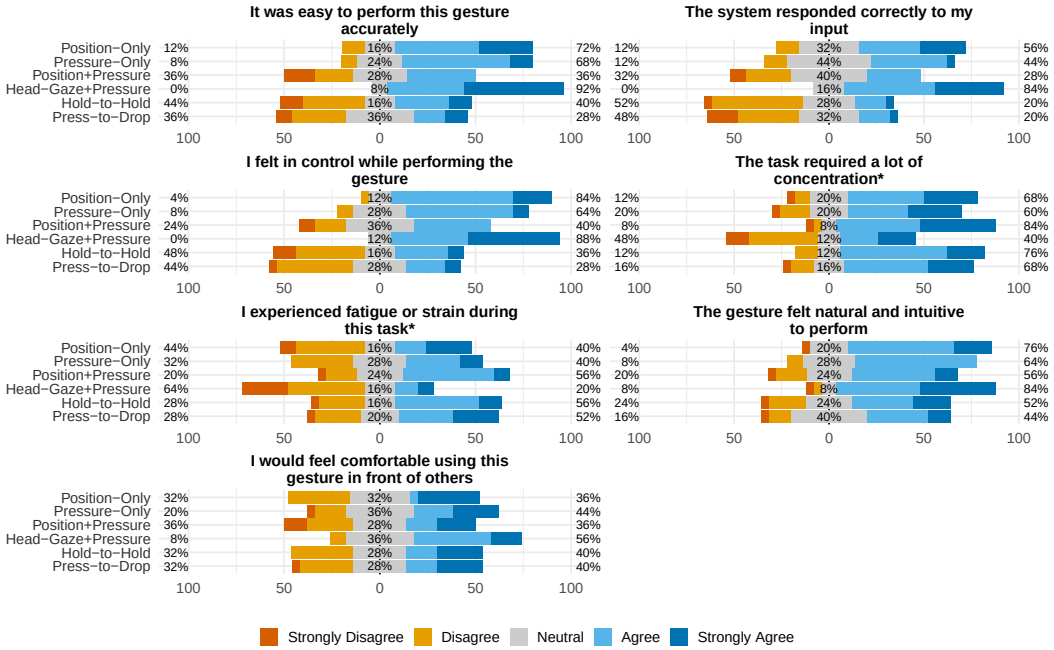


Fig. 6. Subjective ratings (estimated marginal means ± 95% CI). Items marked with * are reverse-coded.

faster completion times than Hold-to-Hold ($\Delta = -5.17s$, $p < .001$), though error rates and success probabilities were statistically indistinguishable.

Model Performance. Offline classifier performance (Table 5) showed high positional accuracy (85.7%, MCC = 0.796). Pressure classification achieved a balanced accuracy of 64.6% but maintained high ordinal integrity (1-Off Accuracy = 96.7%), indicating errors were overwhelmingly confined to adjacent levels. Participants consistently outperformed the classifier's predicted limits (actual error 17.4% vs. predicted 41.2%), suggesting a human–model gap where users compensated for system noise.

Subjective Feedback. Subjective measures (Table 4 and Fig. 6) aligned closely with performance. Participants rated Head-Gaze+Pressure superior to Position+Pressure across nearly all dimensions, including Ease of Use ($p < .001$), System Response ($p < .001$), and Sense of Control ($p < .001$). Notably, despite requiring visible head movements, Head-Gaze+Pressure was rated significantly higher in Social Comfort than the invisible tongue-only condition ($p < .05$). In the Integration Block, Position+Pressure was found significantly more mentally demanding ($p < .001$) and physically fatiguing ($p < .01$).

Qualitative Feedback. Participants responded positively to the novel input, describing it as *"sci-fi tech"* (P10) or imagining social party games (P19). Practical applications were also highlighted: P6 and P22 envisioned micro-interactions like *"skipping songs"* or ' *'minor secondary interactions,"* while P8 and P6 noted accessibility potential. P11 described the *"swiping across teeth"* interaction as particularly natural. Several participants (P2, P4, P14, P21, P24) identified distinguishing between three pressure levels as a challenge, suggesting a reduction to two levels. A strong desire for richer feedback emerged: P12, P17, P20, and P24 requested enhanced audio-visual cues (e.g., *"glowing light," "sound when task is done"*) to aid calibration and sense of effort. P8 found sustaining precise

Measure	**Selection** (Pos-Only − Pres-Only)	**Integration** (Pos+Pres − Head-Gaze+Pres)	**Drag-and-Drop** (Hold-to-Hold − Press-to-Drop)
Ease of Use	Δ = 0.125 *[-0.647, 0.897], $p = .997$*	Δ = **-1.625** *[-2.397, -0.853], $p < .001$*	Δ = 0.042 *[-0.730, 0.814], $p = 1.000$*
System Response	Δ = 0.292 *[-0.424, 1.007], $p = .845$*	Δ = **-1.292** *[-2.007, -0.576], $p < .001$*	Δ = 0.083 *[-0.632, 0.799], $p = .999$*
Sense of Control	Δ = 0.333 *[-0.415, 1.082], $p = .789$*	Δ = **-1.250** *[-1.998, -0.502], $p < .001$*	Δ = -0.042 *[-0.790, 0.707], $p = 1.000$*
Mental Demand	Δ = 0.250 *[-0.458, 0.958], $p = .909$*	Δ = **1.042** *[0.333, 1.750], $p < .001$*	Δ = 0.042 *[-0.667, 0.750], $p = 1.000$*
Physical Demand	Δ = -0.042 *[-0.772, 0.689], $p = 1.000$*	Δ = **0.958** *[0.228, 1.689], $p < .01$*	Δ = -0.083 *[-0.814, 0.647], $p = .999$*
Intuitiveness	Δ = 0.292 *[-0.386, 0.970], $p = .813$*	Δ = -0.667 *[-1.345, 0.011], $p = .057$*	Δ = 0.083 *[-0.595, 0.761], $p = .999$*
Social Comfort	Δ = -0.125 *[-0.688, 0.438], $p = .987$*	Δ = **-0.583** *[-1.146, -0.021], $p < .05$*	Δ = 0.042 *[-0.521, 0.604], $p = 1.000$*

Table 4. Estimated differences (Δ) for subjective ratings (Likert 1-5), with 95% confidence intervals in brackets. Bold values indicate statistical significance ($p < .05$).

	Accuracy		**Evaluation**	
Modality	Balanced	1-Off	Score	Metric
Position	0.857	0.965	0.796	(MCC)
Pressure	0.646	0.967	0.604	(QWK)
Combined	0.588	*0.934*	0.528	(MCC)

Table 5. Offline classifier performance. Balanced Accuracy reflects the average recall per class. For nominal Position classes, we report the Matthews Correlation Coefficient (MCC); for ordinal Pressure levels, we use Quadratic Weighted Kappa (QWK). MCC ≈ 0.80 indicates strong predictive quality, while QWK ≈ 0.60 corresponds to borderline substantial agreement. The 1-Off column shows that nearly all pressure errors were limited to adjacent levels.

pressure physically demanding, while P1 noted the session was *"not enough to acquire fine control,"* though others (P10, P4, P16) remained optimistic about hardware improvements.

In pressure tasks, participants adopted diverse strategies for side selection (Left/Right/Front). Some aligned with handedness (P4, P15, P16), while others prioritized ease (P5, P6, P7), control (P9, P10, P13), or naturalness (P1). When choosing between Position-Only and Pressure-Only, the majority (18/24) preferred Position for its precision (P15, P16). However, some (P6, P19) found pressure more *"embodied and natural,"* despite the control challenges. In the Integration Block, preference for Head-Gaze+Pressure was nearly unanimous. P19 noted it offered *"more accuracy and better feedback,"* while P5 and P8 found the alternative Position+Pressure *"tiresome."* For Drag-and-Drop, preferences were split. 13 participants preferred Hold-to-Hold, describing it as *"relaxing"*

(P22) or intuitive (P6, P7, P8, P17), despite worse performance metrics. Conversely, those favoring Press-to-Drop (P1, P15, P18, P20) cited better control and a clearer mental model of confirmation.

