Design of a Continuous-Output Switch-Access Interface for Users Whose Structured Motor Signature Is Sustained Engagement with Modulation: Detection Framework and Channel-Capacity Argument

Cara Jade Catalano

Abstract

Objective. Switch-access interfaces for augmentative and alternative communication (AAC) assume the user can produce discrete tap-like activations. When a candidate user cannot, the standard clinical inference is that voluntary motor control sufficient for switch access is absent. This paper addresses a different *structured motor signature*—sustained engagement with within-engagement modulation (canonically, sustained finger-pad contact with within-contact force modulation)—as a viable control signal for a redesigned interface, under the taxonomy in which sustained-hold-with-modulation is distinguishable from pure-tap activation, undifferentiated resting contact, and rhythmic-tap (Morse-style) sequences. *Methods.* The paper specifies a detection framework for identifying the sustained-contact-with-modulation signature from video-and-audio recordings of pad interaction: MediaPipe-based fingertip landmark extraction yields a continuous contact-state estimate; two independent operationalizations of "press" are defined (velocity-reversal kinematic; contact-interval geometric); librosa spectral-flux onset detection provides an independent acoustic channel; and the sustained-contact ratio $R_{SC} = N_{onsets} / N_{presses}$ quantifies how many acoustic events occur per press, with $R_{SC} \approx 1$ the tap-like null and $R_{SC} >> 1$ the sustained-hold-with-modulation regime. *Proposed detection framework.* For recordings made while a device emits its own periodic audio (a drum-machine loop, a metronome, any stereotyped acoustic pattern), a loop-subtraction technique is specified that partitions in-contact audio onsets against the observed loop grid and yields a strict lower bound on subject-initiated events. *Channel capacity.* A structural argument grounds the interface specification: A binary-closure switch transmits at most one bit per contact event, whereas a force-sensing pad sampling at f_s with b-bit resolution encodes $b \cdot f_s \cdot T$ bits per contact of duration T. At 8 bits, 100 Hz, and T = 1 s, the nominal capacity ratio is 800:1; after Shannon–Hartley correction for realistic noise (20 dB SNR) and motor bandwidth (10 Hz) the effective ratio is ≈ 66:1, falling to ≈ 17:1 only under aggressively pessimistic assumptions (10 dB SNR, 5 Hz bandwidth). *Interface contribution.* A continuous-output interface is specified, with the force-sensing pad (capacitive or piezoresistive) as the canonical transducer, to encode engagement onset, engagement duration, and within-engagement modulation, with a signal-processing pipeline that maps these dimensions onto a small AAC command alphabet and a proposed binary (yes/no) bootstrap protocol derived from modulated-vs-silent sustained holds. The non-thresholded output (≥ 100 Hz sampling, no in-device binarization) is the critical architectural choice. *Deployment scenarios.* Deployment is discussed separately for static-etiology populations (hypoxic-anoxic brain injury, post-status-epilepticus syndromes, severe cerebral palsy with preserved tactile response), where the interface may be the first channel with sufficient bandwidth to register the user's motor signature across any assessment in their lifespan, and progressive populations (notably late-stage amyotrophic lateral sclerosis), where it serves as a bridge channel that maintains direct-selection AAC usability during the period when tap fidelity degrades before sustained-contact modulation does. Additional deployment scenarios address users whose primary constraint is distal upper-extremity weakness or atypical wrist kinematics (contractures, hypermobility, congenital variants) and integration with standard clinical assessment batteries (CRS-R, Motor Behavior Tool, AAC timing protocols). The framework is stated in sensor-agnostic form: The force-sensing pad is the canonical transducer, but the channel-capacity argument and detection pipeline transfer to non-contact transducers (capacitive proximity, infrared / time-of-flight optical, millimeter-wave radar) for users for whom physical contact is contraindicated by tactile hypersensitivity, skin compromise, or pressure-sore risk, and to bio-signal transducers (surface electromyography from residual musculature or targeted muscle reinnervation sites) for users with upper-extremity amputation or prosthetic-limb deployment, for whom the multi-channel structure of the EMG record expands the achievable symbol alphabet by one axis per electrode. Combat blast-injury survivors (burn scarring, upper-extremity amputation, and blast-induced TBI frequently co-occurring in the same patient) are discussed as a cross-cutting clinical population for whom preserved visuomotor coordination typically renders the non-contact variant's distance-dimension vocabulary gain immediately usable. Epilepsy and rhythmic-involuntary-motor populations are treated as a first-class deployment scenario, including the severe case of continuous subclinical (electrographic) seizure activity in which involuntary micro-twitches are superimposed on every sustained-contact interval; frequency-domain separation of involuntary rhythmic contamination from intentional modulation, supported by a mandatory per-session diagnostic baseline and a ≥ 25 Hz Nyquist sampling floor that covers the full 1–12 Hz involuntary band, is the operational key to serving this cohort. *Scope.* The paper is a design and methods contribution; empirical validation of the detection framework and the bootstrap protocol on a recruited cohort is flagged as future work, and prevalence of the target signature in candidate populations is an open empirical problem.