3.7 Discussion

Study 1 established non-invasive sEMG tongue pressure as a viable input method for XR under specific role assignments and modality pairings. Participants achieved stable control and high success rates when tongue input was assigned a single role in selection, either through tongue position or pressure. In contrast, the intramodality mapping that required simultaneous, granular control of both position and pressure (Position+Pressure) exceeded the effective resolution of the sensing setup, resulting in increased error, instability, and fatigue. Moreover, while participants achieved high 1-Off accuracy (96.7%), their feedback suggested that fewer discrete levels would better support interaction.

The combination of head-gaze and tongue pressure emerged as the most robust configuration. Assigning the primary selection role to head orientation while reserving tongue pressure in the modifier role for confirmation allowed the interaction to absorb classification error and enabled users to adapt to the predictable manner in which the error manifested (e.g., similar to adapting to a noisy joystick). This adaptation was reflected in the data as a consistent gap between offline model accuracy and real-time player performance.

The Social Comfort result also requires a deeper look. Contrary to the intuitive expectation that an invisible interaction such as tongue-only input would be rated higher in social comfort than a more visible modality, we observed the opposite: Head-Gaze+Pressure was rated significantly higher. We attribute this to the strong coupling between ease of use and perceived social acceptability. The tongue-only condition demanded visible concentration to maintain simultaneous control, which participants may have felt to be more conspicuous despite the input itself being hidden. Head-Gaze+Pressure, by contrast, leverages a natural looking behaviour as its primary channel, making the overall interaction appear more like passive observation than deliberate input effort.

The evaluations of Hold-to-Hold and Press-to-Drop are similarly explained by the demands of simultaneous control. Both techniques required users to maintain accurate tongue pressure while actively changing position, placing two parallel high-precision demands on the same effector. Overlapping muscle activations caused crosstalk that the regressor would sometimes not resolve correctly, producing erroneous predictions and a sense of lost control. Press-to-Drop was better received than Hold-to-Hold despite involving the same two channels, because it structured position and pressure sequentially rather than simultaneously: medium pressure pinned the object, freeing the user to attend to position, and then high pressure resolved the action. This sequential decomposition kept each role bounded and clear, reducing the cognitive and physical load that made the parallel conditions frustrating.

However, Study 1 treated physiological variability strictly through a lens of control error, with residual variability manifesting as cursor jitter and unintended crossings. To isolate feasibility and support stable qualitative assessment for Study 2, the study additionally imposed a dwell time of 3*s* and an end-to-end latency of 1.25*s*. This latency is not a hardware constraint but an analytical one as non-overlapping windows were used in order to have statistically independent error attribution per trial. Similarly, the 3*s* dwell requirement was a methodological safeguard to minimize false positives and false negatives during feasibility assessment and not a feature of the interaction design. The purpose of Study 1 was to establish what the tongue can do under controlled measurement conditions, and not predict gaming performance. While this framing was appropriate for establishing viability, it leaves unexamined how physiological variability might be interpreted once it decouples from the control layer.

In particular, Study 1 does not address how tongue input feels in gameplay contexts when discrete pressure levels are lowered, dwell time is removed, overlapping windows are used to reduce latency, or when residual variability is used as a texture to drive game feel. These unresolved questions motivate Study 2, which reframes physiological variability not as noise to be suppressed, but as a resource for play.

4 STUDY 2: TONGUE IN PLAY

Study 1 established the feasibility of tongue pressure interaction measured with non-invasive EMG sensors. It was demonstrated that when combined with head-gaze, tongue pressure created an efficient and natural interaction pattern where head orientation determined the selection, while the tongue confirmed it. Study 2 asks "what does this combination and the residual variability from EMG afford in play?" Study 2 follows a within-participant qualitative probe design utilizing a 2×2 factorial structure: Temporal Context (Anticipatory vs. Reactive) × Locus of Feedback (Egocentric vs. Exocentric). Anticipatory feedback is rendered *before* action resolution: residual variability builds visible tension that fires once a threshold is reached (as in drawing a bow). Reactive feedback, on the other hand, is rendered *during and after* the action: the output responds immediately to ongoing effort without a discrete threshold (as in throttle vibration). It intentionally contrasts with Study 1, where the continuous pressure signal was mapped directly to cursor position, making any signal variability visible as control noise (e.g., cursor jitter). In Study 2, we reframe this variability. Instead of displaying it as control error, we isolate the residual variability and map it to feedback (e.g., tremor, rumble), asking how the spatial assignment of this feedback changes the player's interpretation of effort and agency.

4.1 Participants

Twelve participants (6 female, 6 male), aged 22–34 years ($M = 27.4$, $SD = 3.5$), were recruited. VR experience ranged from 1 to 3.5 on a scale of 1–5 ($M = 2.6$, $SD = 0.7$), while weekly gaming hours ranged from 0 to 20 ($M = 5.0$, $SD = 6.0$). Two participants had previously taken part in Study 1 conducted several months prior. The study followed institutional ethical guidelines, and informed consent was obtained from all participants prior to the session and they received a cinema ticket as compensation.

4.2 Apparatus and Interaction Configuration

Participants wore a Meta Quest 3 headset. Head-gaze served as the primary spatial control, while tongue pressure served as the complementary control channel. Consistent with the flexible strategies used in Study 1, participants were free to apply pressure against the position of their choice (Left, Front, or Right). To support both reliable game mechanics and seamful [10] feedback, we implemented a dual-stream architecture. We use the term residual variability to denote fluctuations or variations in the normalized EMG amplitude envelope that remain before discretizing the control signal into stable intent states. Therefore, residual variability reflects moment-to-moment instability in exertion rather than environmental or electrical noise.

Control Stream: Raw time-series data were streamed to the custom Python pipeline (identical to Study 1 but optimized with 200 ms [59] overlapping windows). This stream prioritized gameplay stability: it used the trained Random Forest regressor to estimate the user's discrete intent. Informed by the high 1-Off accuracy (96.7%) in Study 1, we mapped this output to simplified bins (2 levels for the anticipatory prototype, 3 levels for the reactive prototype), effectively absorbing classification jitter for reliable input.

Feedback Stream: We utilized the standard "EMG Joystick" computed channel from the OpenBCI acquisition GUI. This stream provides a normalized amplitude envelope that is already cleaned

of electrical artifacts (via hardware notch and band-pass filters) but retains the high-frequency physiological variability of the muscle's exertion. We mapped this "raw effort" signal directly to visual texture (tremor/rumble), allowing the player to feel the continuous struggle of the muscle that the Control Stream intentionally suppressed.

4.3 Game Prototypes

We developed two VR game prototypes – Frog Game and Kart Game (Figure 7) – to test how temporal and spatial manipulation of feedback is perceived and interpreted, resulting in a total of four conditions. The direction of the jitter was randomized (per frame and per object), while its amplitude was scaled by the participant's normalized EMG signal. As a result, players could not predict or control the direction of movement, but changes in effort directly increased or decreased the intensity of the feedback.

Frog Game (Anticipatory Feedback). Participants played as a frog catching wasps, during which head-gaze determines aim. Tongue pressure increases tension in the frog's tongue and once a fixed threshold is reached, the tongue snaps forward to catch the target.

- *Egocentric Condition:* Residual pressure variability is mapped to visual tremors in the frog's stuck-out tongue avatar. This mapped effort internally, reinforcing the proprioceptive illusion of a straining muscle.
- *Exocentric Condition:* Residual variability is mapped to the target wasp, causing it to jitter in mid-air relative to the player's struggle. This mapped effort externally to the target's stability.

Kart Game (Reactive Feedback). Adapted from the Unity Kart Microgame template [1] (© Unity Technologies), this prototype required participants to drive a kart around a track. Head-gaze controls steering, while tongue pressure controls throttle (rest, cruise, boost).

- *Egocentric Condition:* Residual pressure variability corresponds to chassis vibration; spikes in tongue effort at high speeds intensified the rumble into a jolt and a jump. This linked effort internally to the directly controlled vehicle.
- *Exocentric Condition:* Residual variability is mapped to environmental instability. Specifically, the road, trees, arches, and stones vibrate in sync with the player's effort. This mapped effort externally to the world that becomes turbulent at high speeds.

4.4 Procedure

Each session lasted approximately 50 minutes. Participants first completed a model training phase (approx. 10 minutes), and before each game block, they completed a short familiarization phase (approx. 1-2 minutes) in which they practiced the Head-Gaze+Pressure interaction in the respective games without feedback. Participants completed four gameplay blocks (2 Games × 2 Conditions). Game order was fixed to scaffold complexity (Frog to Kart), while the locus of feedback conditions (Egocentric vs. Exocentric) was counterbalanced within each game. For each game, participants played both the Egocentric and Exocentric versions (2 minutes each) back-to-back to facilitate direct comparison. No video replay or concurrent think-aloud was used, as speaking would interfere with tongue-based input and disrupt gameplay, and experiential data were therefore elicited post-task [8, 31] immediately. To structure the elicitation, we developed open-ended probes grounded in theories of Game Feel [71] and Sense of Agency [25]. For each game, questions probed the general experience and physical sensation of control (e.g., "How would you describe the physical sensation of using your tongue to control the game?") to allow participants to discuss difficulty or fatigue naturally. Subsequent probes narrowed focus to the connection with the avatar or vehicle (e.g., "Did the tension you felt in the game match the tension in your mouth?"). The probes then addressed the locus of variability. Participants were asked to explicitly contrast the Egocentric and

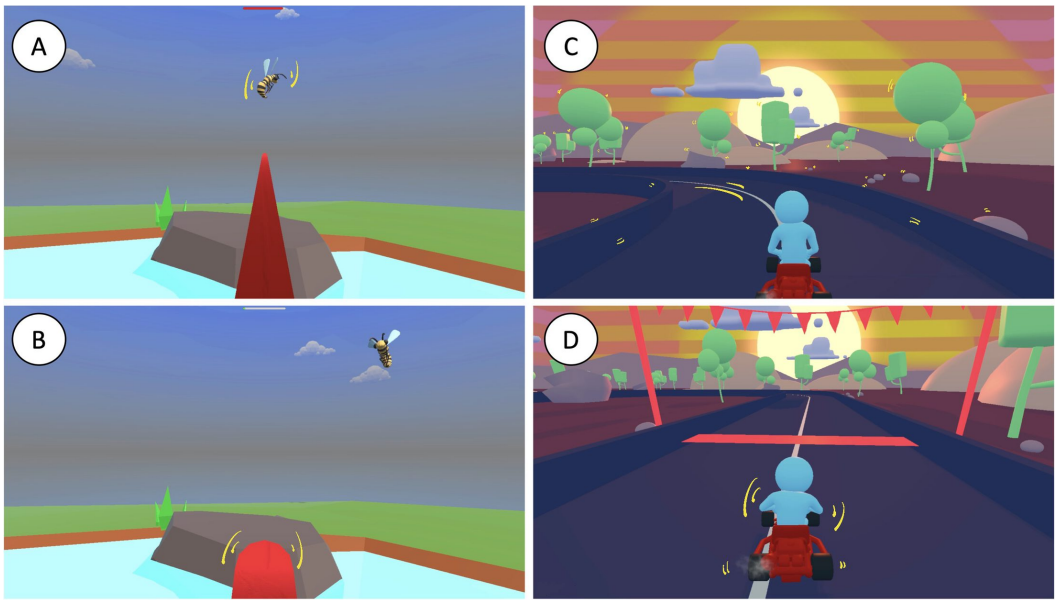


Fig. 7. We explored non-invasive tongue pressure input for gaming in egocentric and exocentric conditions: (A) Frog game/exocentric, in which residual variability is mapped to the target wasp, causing it to jitter in mid-air relative to the player's struggle , (B) Frog game/egocentric, in which residual pressure variability is mapped to visual tremors in the frog's tongue, (C) Kart game/exocentric, in which residual variability is mapped to environmental instability, (D) Kart game/egocentric, in which residual pressure variability corresponds to chassis vibration. Kart game (C, D) adapted from the Unity Learn Karting Microgame, © Unity Technologies.

Exocentric conditions, differentiating whether the feedback felt like an internal struggle or external resistance, and how this attribution influenced their perception. Finally, participants were asked to reflect on how knowing the physiological source of the vibrations from the beginning changed their interpretation of the feedback in both games.

4.5 Data Analysis

Interview data were analyzed using reflexive thematic analysis [6], which was interpretive and theoretically informed by game feel and embodied interaction. The analysis focused on how participants attributed variability and controlled bodily input. Analysis was done iteratively by the first author, starting with transcribing and reading all interviews multiple times for familiarization. For each participant, memos were written to capture immediate reflections, tensions, and patterns between the four conditions in how variability was described. Coding was carried out in multiple stages, with an initial, expansive, inclusive coding capturing a broad set of preliminary features in the data. Initial codes stayed close to the participant level and captured both semantic content (e.g., descriptions of fatigue, fear, instability) and moved on to more latent interpretations (e.g., redistribution of agency, bodily reorganization under discomfort). As such, coding was not a fixed categorization process; the codes were repeatedly revised, merged, reframed, and discarded in each iteration. After the code generation phase, the codes and memos were clustered into candidate preliminary themes and patterns. The process was guided by asking how the temporal and spatial context of the feedback affected participants' interpretation of physiological variability. As in the previous phase, decisions were documented through ongoing memoing, especially when generating

higher-level conceptual categories. Final themes were generated as interpretive claims about how players made sense of residual EMG variability in play. In line with reflexive thematic analysis and its epistemic standing, themes are not summaries, are not presented as frequency-based findings, and no saturation or inter-code reliability was sought, which would be more fit for other forms of thematic analysis. The themes reflect a situated, iterative engagement with the dataset to articulate how physiological variability is reorganized as effort, threat, or controllability under our factorial conditions.

4.6 Results

Study 2 adopts an interpretive lens to examine how players experienced and made snese of residual physiological variability when rendered as visual shake in four conditions (Anticipatory/Reactive × Egocentric/Exocentric). Analysis produced four themes, each addressing a distinct dimension of the play experience: (1) **Physiological Ventriloquism**: how players attributed self-generated variability to external game entities; (2) **Anticipatory and Sustained Reaction**: how temporal context shaped the experience of effort; (3) **From Binary Pressing to Graded Control**: how variability feedback drove motor learning; and (4) **Emergent Control Metaphors**: how players reorganized the tongue into a meaningful control surface. Each theme is reported in a dedicated subsection below.

4.6.1 ***Physiological Ventriloquism.*** In the classic ventriloquism illusion, auditory spatial perception is captured by temporally aligned visual information, such that sounds are mislocalized toward a visible source (e.g., a dummy) rather than their true acoustic origin (e.g., the ventriloquist), even when observers explicitly know the sound's actual source [16]. Similarly, in our study, although participants were explicitly informed that the visual shake was generated by their own tongue muscle activity, this knowledge did not constrain their experiential attribution. Participants' moment-to-moment sense-making treated physiological variability as an external property of the game world rather than as their own bodily variability. We invoke the ventriloquism parallel not as a mechanistic claim, but a metaphor. Both phenomena involve misattributing an internal source to an external object despite retained knowledge of its true origin, but our phenomenon concerns a disruption in the *feeling of agency* [5, 25] over self-generated effort rather than auditory spatial capture [16]. This was most evident in exocentric mappings (particularly when the wasp jittered), where the visual shake was frequently interpreted as hostility or intent. For example, in the frog game, the participants described the wasp as *"angry"* (P12) and *"about to attack"* (P03). P02 reported a moment of panic because the physiological tremor made the shaking feel "very realistic", while P03 noted that in the beginning it affected their decision to move their tongue until they "reminded myself it's just a game" (P03). Interestingly, in the egocentric and neutral conditions, the low-poly wasp alone did not elicit strong reactions, suggesting the variability provided animacy and intentionality. In the kart game, attribution in exocentric conditions was also temporally unstable. For some participants, awareness of self-causation was salient at the beginning and faded during gameplay, while for others, the realization surfaced only late in play. This suggests that causal knowledge remained cognitively accessible throughout but was not sufficient to anchor attribution under affective or attentional load. Egocentric mapping, on the other hand, was readily owned in both games; however, it was described as disruptive in the kart game, owing mostly to the effort spikes causing the vehicle to jolt and jump. Participants described the jumps as *"annoying"* (P03) or making them feel "out of control" (P01) even while knowing their effort was the source. These patterns suggest a form of physiological ventriloquism where self-generated variability was experienced and reacted to as if it belonged to external entities, despite retained causal knowledge of its origin.