Keywords. sustained contact, switch access, AAC, brain-computer interface, force-sensing, non-contact transduction, proximity sensing, myoelectric control, surface electromyography, prosthetic-limb interface, combat-related limb loss, polytrauma, veteran rehabilitation, blast injury, epilepsy, seizure, involuntary motor, rhythmic motor, myoclonus, continuous subclinical seizure activity, interface specification, information capacity, AAC bootstrap, MediaPipe, rehabilitation engineering, Shannon–Hartley capacity, tactile contraindications.

---

1. Introduction

1.1 The Tap-Activation Assumption

For a subset of users with severe motor disability, standard switch-access assessment cannot detect communication potential that is in fact present: The user's voluntary motor output takes the form of sustained finger contact with within-hold force modulation rather than discrete taps, and binary-closure switches by construction discard this signature. Absence of tap-based switch evidence is not evidence of absent voluntary motor control; it is evidence that the assessment channel has a capacity of one bit while the user's motor output lives on a higher-dimensional channel. This paper specifies a continuous-output switch-access interface—a force-sensing pad as the canonical instantiation, with non-contact and bio-signal transducer variants developed in Section 2.4 and Section 5.8.5–Section 5.8.6—and a detection framework designed to recover that signature.

Switch-access AAC is the low-resource lingua franca of assistive communication. The user closes a switch; the switch emits a binary event; a scanning interface uses the event to select from a menu. The approach extends to eye-blink switches, sip-and-puff, capacitive touch pads, and most BCI front ends. In nearly every instance the switch is architected around *momentary closure*: The event is the transition from open to closed, not the duration of closure or any modulation within it.

This architecture presumes the user can (i) generate a brief, well-formed, repeatable activation, (ii) terminate it on demand, and (iii) separate adjacent activations in time. When clinical assessment finds no reliable tap—as in severe hypoxic-anoxic brain injury (HAI), post-status-epilepticus syndromes, or severe cerebral palsy with preserved tactile response—the common inference is that switch access itself is not viable for that user, and the cost of this inference is borne most heavily by users in whom the tap-failure is static: Every binary-switch assessment administered across their lifespan projects a higher-dimensional motor signature to a single bit, and the user is recorded, across years or decades, as having no reliable voluntary output. In progressive neuromuscular disease, most notably late-stage amyotrophic lateral sclerosis (ALS), the same inference is drawn at the disease stage at which tap reliability crosses below clinical threshold, even though sustained-contact motor output with within-hold force modulation is typically still well-preserved at that stage and remains available for months to a year or longer [10, 11]. In either case the interface failure is the same—a binary channel cannot register a higher-dimensional signature—but the deployment story differs (Section 5.8).

1.2 The Sustained-Contact Alternative

A motor signature tap-based architecture cannot capture: The user places a finger on the target and holds, modulating pressure or position within the hold rather than releasing and re-striking. A binary switch registers this as a single activation at the onset of contact and silence thereafter, or, if any micro-release occurs, as a noisy burst of transitions. The within-contact modulation—the actual information-bearing structure of the gesture—is discarded at the interface boundary.

If this signature exists in the candidate population, a tap-based assessment records its presence as equivalent to its absence. A user whose motor output lives on the within-contact modulation dimension cannot be distinguished from a user with no structured motor output of *any* kind. Absence of switch-access evidence becomes, in the clinical record, evidence of absence.

1.3 Contribution of This Work

This paper makes three structural contributions.

Detection framework. A proposed methodology for quantifying sustained-contact-with-modulation signatures from video-and-audio recordings of pad interaction. The pipeline combines MediaPipe fingertip-landmark extraction with librosa spectral-flux onset detection, defines two independent operational press-event counts (a velocity-reversal kinematic definition and a contact-interval geometric definition), and reduces to the single scalar $R_{SC} = N_{onsets} / N_{presses}$, whose null value under pure-tap activation is 1 by construction and whose alternative hypothesis (sustained hold with within-contact modulation) is $R_{SC} >> 1$. For recordings in which the target device emits its own periodic acoustic pattern (drum-machine loop, metronome, pacing stimulus), a loop-subtraction technique is specified that partitions in-contact onsets against the observed loop grid and yields a strict lower bound on subject-initiated activity.