4.6.2 ***Anticipation and Sustained Reaction.*** Anticipatory (frog tongue snap) and sustained reactive (kart acceleration and rumble) feedback produced qualitatively different experiences among participants. Anticipatory variability was consistently interpreted as potential energy or tension buildup, whereas sustained reactive variability was interpreted as resistance or instability. In the frog game, shaking was often experienced as dramatic tension across different perspectives. For P05, the internal shake felt like "charging up" their shot, whereas P03 interpreted the same signal externally as the wasp entering an "attack phase," and P02 felt it was "coming toward me." Across egocentric and exocentric framings, anticipatory feedback retained a metaphor of accumulation and readiness. In contrast, sustained variability in the Kart game was perceived as instability. Rather than conveying speed or power, the egocentric jolts and jumps of the vehicle at signal spikes functioned as a mechanism that penalized high-intensity effort. As P09 described, the jumps were "obstructing the race", noting that unless the game was designed with obstacles to clear, the feedback simply disrupted the flow of driving. More importantly, this was often framed as a calibration issue rather than a conceptual one. P01 noted that the mechanics were engaging, but the feedback gain was *"too much"*, preferring a reduced signal. P03 similarly reported learning to *"hold back"* to prevent the car from jumping. Thus, signal used for sustained reactive feedback requires filtration or attenuation to prevent the user's own effort from becoming a disruptor. However, this interpretation changes once the novelty of the input modality is taken into account.

4.6.3 ***From Binary Pressing to Graded Control.*** Many participants reported that they had never used their tongue in this way and were largely unaware of its utility for digital interaction prior to the study. Because tongue pressure is rarely employed for continuous graded control, the interface placed the participants in a state of novel motor learning, eliciting patterns characteristic of early-stage skill acquisition. Most notably, we observed motor overflow, where effort directed at the tongue spread into the jaw, neck, and torso, and in extreme cases the entire upper body. P03 reflected: *"I was trying to remind myself that it's just the tongue making the movement, not my body going forward or the head going forward or something like that."*. P11 similarly reported unexpected bodily recruitment, expressing surprise: "I noticed that my neck make involuntarily movement without me, without me doing, uh, anything". This overflow manifested as a freezing of degrees of freedom [23]. P08 described that *"my arms get very stiff and, like, very tense"* and that *"my body is, like, tense"* during high-speed sections, effectively recruiting the upper body to stabilize tongue input. P11 also reported difficulty decoupling the tongue from the throat and mouth. The fatigue reported in the Kart task can therefore be understood not as tongue fatigue alone, but as the cost of whole-body recruitment. The same feedback that was experienced as disruptive in the previous theme functioned here as a learning signal. By visualizing excessive effort through vehicle jumps, the system exposed the invisible overexertion. P02 described that "instead of blindly pressing my tongue… I'm, like, thinking a little more mindful about the action." to reduce the fatigue. P03 similarly recognized that it was "an option not to keep going forward and, and keep pressing the acceleration". In this context, feedback intensity paradoxically revealed the structure of control and the cost of overexertion. Disruption forced attention onto an internal, invisible input modality that enabled participants to transition from whole-body recruitment to isolated, graded modulation of tongue effort.

4.6.4 ***Emergent Control Metaphors.*** The tongue interface was designed to register pressure across multiple contact locations without enforcing a single contact point or gesture. Rather than experiencing this flexibility as ambiguity, participants reorganized the tongue into a meaningful control surface, imposing structure through bodily exploration rather than explicit instruction. Spatial remapping was common across tasks and feedback conditions. Participants sought congruence between tongue placement and in-game events, often aligning lateral tongue pressure with

directional control. P04 described steering reflexively: *"I press my tongue on the left side whenever I need to turn left, and then the other way when I need to turn right."* Similarly, P05 described mirroring movement direction with steering while deliberately shifting muscle use to *"relax the muscles... and use maybe different muscles too."* In the wasp task, P12 articulated a comparable strategy: *"Left meant left pressure, right meant right. Neither was in the middle."* Beyond spatial alignment, participants frequently invoked familiar control metaphors, describing the mouth as a set of "buttons". These metaphors drew on prior interaction experience to make sense of an otherwise unfamiliar input space and emerged despite not being required by the system. Fatigue played a central role in shaping these remappings. As described in the previous theme, early interaction was characterized by excessive effort and whole-body recruitment. Over time, participants redistributed pressure across different tongue regions to reduce strain and sustain control. Sustained forward pressure in the kart task was widely described as tiring, whereas the frog task, which required brief, anticipatory effort, was often characterized as comfortable or even relaxing (P03, P12). These remapping behaviors point to a coupling between bodily action and game demands. Tongue activation did not occur in isolation but was experienced as coordinated with head orientation, visual attention, and task goals. Participants described adjusting tongue placement in ways that felt congruent with where they were looking and what the game demanded in that moment, suggesting that control emerged through the coordination of multiple bodily systems rather than through isolated muscle activation. Driven by sensing flexibility and participants' existing motor histories, the tongue interface became usable not through following instructions or learning a set of gestures, but through the projection of familiar spatial and bodily logics onto the system. In this way, participants transformed an unfamiliar physiological signal into a controllable and meaningful game input.

5 DISCUSSION AND FUTURE WORK

This work demonstrates that the tongue offers a unique resource for play. Our results suggest that sEMG tongue pressure supports a dual role in XR gaming: functioning as a viable hands-free input channel when constrained, while serving as a rich source of biological physicality that surfaces "effort" to deepen engagement.

5.1 Separating Control from Perception

Study 1 demonstrated a configuration of sEMG tongue pressure that is viable for XR interaction by discretizing its levels and constraining its role to modifier. It revealed the limits of the sensing device and the physiological input gained from it. Mappings that demanded simultaneous position and granular pressure control amplified error through two mechanisms: cross-talk from overlapping muscle activations, and signal resolution limits of the device. Residual variability in the control layer manifested as instability and cursor jitter. However, visualizing the regression output on the perceptual layer as a moving cursor allowed users to perceive the signal jitter, stabilize their input dynamically, and perform better than the offline model accuracy would have predicted. Based on these findings, Study 2 used a more responsive and cleaner version of the signal used in Study 1 and reframed this same residual variability by relocating it from control to perception. When rendered as a visual shake rather than used as part of actuation, variability in muscle effort was no longer experienced as error nor did they affect the main control of the game. Instead, they were interpreted as animacy, tension, resistance, and a learning signal. Across themes, participants reorganized physiological variability through attribution, temporal framing, motor calibration, and spatial remapping. The findings suggest that physiological variability, once filtered for noise, does not need to be suppressed or eliminated. Rather, it can be separated from the main control: stabilized or discretized for control, while preserved and rendered for game feel.

5.2 Physiological Ventriloquism and Attribution Instability

One unexpected finding of Study 2 was that spatial framing of feedback (i.e., exocentric vs. egocentric) influenced the sense of ownership. Even when participants were explicitly infromed before gameplay that the visual shake originated from their own tongue effort, they often experienced the variability as external, particularly under exocentric mappings. Cognitive knowledge of causality (*judgement of agency*) remained intact, but it did not stabilize the experiential sense of agency (*feeling of agency*) [5, 25]. We term this phenomenon *physiological ventriloquism*: a disruption in the *feeling of agency* [5, 25] whereby one's own physiological effort dynamics are misattributed to an external object or agent. We use the ventriloquism metaphor not to invoke the specific auditory-spatial perceptual illusion of multisensory capture [16], but to describe the analogous displacement of an internal "voice" (bodily effort) to an external "puppet" (the in-game entity). Participants retained *judgement of agency* (they knew they caused the shake) but lost the *feeling of agency* (the effort felt like it belonged to the game world). Baseline variability in the sEMG signal likely facilitated this externalization. According to the Comparator Model, agency relies on a precise match between motor prediction and sensory outcome [25]. For example, in the frog game, where participants controlled the avatar, egocentrically mapped jitter was experienced as internal tension or "charging up." The jitter matched both the biological expectation of a straining tongue and (as the art-style suggested) the comically eager frog avatar, so the predictive model accepted it as self-generated. In contrast, when mapped exocentrically to the wasp which was not controlled by the player, the same jitter was interpreted as aggression. The brain may have resolved the prediction error by displacing the participant's effort into the wasp's psychological state, and rather than perceiving the jitter as their own muscle effort, the player felt the wasp acting aggressively. These findings suggest that the brain continuously negotiates agency attribution. Baseline jitter primes externalization, but whether judgment and feeling of agency align depends on both avatar ownership and the spatial and metaphorical framing of the signal.