Structural argument. A continuous-output interface encoding engagement onset, engagement duration, and within-engagement modulation is required to preserve the information content of this signature. A binary-closure switch cannot. The argument is information-theoretic (Section 5.6): Channel-capacity math gives the capacity ratio directly, rendering *information-destructive* a quantitative label rather than a rhetorical one. The nominal capacity ratio is 800:1 per 1-second contact at 8-bit, 100 Hz sampling; the Shannon–Hartley-corrected ratio under realistic pad electronics is ≈ 66:1; under aggressively pessimistic assumptions the ratio is still ≈ 17:1. The argument does not depend on prevalence of the target signature in any population.

Deployment-scenario separation. The paper carries the design through eight deployment scenarios in Section 5.8, each making distinct demands on the signal-processing pipeline and clinical workflow: static-etiology users (HAI, post-status-epilepticus, severe CP with preserved tactile response) for whom the interface is a first channel and the bootstrap is a code-book invention from zero (Section 5.8.1); progressive neuromuscular disease (ALS) for whom it is a bridge channel with transfer bootstrap and scheduled re-calibration (Section 5.8.2); distal upper-extremity weakness and atypical wrist kinematics (Section 5.8.3); integration with standard clinical assessment batteries (Section 5.8.4); a non-contact transducer variant for users with tactile contraindications, exposing an additional distance dimension (Section 5.8.5); prosthesis-mediated and direct myoelectric deployments for users with upper-extremity amputation, in which multi-channel EMG expands the symbol alphabet (Section 5.8.6); combat blast-injury survivors as a cross-cutting population drawing on Section 5.8.5 and Section 5.8.6 simultaneously (Section 5.8.7); and epilepsy and rhythmic-involuntary-motor populations, including the severe case of continuous subclinical seizure activity, for whom frequency-domain separation of involuntary rhythmic contamination from intentional modulation is the operational key (Section 5.8.8).

This is a design and methods paper. Empirical validation of the detection framework on a recruited cohort, behavioral validation of the bootstrap protocol, and the prevalence question for candidate clinical populations are flagged as open empirical problems in Section 6 and are outside the scope of the present contribution.

1.4 Reading Guide

Readers may locate the sections most relevant to their role:

- **AAC clinicians and SLPs.** Sections 1 (problem statement), 2 (background on switch-access and force-sensing precedents), 5.4–5.5 (interface behavior specification and the binary-channel bootstrap protocol), 5.8 (clinical deployment across eight scenarios: static-etiology Section 5.8.1, progressive-disease Section 5.8.2, distal upper-extremity weakness and atypical wrist kinematics Section 5.8.3, CRS-R integration Section 5.8.4, non-contact and myoelectric variants Section 5.8.5–5.8.6, combat blast-injury cross-cutting population Section 5.8.7, and epilepsy and rhythmic-involuntary-motor populations Section 5.8.8), and 6 (design limitations, open empirical questions, and proposed validation studies).
- **AAC users' families and advocates.** Sections 1.1–1.2 (the assessment-channel-capacity problem stated plainly), 5.8 (what deployment looks like for static-etiology and progressive-disease users), and 7 (what this work is and is not).
- **Rehabilitation engineers and HCI designers.** Sections 2.4 (engagement-state generalization across contact, non-contact, and bio-signal transducer classes), 4 (detection framework, including press operationalizations, R_{SC}, a worked numerical example at Section 4.5.2, and loop subtraction), 5.2–5.3 (sensor and signal-processing specifications), 5.6–5.7 (channel-capacity math and Shannon–Hartley envelope), Section 5.8.5 (non-contact transducer variants for users with tactile contraindications, including the vocabulary-expansion implications of an added distance dimension), Section 5.8.6 (prosthesis-mediated and direct myoelectric deployments, with multi-channel EMG expanding the symbol alphabet by one axis per electrode), Section 5.8.7 (combat blast-injury deployment, including dual-transducer staging across healing phases and integration with VA/DoD rehabilitation pipelines), and Section 5.8.8 (epilepsy and rhythmic-involuntary-motor populations, including frequency-domain separation of involuntary rhythmic contamination from intentional modulation, per-user power-spectral baselining, and the ≥ 25 Hz Nyquist sampling requirement that covers the full 1–12 Hz involuntary band).
- **Statisticians and methods reviewers.** Section 4 (R_{SC} as a scalar statistic with a data-free null; contact-interval press definition; loop-subtraction lower bound), Section 6.4 (proposed validation study designs), and Appendix A (reproducibility notes).

---

2. Background

2.1 Switch-Access Interfaces

Switch-based AAC emerged from the 1970s–80s assistive-communication literature, canonical in Beukelman and Mirenda [1]. The interface presumes a binary input paired with a scanning selection method (row–column scan, auditory scan, directed scan). Force-sensing variants exist—BIGmack and successors register contact force—but force is typically thresholded at device firmware into a binary event before reaching the application. The force channel is collapsed to binary at the earliest possible stage.

2.2 Continuous-Signal Inputs in Clinical Assistive Technology

Continuous-signal inputs are established in prosthetics (myoelectric control), pressure-mapping wheelchairs, and sip-and-puff interfaces, where the transducer output is a graded signal rather than a trigger. These precedents confirm that continuous inputs are mechanically and clinically feasible; they have not yet extended systematically into switch-access AAC for users without tap capability.

These precedents span distinct physical quantities—contact force, interface pressure, myoelectric voltage—unified not by modality but by the continuous, graded character of the transduced signal; the present work treats force-on-pad as one instantiation of that broader class, with the generalization made explicit in Section 2.4.

2.3 Agency Detection in HAI

Assessment batteries for severe HAI include the Coma Recovery Scale–Revised (CRS-R) [2] and Motor Behavior Tool (MBT) [3]. These scales record voluntary motor output categorically (present / absent / ambiguous) and are sensitive to reliable binary response. They are not designed to capture modulation within a held contact, and cannot refute the null hypothesis "no voluntary motor control" when a candidate user's actual motor output is sustained rather than phasic.

2.4 Engagement-State Generalization: Contact Is an Instantiation

The remainder of this paper is written around a force-sensing pad as the canonical transducer, because the clinical populations of greatest interest in Section 5.8 benefit from a tactile anchor and because pressure-transduction electronics are inexpensive, robust, and well characterized. The underlying information-theoretic argument, however, is not about pressure specifically. It is about a continuous, adequately sampled, sufficiently-high-resolution signal that registers a time-varying *engagement state* between the user's effector and the interface—where, in the force-on-pad case, "engagement state" reduces to applied normal force, and the analogous quantity for a contactless transducer is fingertip distance to the sensor plane, fingertip occupancy of a capacitive field, or returned-signal amplitude from an optical or radar target.

Three broad transducer classes instantiate the engagement-state abstraction. The first is the *contact* class: force-sensing pads (the canonical case of Section 5.2), pressure-sensitive switches retaining graded output, and instrumented touch surfaces retaining sub-threshold response. The second is the *non-contact* class: capacitive proximity sensors (Theremin-style field measurement; the air-touch DJ controllers in which a hovering hand modulates a continuous audio parameter are a familiar consumer instance), infrared reflective and time-of-flight optical rangers, structured-light depth cameras used standalone from the MediaPipe video pipeline of Section 4, and millimeter-wave radar front ends (Google Soli and its successors). The third is the *bio-signal* class: surface electromyography (sEMG) electrodes placed over residual musculature in users with upper-extremity amputation, targeted-muscle-reinnervation (TMR) sites, and (in research contexts) intramuscular or implanted electrodes delivering higher signal-to-noise recordings. Each class delivers a continuous scalar or low-dimensional vector per frame at rates at or above the 100 Hz assumed in Section 5.2; each admits the same onset, duration, and within-engagement modulation decomposition as the force-pad case, with units and noise characteristics substituted. For the bio-signal class the scalar is the rectified and envelope-filtered EMG amplitude per electrode, and the multi-channel generalization replaces the scalar with a vector whose dimensionality is the electrode count—a structural difference from the other two classes that the symbol alphabet of Section 5.4 exploits in Section 5.8.6. The detection framework of Section 4, the command alphabet of Section 5.4, the bootstrap protocol of Section 5.5, the channel-capacity argument of Section 5.6, and the loop-subtraction technique of Section 4.7 are stated in terms that remain valid under this substitution. The translation is taken up concretely in Section 5.8.5 (non-contact deployment) and Section 5.8.6 (prosthesis users and direct myoelectric deployment), where the non-force-pad variants' clinical roles are discussed alongside the force-pad deployments of Section 5.8.1–Section 5.8.4. A fourth transducer class—electrocorticographic (ECoG) and intracortical brain-computer interfaces producing continuous neural signals—is implicitly compatible with the engagement-state generalization but falls outside the scope of this paper, which is written around non-invasive transducers deployable in standard rehabilitation and AAC clinical contexts; the channel-capacity argument of Section 5.6 carries through to such signals with the substituted noise and bandwidth parameters.

A single term is used throughout the rest of the paper: *contact-state* signals and quantities refer to the force-pad case and its direct analogues on a physical transducer; *engagement-state* is used when the statement is intended to cover any of the three transducer classes. Where only the contact term appears, the extension to the non-contact and bio-signal cases follows by the substitution described in this subsection.

3. Motor-Signature Taxonomy

The detection framework is motivated by a four-class taxonomy of pad-interaction motor output. The taxonomy is stated prior to the detection pipeline so that the operational definitions in Section 4 can be read as implementations of specific distinctions within this taxonomy rather than as arbitrary thresholds.