5.3 Beyond Direct vs Indirect Biofeedback

While physiological variability can be architecturally separated within an interaction loop, this also reframes how biofeedback signals are classified in games. Traditional distinctions between "direct" (voluntary, explicitly controlled) and "indirect" (involuntary) biofeedback [52, 62] often determine whether signals drive mechanics (Action) or modulate aesthetics (Context). Our findings suggest that this is incomplete: a signal's role also depends on whether its residual effort-coupled variability is leveraged perceptually. In our case, tongue pressure, a voluntary input, acted as a reliable actuator in Study 1 when discretized and stabilized. In Study 2, the same signal's residual variability was rendered as visual feedback, creating tension, instability, and animacy. The design question, therefore, shifts. It is not simply whether a physiological signal is voluntary, but whether its post-filtered variability remains meaningfully tied to user effort and can be exploited for game feel. Short-time variability in tongue effort were sufficient to generate aesthetics of control and sensory polish without procedural simulation, showing that biological effort itself can be a source for game feel.

5.4 Future Physiologically Controlled Games

Since tongue pressure is rarely used for continuous digital control, participants initially exhibited motor overflow, recruiting the jaw, neck, and upper body to stabilize their input. This whole-body recruitment caused significant fatigue, particularly in the kart game where sustained pressure was required. However, the feedback manifesting as repeating vehicle jolts at high pressures served to expose overexertion. This disruption forced participants to become aware of the multiple levels of

pressure and lower their effort. These findings suggest that rendering physiological variability not just as feedback, but as resistance, can scaffold calibration in novel interfaces. Future longitudinal studies are needed to disentangle the contributions of fatigue and feedback in shaping motor adaptation and determine which factor is most effective.

Residual physiological variability opens up gameplay mechanics and experiences that were previously not possible to achieve with conventional controllers. Effort dynamics can be mapped to other game feel techniques, audio, lighting, or haptic feedback, allowing the texture of the player's physical effort to be felt in different ways. These signals could support dexterity and motor learning, or even be applied in rehabilitation-focused games to increase gaming accessibility, where subtle variability guide precise control. In traditional gaming contexts, effort could be projected onto NPCs, crowds, or environmental objects, amplifying horror or tension, or increasing realism, such as the tremor felt when holding a weapon, directly linked to the player's exertion. More experimental or artistic applications are also possible. Players could sculpt or paint with their effort, or leave persistent traces of exertion in virtual spaces reminiscent of "dark souls"-like environments. Finally, future work could explore the inverse of the physiological ventriloquism phenomenon. Just as ventriloquism involves a breakdown in the *feeling of agency* over self-generated signals, the inverse would involve inducing a *feeling of agency* over externally generated signals. By synchronizing passive bio-signals (e.g., respiration or heart rate) with the movement of an autonomous ally or prosthetic limb, games could induce a feeling of agency over independent agents.

Together, these examples illustrate that physiological inputs, when leveraged with their residual, expressive variability, are no longer a redundant or auxiliary controller signal. Instead, they can contribute a unique layer of game feel, enriching gameplay through direct, embodied, and affective interaction that conventional controllers cannot provide.

6 LIMITATIONS

Our Study 1 results indicate a clear performance hierarchy: while tongue pressure functioned exceptionally well as a discrete modifier channel, when combined with tongue position in primary role it suffered from higher error rates and fatigue compared to head-gaze. This suggests that with current commercially available sEMG sensing resolution, the tongue is best utilized as a complementary channel (e.g., "aim with head, shoot with tongue") rather than a primary spatial controller. Future work using high-density EMG grids could potentially resolve this limitation, enabling more robust continuous tracking. The submental electrodes used in both studies are a controlled-lab instrumentation choice that reflects standard practice for sEMG feasibility work, and not a design target for deployment. Establishing feasibility under controlled conditions is a necessary precursor to work on miniaturized wearable alternatives and we do not claim that the current form factor is socially acceptable or ergonomically viable at scale. As observed in Study 2, the novelty of graded tongue pressure control elicited significant motor overflow, where participants recruited jaw, neck, and upper-body muscles to stabilize their input. However, these observations are constrained by the short duration of our experimental sessions. While we frame this as a stage of early motor learning, it presents a potential ergonomic risk. If this whole-body recruitment does not fade with longitudinal practice, sustained use could lead to fatigue or strain. Future longitudinal studies are required to verify that users can transition to isolated, efficient tongue control over time.

Finally, the phenomenon of physiological ventriloquism relies on a complex constellation of factors that our current design could not disentangle. While we propose that baseline signal jitter primed the externalization, we cannot determine which specific combination of variables is required for the effect to emerge. Our study design also combined spatial locus (where the feedback appears), narrative scaffolding (the metaphor of the angry wasp or turbulent kart), and attentional load (the

difficulty of the task). Furthermore, our findings are drawn from a limited sample size within a controlled laboratory setting. As a result, it remains unknown if our architecture alone is sufficient to drive this externalization across a broader demographic, or if the illusion is fragile and dependent on the specific alignment of a diegetic metaphor and high task demand.

7 CONCLUSION

Our research proposes a dual contribution: establishing sEMG tongue pressure as a viable controller for XR gaming, and demonstrating its capacity to drive a new category of biological game feel. Our experiments revealed that while tongue pressure functions best as a discrete modifier for head-gaze selection, its continuous variability, although too volatile for precision control, holds untapped experiential value. Building on the viable configuration established in Study 1, we uncovered physiological ventriloquism: a phenomenon where the raw jitter of exertion is externalized, allowing players to perceive their own physical effort as the animacy or resistance of the game world. These findings expand biofeedback interaction beyond the binary of "voluntary active control" versus "involuntary passive tracking," offering a new design pattern for embodied play. We show that by architecturally separating "intent" (stable control) from "texture" (residual variability), designers can leverage the variability inherent in biological input rather than suppress it, substituting procedural simulation with biological physicality.

Acknowledgments

This work was supported by the Digital Futures project XR Horizons. We thank Renan Guarese for their feedback and assistance, Elena Moser for the illustration in Fig. 1, and all our study participants for their time and engagement.

References

[1] 2023. Unity Learn | Karting Microgame | URP | Templates | Unity Asset Store. https://assetstore.unity.com/packages/templates/unity-learn-karting-microgame-urp-150956

[2] Isayas Berhe Adhanom, Paul MacNeilage, and Eelke Folmer. 2023. Eye Tracking in Virtual Reality: a Broad Review of Applications and Challenges. *Virtual Reality* 27, 2 (June 2023), 1481–1505. https://doi.org/10.1007/s10055-022-00738-z

[3] R. H. Baayen, D. J. Davidson, and D. M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59, 4 (Nov. 2008), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

[4] Felix Born, Sophie Abramowski, and Maic Masuch. 2019. Exergaming in VR: The Impact of Immersive Embodiment on Motivation, Performance, and Perceived Exertion. In *2019 11th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*. 1–8. https://doi.org/10.1109/VS-Games.2019.8864579 ISSN: 2474-0489.

[5] Davide Fausto Borrelli, Matteo Tonna, and Reuven Dar. 2024. An investigation of the experience of control through the sense of agency in people with obsessive–compulsive disorder: a review and meta-analysis. *CNS Spectrums* 29, 4 (Aug. 2024), 224–232. https://doi.org/10.1017/S1092852924000117

[6] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health* 11, 4 (Aug. 2019), 589–597. https://doi.org/10.1080/2159676X.2019.1628806 _eprint: https://doi.org/10.1080/2159676X.2019.1628806.