- **Pure tap.** Contact initiation is followed by release before the next initiation. Each contact generates at most one acoustic event on a pressure-triggered device. $R_{SC} \approx 1$ by construction.
- **Sustained contact without modulation (silent hold).** Contact is maintained for an extended duration without within-contact force or position change sufficient to retrigger the device. Contact duration is long; acoustic events per contact approach zero on a pressure-triggered device, though contact is directly measurable on a force-sensing pad.
- **Sustained contact with modulation.** Contact is maintained for an extended duration, and within that contact the user produces force deflections, finger slides, or adjacent-cell re-activations that retrigger the device without a full lift. Acoustic events per contact are well above 1; the information-bearing content of the gesture lives in the within-contact modulation series. Voluntary force modulation is bandlimited: Deliberate finger-force adjustment is typically produced below ~ 8 Hz, with most volitional content concentrated below 5 Hz. Modulation energy above ~ 8 Hz overlaps the frequency bands of involuntary motor phenomena (physiological tremor, clonus, myoclonic activity) and is treated as the confound rather than the signal (see the involuntary-motor paragraph below and Section 5.8.8).
- **Rhythmic or repetitive tap (twitch burst).** Contact takes the form of brief, discrete initiations (tap duration <300 ms, contact-to-release) produced in structured sequences. The relevant timing channel has three operationally distinct intervals: tap duration itself (<300 ms); intra-burst interval between consecutive taps within a sequence (≥200 ms, per-user calibrated from observed volitional rate, with no universal upper bound); and boundary pause separating sequences (≥3× the user's mean intra-burst interval, following the Morse convention for word/symbol delimiters—robust across users regardless of individual tapping speed). Intervals below 200 ms fall into involuntary-motor frequency ranges (clonus 5–8 Hz, physiological tremor 4–12 Hz) and are screened against per the intentional-modulation requirement below. $R_{SC} \approx 1$ per contact as in pure tap, but information is encoded in burst rate, count, and inter-tap interval variance—a continuous timing channel that standard switch-access collapses by treating each contact as an isolated binary event. Morse-code AAC input, long used by individuals with severe motor impairment, is the established linguistic precedent: vocabulary transmitted entirely through tap-timing structure—with the additional clinical relevance that populations carrying prior Morse fluency from military signals training or amateur-radio service (see Section 5.8.7) acquire the vocabulary at zero training cost. Users with intentional twitch capability but limited sustained-contact smoothness (certain spastic, clonic, or distal-weakness presentations) fall into this class.

Two distinct continuous channels are therefore in scope: within-contact modulation (sustained-contact class) and inter-contact interval structure (rhythmic-tap class, with Morse as linguistic precedent). Both are collapsed by binary-event switch-access; both carry vocabulary-bearing information under the framework of Sections 4–5.

The sustained-contact-with-modulation class is the focus of this paper. The detection framework in Section 4 is designed to identify it from video-plus-audio recordings; the interface specification in Section 5 is designed to capture it without the information loss inherent in binary closure. The silent-hold class is acknowledged as a second legitimate motor state: Any interface that admits only modulated holds would misclassify a silent hold as non-intent, a failure mode explicitly guarded against in the command alphabet of Section 5.4.

The taxonomy presumes modulation is intentional rather than the product of involuntary motor phenomena, which span a clinically heterogeneous range: clonus (rhythmic 5–8 Hz oscillation at velocity-sensitive stretch), physiological tremor (4–12 Hz), clonic seizure activity (typically 2–3 Hz rhythmic jerking), myoclonic jerks (brief 1–5 Hz discrete contractions, generalized or multifocal), interictal and post-ictal myoclonus, post-anoxic (Lance-Adams) myoclonus, progressive myoclonic epilepsies (Unverricht-Lundborg, Lafora body disease, neuronal ceroid lipofuscinoses, Ramsay Hunt syndrome), dystonia, athetoid overflow, spasticity-driven fasciculation, and—in the clinically severe case—continuous subclinical (electrographic) seizure activity producing many mini-seizures per second, in which involuntary micro-twitches are superimposed on every sustained-contact interval and can occur even within a single intentional press. Each has a characteristic frequency signature and task-contingency profile. Operational separation of intentional modulation from involuntary contamination relies on task-contingent contrast—modulation statistics during caregiver-prompted engagement versus ambient periods—combined with frequency-domain screening against the per-user involuntary spectrum. This is a required per-user screening step for populations with prominent involuntary motor phenomena, and for epilepsy and rhythmic-involuntary-motor populations is developed as a full deployment subsection (Section 5.8.8).