[7] John Brooke. 1995. SUS: A quick and dirty usability scale. *Usability Eval. Ind.* 189 (Nov. 1995).

[8] Paul Cairns, Anna Cox, and A. Imran Nordin. 2014. Immersion in Digital Games: Review of Gaming Experience Research. In *Handbook of Digital Games*. John Wiley & Sons, Ltd, 337–361. https://doi.org/10.1002/9781118796443.ch12 Section: 12 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118796443.ch12.

[9] Camila R. Carvalho, J. Marvin Fernández, Antonio J. del Ama, Filipe Oliveira Barroso, and Juan C. Moreno. 2023. Review of electromyography onset detection methods for real-time control of robotic exoskeletons. *Journal of NeuroEngineering and Rehabilitation* 20, 1 (Oct. 2023), 141. https://doi.org/10.1186/s12984-023-01268-8

[10] M. Chalmers, I. MacColl, and M. Bell. 2003. Seamful design: showing the seams in wearable computing. In *2003 IEE Eurowearable*. 11–16. https://doi.org/10.1049/ic:20030140 ISSN: 0537-9989.

[11] Victor Chen, Xuhai Xu, Richard Li, Yuanchun Shi, Shwetak Patel, and Yuntao Wang. 2021. Understanding the Design Space of Mouth Microgestures. In *Designing Interactive Systems Conference 2021*. ACM, Virtual Event USA, 1068–1081. https://doi.org/10.1145/3461778.3462004

[12] Jingyuan Cheng, Ayano Okoso, Kai Kunze, Niels Henze, Albrecht Schmidt, Paul Lukowicz, and Koichi Kise. 2014. On the tip of my tongue: a non-invasive pressure-based tongue interface. In *Proceedings of the 5th Augmented Human International Conference* (Kobe, Japan) *(AH '14)*. Association for Computing Machinery, New York, NY, USA, Article 12, 4 pages. https://doi.org/10.1145/2582051.2582063

[13] C.-H. Chou, Y.-S. Hwang, C.-C. Chen, S.-C. Chen, S.-W. Chou, and Y.-L. Chen. 2016. Noninvasive tongue-motion controlled computer mouse for the disabled. *Technology and Health Care* 24, 3 (2016), 401–408. https://doi.org/10.3233/THC-161137 arXiv:https://journals.sagepub.com/doi/pdf/10.3233/THC-161137 PMID: 26835730.

[14] Carlo J De Luca. 2002. Surface electromyography: Detection and recording. *DelSys Incorporated* 10, 2 (2002), 1–10.

[15] Ovidiu Draghici, Izmail Batkin, Miodrag Bolic, and Ian Chapman. 2013. The MouthPad: A tongue-computer interface. In *2013 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. 315–319. https://doi.org/10.1109/MeMeA.2013.6549759

[16] J. Driver and C. Spence. 1998. Crossmodal attention. *Current Opinion in Neurobiology* 8, 2 (April 1998), 245–253. https://doi.org/10.1016/s0959-4388(98)80147-5

[17] Ethan Eddy, Erik J Scheme, and Scott Bateman. 2023. A Framework and Call to Action for the Future Development of EMG-Based Input in HCI. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–23. https://doi.org/10.1145/3544548.3580962

[18] Pablo Gallego Cascón, Denys J.C. Matthies, Sachith Muthukumarana, and Suranga Nanayakkara. 2019. ChewIt. An Intraoral Interface for Discreet Interactions. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300556

[19] William W. Gaver, Jacob Beaver, and Steve Benford. 2003. Ambiguity as a resource for design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. Association for Computing Machinery, New York, NY, USA, 233–240. https://doi.org/10.1145/642611.642653

[20] Tan Gemicioglu, R. Michael Winters, Yu-Te Wang, Thomas M. Gable, Ann Paradiso, and Ivan J. Tashev. 2023. Gaze & Tongue: A Subtle, Hands-Free Interaction for Head-Worn Devices. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, 1–4. https://doi.org/10.1145/3544549.3583930

[21] Tan Gemicioglu, R. Michael Winters, Yu-Te Wang, Thomas M. Gable, and Ivan J. Tashev. 2023. TongueTap: Multimodal Tongue Gesture Recognition with Head-Worn Devices. In *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION*. ACM, Paris France, 564–573. https://doi.org/10.1145/3577190.3614120

[22] Mayank Goel, Chen Zhao, Ruth Vinisha, and Shwetak N. Patel. 2015. Tongue-in-Cheek: Using Wireless Signals to Enable Non-Intrusive and Flexible Facial Gestures Detection. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, Seoul Republic of Korea, 255–258. https://doi.org/10.1145/2702123.2702591

[23] Anderson Nascimento Guimarães, Herbert Ugrinowitsch, Juliana Bayeux Dascal, Alessandra Beggiato Porto, and Victor Hugo Alves Okazaki. 2020. Freezing Degrees of Freedom During Motor Learning: A Systematic Review. *Motor Control* 24, 3 (March 2020), 457–471. https://doi.org/10.1123/mc.2019-0060

[24] Kutlucan Görür, Mehmet Recep Bozkurt, Muhammed Serdar Bascil, and Feyzullah Temurtas. 2021. Tongue-Operated Biosignal over EEG and Processing with Decision Tree and kNN. *Academic Platform Journal of Engineering and Science* 9, 1 (Jan. 2021), 112–125. https://doi.org/10.21541/apjes.583049

[25] Patrick Haggard. 2017. Sense of agency in the human brain. *Nature Reviews. Neuroscience* 18, 4 (April 2017), 196–207. https://doi.org/10.1038/nrn.2017.14

[26] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9 ISSN: 0166-4115.

[27] Robert Horne, Steve Kelly, and Paul Sharp. 2015. Resistopalatography as an assistive technology for users with spinal cord injuries. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 4367–4370. https://doi.org/10.1109/EMBC.2015.7319362

[28] Xueliang Huo, Chihwen Cheng, and Maysam Ghovanloo. 2009. Evaluation of the Tongue Drive System by Individuals with High-Level Spinal Cord Injury. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference* 2009 (2009), 555–558. https://doi.org/10.1109/IEMBS.2009.5334555

[29] Patricio Isbej and Francisco J. Gutierrez. 2021. Gaze as a Navigation and Control Mechanism in Third-Person Shooter Video Games. In *HCI in Games: Experience Design and Game Mechanics*, Xiaowen Fang (Ed.). Springer International Publishing, Cham, 45–56. https://doi.org/10.1007/978-3-030-77277-2_4

[30] Ali Jafari, Nathanael Buswell, Maysam Ghovanloo, and Tinoosh Mohsenin. 2018. A Low-Power Wearable Stand-Alone Tongue Drive System for People With Severe Disabilities. *IEEE transactions on biomedical circuits and systems* 12, 1 (Feb. 2018), 58–67. https://doi.org/10.1109/TBCAS.2017.2757031

[31] Charlene Jennett, Anna Cox, Samira Dhoparee, Andrew Epps, Tim Tijs, and Alison Walton. 2008. Measuring and Defining the Experience of the Immersion in Games. *International Journal of Human-Computer Studies* 66 (Sept. 2008), 641–661. https://doi.org/10.1016/j.ijhcs.2008.04.004

[32] Yijing Jiang, Julia Kleinau, Till Max Eckroth, Eve Hoggan, Stefanie Mueller, and Michael Wessely. 2024. MouthIO: Fabricating Customizable Oral User Interfaces with Integrated Sensing and Actuation. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3654777.3676443

[33] Arata Jingu, Yudai Tanaka, and Pedro Lopes. 2023. LipIO: Enabling Lips as both Input and Output Surface. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 695, 14 pages. https://doi.org/10.1145/3544548.3580775

[34] Dominic Kao, Nick Ballou, Kathrin Gerling, Heiko Breitsohl, and Sebastian Deterding. 2024. How does Juicy Game Feedback Motivate? Testing Curiosity, Competence, and Effectance. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3613904.3642656