---

4. Detection Framework

4.1 Video Preprocessing

Input recordings are re-encoded to H.264 MP4 at native frame rate using FFmpeg; audio is extracted as 22.05 kHz mono WAV. No visual enhancement is applied. For landmark extraction, frames are decimated to 10 fps (sampling period $\Delta t = 0.1$ s). This rate resolves the slowest characteristic motor events of the sustained-contact-with-modulation signature with several frames per event while keeping the analysis tractable on commodity hardware; higher sampling is permissible and does not change the framework, but is not required for the primary statistic.

4.2 Hand Landmark Extraction

Per-frame 21-point hand landmarks are extracted with the MediaPipe Hands solution [4] (model complexity 1, minimum detection confidence 0.5, minimum tracking confidence 0.5). The index fingertip (landmark 8) and wrist (landmark 0) are the primary points used downstream. Hand-detection rate is expected to approach > 95% of decimated frames under unobstructed camera framing with adequate ambient illumination, and to degrade under partial occlusion (camera angle, own-body occlusion by a user's contralateral hand or wrist, or device body intervening between the fingertip and the camera). Detection-rate sensitivity is addressed in Section 6 as an open empirical question for any validation cohort. Undetected frames are left as missing values; no interpolation is applied.

4.3 Press Detection (Velocity-Reversal Definition)

Let $y(t) \in [0, 1]$ denote the normalized vertical image position of the index fingertip, with y increasing downward. The first-order discrete time derivative at frame i is

> $$v_y(i) = [y(i) - y(i-1)] / \Delta t \quad \approx \quad dy/dt \text{ evaluated at } t = i\Delta t.$$

A press candidate at frame i is flagged when

> $$(1) \quad v_y(i) > \tau_v \quad \text{(downward velocity exceeds threshold).}$$

A second condition upgrades confidence:

> $$(2) \quad \exists\, j \in \{1, 2\} \text{ such that } v_y(i + j) < \tfrac{1}{2} v_y(i) \quad \text{(deceleration / reversal).}$$

Condition (2) is a discrete proxy for a sufficiently negative second derivative d^2y/dt^2 over $[i\,\Delta t, (i+2)\Delta t]$—the kinematic signature of contact arresting downward motion. A three-frame refractory period after each detection prevents double-counting of a single gesture. The threshold τ_v is calibrated against a manually-labeled subset of any given corpus, and its sensitivity is assessed via a standard parameter-sweep analysis within a future validation study.

Each press candidate is tagged with a confidence level. *High* requires both deceleration/reversal (condition 2) and audio-onset co-occurrence within ±1 frame (±100 ms at 10 fps). *Medium* requires one of the two (deceleration/reversal or audio co-occurrence, but not both). *Low* requires a velocity spike (condition 1) alone, with neither deceleration/reversal nor audio co-occurrence. All R_{SC} and inter-press-interval statistics use the inclusive count (high + medium + low), to preserve a strictly kinematic operational definition that does not condition on audio co-occurrence—which would make R_{SC}'s numerator and denominator non-independent and bias the ratio downward.

4.4 Audio Onset Detection

Audio onsets are computed with librosa's `onset_detect` [5] (spectral-flux novelty, backtrack enabled). A pressure-triggered pad device produces a percussive acoustic event at each capacitive-pad activation, so onset count is a reasonable proxy for the number of pad-surface activation events, independent of the video-derived fingertip trajectory. Because onset detection uses a different modality and algorithm than press detection, their co-occurrence (or lack thereof) is an independent check on interpretation.

Source decomposition of audio onsets. Raw onset counts include events from three distinct sources: (i) direct pad activations by the user; (ii) in sessions with an active internal loop pattern on the target device, a periodic sequence independent of pad activation; (iii) ambient sound (speech, object handling, room noise) and device sustain or resonance from a single contact. The contact-interval analysis in Section 4.5.1 separates (i) from (ii)+(iii) by co-registering each audio onset against the video-derived fingertip-in-pad-region time series. Onsets falling inside a sustained-contact interval (±100 ms tolerance) are attributed to the user; onsets outside any contact interval are attributed to loop (in loop-active sessions) and to device sustain or environmental sound otherwise. The loop-subtraction procedure in Section 4.7 provides an additional conservative partition within loop-active recordings.