[35] Sukran Karaosmanoglu, Sebastian Cmentowski, Lennart E. Nacke, and Frank Steinicke. 2024. Born to Run, Programmed to Play: Mapping the Extended Reality Exergames Landscape. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–28. https://doi.org/10.1145/3613904.3642124

[36] Mohamed Kari and Christian Holz. 2023. HandyCast: Phone-based Bimanual Input for Virtual Reality in Mobile and Space-Constrained Settings via Pose-and-Touch Transfer. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3544548.3580677

[37] Panagiotis Kourtesis, Evgenia Giatzoglou, Panagiotis Vorias, Katerina Gounari, Eleni Orfanidou, and Chrysanthi Nega. 2025. Examination of Eye-Tracking, Head-Gaze, and Controller-Based Ray-Casting in TMT-VR: Performance and Usability Across Adulthood. *Multimodal Technologies and Interaction* 9 (July 2025). https://doi.org/10.3390/mti9080076

[38] G. Krishnamurthy and M. Ghovanloo. 2006. Tongue drive: a tongue operated magnetic sensor based wireless assistive technology for people with severe disabilities. In *2006 IEEE International Symposium on Circuits and Systems*. 4 pp.–. https://doi.org/10.1109/ISCAS.2006.1693892

[39] Devendra Kumar, Arnav Kothiyal, Raj Kumar, C Hemantha, and Ramya Maranan. 2024. Random Forest approach optimized by the Grid Search process for predicting the dropout students. In *2024 International Conference on Innovations and Challenges in Emerging Technologies (ICICET)*. 1–6. https://doi.org/10.1109/ICICET59348.2024.10616372

[40] Yunfei Lai, Minghui Sun, and Zhuofeng Li. 2024. GazePuffer: Hands-Free Input Method Leveraging Puff Cheeks for VR. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. 331–341. https://doi.org/10.1109/VR58804.2024.00055 ISSN: 2642-5254.

[41] Richard Li, Jason Wu, and Thad Starner. 2019. TongueBoard: An Oral Interface for Subtle Input. In *Proceedings of the 10th Augmented Human International Conference 2019* (Reims, France) *(AH2019)*. Association for Computing Machinery, New York, NY, USA, Article 1, 9 pages. https://doi.org/10.1145/3311823.3311831

[42] Zheng Li, Ryan Robucci, Nilanjan Banerjee, and Chintan Patel. 2015. Tongue-n-cheek: non-contact tongue gesture recognition. In *Proceedings of the 14th International Conference on Information Processing in Sensor Networks* (Seattle, Washington) *(IPSN '15)*. Association for Computing Machinery, New York, NY, USA, 95–105. https://doi.org/10.1145/2737095.2737109

[43] Li Liu, Shuo Niu, Jingjing Ren, and Jingyuan Zhang. 2012. Tongible: a non-contact tongue-based interaction technique. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility* (Boulder, Colorado, USA) *(ASSETS '12)*. Association for Computing Machinery, New York, NY, USA, 233–234. https://doi.org/10.1145/2384916.2384969

[44] Steven G. Luke. 2017. Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods* 49, 4 (Aug. 2017), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

[45] Morten Lund, Hector Caltenco, Eugen Lontis, Henrik Christiensen, Bo Bentsen, and Lotte Struijk. 2009. A Framework for Mouse and Keyboard Emulation in a Tongue Control System. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference* 2009 (09 2009), 815–8. https://doi.org/10.1109/IEMBS.2009.5334055

[46] Kurt Manal, Benjamin Gillette, and Ira Lockwood. 2021. *The Mouth Mouse: An Intuitive Bluetooth Controller of Electronic Systems for Persons with Upper-Limb Impairment*. Medical Devices, Vol. 2021 Design of Medical Devices Conference. https://doi.org/10.1115/DMD2021-1044 _eprint: https://asmedigitalcollection.asme.org/BIOMED/proceedings-pdf/DMD2021/84812/V001T12A009/6805521/v001t12a009-dmd2021-1044.pdf.

[47] Roberto Merletti and Philip Parker. 2004. *Detection and Conditioning of the Surface EMG Signal*. 107–131. https://doi.org/10.1002/0471678384.ch5

[48] Masato Miyauchi, Takashi Kimura, and Takuya Nojima. 2012. Development of a non-contact tongue-motion acquisition system. In *Adjunct Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology* (Cambridge, Massachusetts, USA) *(UIST Adjunct Proceedings '12)*. Association for Computing Machinery, New York, NY, USA, 75–76. https://doi.org/10.1145/2380296.2380329

[49] Pedro Monteiro, Guilherme Gonçalves, Hugo Coelho, Miguel Melo, and Maximino Bessa. 2021. Hands-free interaction in immersive virtual reality: A systematic review. *IEEE Transactions on Visualization and Computer Graphics* 27, 5 (May 2021), 2702–2713. https://doi.org/10.1109/TVCG.2021.3067687

[50] T. Moritani, D. Stegeman, and R. Merletti. 2004. Basic Physiology and Biophysics of EMG Signal Generation. In *Electromyography* (1 ed.), Roberto Merletti and Philip Parker (Eds.). Wiley, 1–25. https://doi.org/10.1002/0471678384.ch1

[51] Florian Müller, Daniel Schmitt, Andrii Matviienko, Dominik Schön, Sebastian Günther, Thomas Kosch, and Martin Schmitz. 2023. TicTacToes: Assessing Toe Movements as an Input Modality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 520, 17 pages. https://doi.org/10.1145/3544548.3580954

[52] Lennart Erik Nacke, Michael Kalyn, Calvin Lough, and Regan Lee Mandryk. 2011. Biofeedback game design: using direct and indirect physiological control to enhance game interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 103–112. https://doi.org/10.1145/1978942.1978958

[53] Takuro Nakao, Yun Suen Pai, Megumi Isogai, Hideaki Kimata, and Kai Kunze. 2018. Make-a-face: a hands-free, non-intrusive device for tongue/mouth/cheek input using EMG. In *ACM SIGGRAPH 2018 Posters (SIGGRAPH '18)*. Association for Computing Machinery, New York, NY, USA, 1–2. https://doi.org/10.1145/3230744.3230784

[54] Diego Navarro, Veronica Sundstedt, and Valeria Garro. 2021. Biofeedback Methods in Entertainment Video Games: A Review of Physiological Interaction Techniques. *Proc. ACM Hum.-Comput. Interact.* 5, CHI PLAY (Oct. 2021), 268:1–268:32. https://doi.org/10.1145/3474695

[55] Phuc Nguyen, Nam Bui, Anh Nguyen, Hoang Truong, Abhijit Suresh, Matt Whitlock, Duy Pham, Thang Dinh, and Tam Vu. 2018. TYTH-Typing On Your Teeth: Tongue-Teeth Localization for Human-Computer Interface. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, Munich Germany, 269–282. https://doi.org/10.1145/3210240.3210322

[56] Wolfgang Nutt, Christian Arlanch, Silvio Nigg, and Gerhard Staufert. 1998. Tongue-mouse for quadriplegics. *Journal of Micromechanics and Microengineering* 8, 2 (June 1998), 155. https://doi.org/10.1088/0960-1317/8/2/028

[57] Rakesh Patibanda, Chris Hill, Aryan Saini, Xiang Li, Yuzheng Chen, Andrii Matviienko, Jarrod Knibbe, Elise van den Hoven, and Florian 'Floyd' Mueller. 2023. Auto-Paizo Games: Towards Understanding the Design of Games That Aim to Unify a Player's Physical Body and the Virtual World. *Proc. ACM Hum.-Comput. Interact.* 7, CHI PLAY (Oct. 2023), 408:893–408:918. https://doi.org/10.1145/3611054

[58] Siyou Pei. 2025. *Embodied Extended Reality: Accelerating the Sensing-Interaction Cycle*. Ph. D. Dissertation. UCLA. https://escholarship.org/uc/item/34q649tw

[59] Angkoon Phinyomark, Rami N. Khushaba, and Erik Scheme. 2018. Feature Extraction and Selection for Myoelectric Control Based on Wearable EMG Sensors. *Sensors* 18, 5 (2018). https://doi.org/10.3390/s18051615