4.5 Sustained-Contact Ratio

Let $c(t) \in \{0, 1\}$ denote instantaneous finger–pad contact state, with $c(t) = 1$ during contact. The number of distinct contact initiations in $[0, T]$ is

> $$N_{presses} = \int_0^T (dc/dt)_+ \, dt$$

where $(\cdot)_+$ denotes the positive part, so only the upward transitions of $c(t)$ are counted. Acoustic onsets are a point process $\{\tau_k\} \subset [0, T]$ with

> $$N_{onsets} = |\{k : \tau_k \in [0, T]\}|.$$

Define the *sustained-contact ratio*:

> $$R_{SC} = N_{onsets} / N_{presses}.$$

Interpretation follows from the physics of a pressure-triggered pad:

- **Pure-tap null.** Each contact initiation is followed by immediate release before the next initiation. One onset per press. $R_{SC} \approx 1$.
- **Pure sustained-hold with modulation.** $c(t)$ remains 1 for an extended hold, during which the finger slides or re-activates adjacent cells without a full lift. Many onsets per press. $R_{SC} >> 1$, scaling with the modulation rate within the hold.
- **$R_{SC} < 1$.** Press motions that did not generate above-threshold acoustic events (press without pad contact, or below-threshold contact). Informative about motor activity invisible to a pure-audio interface.

R_{SC} is a single scalar that discriminates the tap regime from the sustained regime without any parameter beyond the thresholds inherent in contact and onset detection. It is the primary behavioral descriptor produced by the detection framework.

Choice of R_{SC} over alternative scalar descriptors

R_{SC} is chosen as the primary statistic over three candidates that might seem more natural at first:

- **Onset density (N_{onsets} / T, events s^{-1}).** Onset density measures how many acoustic events occur per unit session time, but its null hypothesis is session-specific and depends on how active the user was, not on whether presses are sustained. A user who taps fast and a user who holds with fast internal modulation can produce identical onset densities; the two behaviors are distinguishable only by referencing the press series—most simply by normalizing against press count, the R_{SC} construction, though inter-press interval and mean press duration carry equivalent information. Onset density is also not dimensionless: A 10-minute session and a 2-minute session of identical behavior report different values, obscuring cross-session comparison.
- **Contact fraction ($\int c(t)\,dt / T$).** Contact fraction answers "what share of session time was the finger on the pad?" not "when the finger was down, was it tapping or holding?" It is also dominated by hand-visibility rather than behavior: A session with reduced MediaPipe hand-detection rate owing to occlusion will report depressed contact fraction even when engagement is unchanged. R_{SC} is less video-dependent than contact fraction: Only the denominator (press count) relies on landmarks; the numerator (onset count) is audio-derived and unaffected by detection dropout. Missed presses deflate the denominator while the numerator holds steady, biasing R_{SC} upward—toward the sustained-hold alternative. A true sustained-hold signature therefore cannot be masked as tap-like by dropout; the symmetric risk, that heavy dropout spuriously elevates a tap-like session's R_{SC} above the null, makes MediaPipe detection rate a required co-reported statistic rather than one substituted for by R_{SC}.
- **IOI entropy or IOI coefficient of variation.** These measure periodicity or regularity of audio events, both of which are properties of the acoustic train in isolation. They do not reference the motor signal at all, so a well-behaved loop (low IOI entropy) and a well-behaved tapping user (also low IOI entropy) are indistinguishable under this metric. R_{SC} uniquely couples the two independent channels—video-derived press count and audio-derived onset count—and discriminates tap-like from hold-like regardless of periodicity.

R_{SC} has the additional property that its null value is not data-dependent: Under the pure-tap null, $R_{SC} = 1$ by construction, providing a pre-specified rejection criterion that requires no estimation from the session itself. Alternative ratios (e.g., $N_{onsets} \cdot \text{mean-hold} / T$) have nulls that depend on session-specific quantities, which would weaken the confirmatory status of any inferential use of the statistic. For these reasons R_{SC} is specified as the primary statistic of the detection framework; contact fraction and onset density are reported as descriptive context within a future validation study. One additional candidate—the cross-correlation function between the press-indicator series and the onset-indicator series—would supply timing information about *when* within a contact the onsets occur, but reduces to a scalar only after an arbitrary lag-integration window; as a summary statistic it inherits the same interpretation-ambiguity that R_{SC} avoids.

4.5.1 Contact-Interval Press Definition (Stricter Alternative)

The velocity-reversal press detector (Section 4.3) counts each downward-velocity spike as a press event. A physically complementary definition treats the *maximal continuous interval of finger-in-pad-region* as a single press, regardless of internal velocity-reversal events. Under this definition, a finger lowered onto the pad and held for three seconds—during which the instrument retriggers, decays, and is re-tapped within the same contact—is one press, not many. This definition is closer to the force-sensing-pad segmentation described in Section 5.3 and gives a stricter lower bound on distinct motor-initiation events.