[60] Martin Pichlmair and Mads Johansen. 2022. Designing Game Feel: A Survey. *IEEE Transactions on Games* 14, 2 (June 2022), 138–152. https://doi.org/10.1109/TG.2021.3072241

[61] Gonzalo Ramos, Matthew Boulos, and Ravin Balakrishnan. 2004. Pressure widgets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Vienna Austria, 487–494. https://doi.org/10.1145/985692.985754

[62] Raquel Robinson, Katelyn Wiley, Amir Rezaeivahdati, Madison Klarkowski, and Regan L. Mandryk. 2020. "Let's Get Physiological, Physiological!": A Systematic Review of Affective Gaming. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '20)*. Association for Computing Machinery, New York, NY, USA, 132–147. https://doi.org/10.1145/3410404.3414227

[63] Chris Salem and Shumin Zhai. 1997. An isometric tongue pointing device. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '97)*. Association for Computing Machinery, New York, NY, USA, 538–539. https://doi.org/10.1145/258549.259021

[64] T. Saponas, Daniel Kelly, Babak Parviz, and Desney Tan. 2009. Optically Sensing Tongue Gestures for Computer Input. 177–180. https://doi.org/10.1145/1622176.1622209

[65] Makoto Sasaki, Kohei Onishi, Dimitar Stefanov, Katsuhiro Kamata, Atsushi Nakayama, Masahiro Yoshikawa, and Goro Obinata. 2016. Tongue interface based on surface EMG signals of suprahyoid muscles. *ROBOMECH Journal* 3, 1 (Dec. 2016), 9. https://doi.org/10.1186/s40648-016-0048-0

[66] Lijuan Shi, Xiong Peng, Jian Zhao, Zhejun Kuang, Tianbo An, and Liu Wang. 2023. Non-Invasive Intraoral Stand-Alone Tongue Control System Based on RSIC-V Edge Computing. *Applied Sciences* 13, 17 (Jan. 2023), 9490. https://doi.org/10.3390/app13179490 Number: 17 Publisher: Multidisciplinary Digital Publishing Institute.

[67] Tanay Singhal and Oliver Schneider. 2021. Juicy Haptic Design: Vibrotactile Embellishments Can Improve Player Experience in Games. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3411764.3445463

[68] Ronit Slyper, Jill Lehman, Jodi Forlizzi, and Jessica Hodgins. 2011. A tongue input device for creating conversations. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*. ACM, Santa Barbara California USA, 117–126. https://doi.org/10.1145/2047196.2047210

[69] Misha Sra, Xuhai Xu, and Pattie Maes. 2018. BreathVR: Leveraging Breathing as a Directly Controlled Interface for Virtual Reality Games. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173914

[70] Rujia Sun, Xiaohe Zhou, Benjamin Steeper, Ruidong Zhang, Sicheng Yin, Ke Li, Shengzhang Wu, Sam Tilsen, Francois Guimbretiere, and Cheng Zhang. 2023. EchoNose: Sensing Mouth, Breathing and Tongue Gestures inside Oral Cavity using a Non-contact Nose Interface. In *Proceedings of the 2023 ACM International Symposium on Wearable Computers* (Cancun, Quintana Roo, Mexico) *(ISWC '23)*. Association for Computing Machinery, New York, NY, USA, 22–26. https://doi.org/10.1145/3594738.3611358

[71] Steve Swink. 2008. *Game Feel: A Game Designer's Guide to Virtual Sensation*. CRC Press, Boca Raton. https://doi.org/10.1201/9781482267334

[72] Graham Wilson, Jamie McCready, Euan Freeman, Florian Mathis, Harvey Russell, and Mark McGill. 2025. InterFACE: Establishing a Facial Action Unit Input Vocabulary for Hands-Free Extended Reality Interactions, From VR Gaming to AR Web Browsing. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 607, 23 pages. https://doi.org/10.1145/3706598.3713694

[73] Dennis Wolf, Jan Gugenheimer, Marco Combosch, and Enrico Rukzio. 2020. Understanding the Heisenberg Effect of Spatial Interaction: A Selection Induced Error for Spatially Tracked Input Devices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3313831.3376876

[74] Xuhai Xu, Chun Yu, Anind K. Dey, and Jennifer Mankoff. 2019. Clench Interface: Novel Biting Input Techniques. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–12. https://doi.org/10.1145/3290605.3300505

[75] Shumin Zhai. 1998. User Performance in Relation to 3D Input Device Design. *ACM Siggraph Computer Graphics* 32 (11 1998). https://doi.org/10.1145/307710.307728

[76] Fengze Zhang, Yunxiang Zhang, Xi Peng, Sky Achitoff, Paul M. Torrens, and Qi Sun. 2024. May the Force Be with You: Dexterous Finger Force-Aware VR Interface. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Bellevue, WA, USA, 455–464. https://doi.org/10.1109/ISMAR62088.2024.00060

[77] Qiao Zhang, Shyamnath Gollakota, Ben Taskar, and Raj P.N. Rao. 2014. Non-intrusive tongue machine interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Toronto Ontario Canada, 2555–2558. https://doi.org/10.1145/2556288.2556981

[78] Daniel Zielasko, Sebastian Freitag, Dominik Rausch, Yuen C. Law, Benjamin Weyers, and Torsten W. Kuhlen. 2015. BlowClick: A Non-Verbal Vocal Input Metaphor for Clicking. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction (SUI '15)*. Association for Computing Machinery, New York, NY, USA, 20–23. https://doi.org/10.1145/2788940.2788953

A Tongue Input Design Space

Table 6. Tongue-based systems categorized by position and pressure resolution. † intra-oral, * extra-oral. The shaded cell represents a passive speculative category where natural tongue movements can be measured to detect user context, intent to act, and other implicit signals. This category is outside the scope of this paper.

Position / Pressure	None	Discrete	Continuous
None	—	—	*TonguePoint*† [63]
Discrete	*TongueTap** [21] *EchoNose** [70] *Tongue-n-Cheek** [42] *Tongue-in-Cheek** [22] *Tongue-Touch Keypad (TTK)*† [68] *Optopalatography*† [64] *Tongue Joystick*† [68] *TongueSee** [77] *Tongue Interface** [65] *Tongible** [43] *TongueBoard** [41] *Tongue Control System*† [66] *Resistopalatography*† [27] *Glossokinetic Sensing*† [24] *On the Tip of My Tongue** [12] *Noninvasive tongue-motion** [13] *MouthIO** [32]	***This Paper****	—
Continuous	*Tongue Drive System (TDS)*†* [38] *Inductive Sensing*† [45] *SITA** [48]	—	*MouthMouse*† [46] *Tongue-Mouse*† [56]

B Study 2 Interview Probes

The following probes were used in Study 2. Participants were not shown these questions in advance. Immediate elicitations were conducted verbally after each game block; the reflective interview followed after both games were completed. Follow-up questions were asked at the researcher's discretion to clarify or deepen participant responses.

Immediate Elicitation — Frog Game

(1) "How did playing feel in general?"
(2) "How would you describe the physical sensation of using your tongue to aim and snap?"
(3) "Did the tension you felt in the game match the tension in your mouth?"
(4) "In the version where the wasp shook: How did it feel to know that *your* body tension was making the enemy vibrate?"
(5) "Did you feel more responsible for the shaking when you saw it on the frog or on the wasp?"

Immediate Elicitation — Kart Game

(1) "How did playing feel in general?"
(2) "How did it feel to drive using your tongue as the gas pedal?"
(3) "How did the vibration affect your sense of speed?"
(4) "In the version where the world shook: How did it feel that your physical effort was causing the whole environment to shake?"

(5) "Which felt more natural: feeling the car shake and jump, or seeing the world shake? Why?"

Reflective Interview (Post-Both-Games)

(1) "After playing both games, how did you feel overall, and how, if at all, did that change across the session?"
(2) "You knew from the start that the shaking was caused by your own muscles as measured by the device. How did it feel to see your effort in the game?"
(3) "Did seeing the world vibrate make you feel more powerful — like you were affecting reality — or less stable, like you were losing control?"

Received February 2026; revised June 2026; accepted July 2026