Formally, define the contact-state function

$$c(i) = \mathbf{1}[\mathrm{hand\_present}(i) = 1] \cdot \mathbf{1}[y(i) > y_{\mathrm{contact}}]$$

where y_{contact} is a per-session threshold on the index-fingertip vertical coordinate, specified as the 25th percentile of y across frames in which the hand is present (median-based thresholds are rejected on geometric grounds; see Appendix A.2). Sustained-contact intervals are maximal runs of $c(i) = 1$ with a gap tolerance of up to two consecutive zero-frames (≤ 200 ms at the 10-Hz sampling rate, inclusive) absorbed as MediaPipe detection dropouts within a continuing hold. The number of contact intervals $N_{\mathrm{intervals}}$ is the alternative press count. Onsets are assigned to an interval if they fall within ±100 ms of the interval, which accommodates finite-window latency in both onset detection and landmark-derived contact timing.

A second, distinct gap-tolerance rule handles involuntary motor interruption within a sustained hold: A velocity-reversal episode shorter than the Section 3 tap-duration ceiling (< 300 ms), bracketed by sustained-contact states on both sides, is bridged as a single continuing contact interval and separately flagged as an involuntary-twitch candidate rather than counted as a new press or as intentional within-contact modulation. The < 300 ms threshold mirrors the tap-duration ceiling: A genuine new volitional contact would itself require completion within this gap, which is inconsistent with a bracketed hold. The bracketing requirement ensures the micro-release did not terminate the hold. Counts of bridged episodes per session are reported alongside R_{SC} and press counts as a session-level involuntary-motor indicator, and their frequency distribution feeds the involuntary-motor screening of Section 3 and the diagnostic protocol of Section 4.8.

Because this definition is strictly less permissive than the velocity-reversal definition (multiple velocity spikes can occur within a single contact interval), $N_{\mathrm{intervals}} \leq N_{\mathrm{presses}}$ is expected. The contact-interval ratio is denoted $R_{SC}^{*} \equiv N_{\mathrm{onsets\text{-}inside}} / N_{\mathrm{intervals}}$, where $N_{\mathrm{onsets\text{-}inside}}$ is the count of audio onsets falling within ±100 ms of any contact interval; this notation distinguishes the stricter statistic (with asterisk) from the unadorned R_{SC} of Section 4.5, which uses the velocity-reversal denominator and a raw-onset numerator. If the sustained-hold-with-modulation hypothesis holds for a given user, R_{SC}^{*} should substantially exceed R_{SC} and remain > 1; empirical validation is future work.

4.5.2 Worked Numerical Example (Hypothetical)

> **Illustrative scenario—no real recording is described.** *Consider a hypothetical 60-second recording in which a user makes three finger-pad contacts on a pressure-triggered pad: a 100 ms tap, a 2-second sustained hold with two clear within-hold force re-pressings that re-trigger the pad, and a 10-second sustained hold with continuous low-amplitude force oscillation near 3 Hz that repeatedly crosses the pad's re-trigger threshold.*
>
> *Under the velocity-reversal press definition (Section 4.3), the kinematic detector fires once on the tap, once on initial contact of the 2-second hold plus twice within it (on each visible fingertip-depth re-advance), and once on the 10-second hold's initial contact—giving* $N_{presses} = 5$*. Audio onsets fire once on the tap, three times on the 2-second hold (initial activation plus two re-triggerings), and approximately 30 times during the 10-second oscillating hold—giving* $N_{onsets} \approx 34$*. Thus* $R_{SC} \approx 34/5 \approx 6.8$*, well above the tap-like null of 1.*
>
> *Under the contact-interval press definition (Section 4.5.1), the three contacts collapse to three maximal in-pad-region intervals (*$N_{intervals} = 3$*) and* $N_{onsets\text{-}inside} \approx 34$ *(all onsets occur within ±100 ms of a contact interval), yielding* $R_{SC}^{*} \approx 34/3 \approx 11.3$*. Both statistics reject the tap-like null by roughly an order of magnitude; the rough agreement on the sustained-hold inflation across the two definitions is the cross-check the two definitions exist to supply. A pure-tap session with the same* $N_{contacts} = 3$ *and one onset per contact would give* $R_{SC} = R_{SC}^{*} = 1$ *by construction. The numerical values above are illustrative only and are not drawn from any recording. A rhythmic-tap (twitch-burst) session—a sequence of brief discrete contacts carrying information in inter-tap interval structure rather than within-contact modulation—would under the same statistics produce* $N_{onsets} \approx N_{presses}$ *and therefore* $R_{SC} \approx 1$*, indistinguishable from the pure-tap null. This is consistent with Section 3: The rhythmic-tap class is detected not by* R_{SC} *but by statistics on the inter-press interval series (burst rate, rhythm autocorrelation, Morse-symbol decoding), deferred to future work.*

4.6 Rotation Detection

Body rotations toward a person or object in frame are detected from MediaPipe Pose shoulder and head landmarks. A raw "sweep" is defined as three or more consecutive frames of same-sign horizontal head-landmark velocity with cumulative displacement > 0.04 in normalized image coordinates. Raw sweeps are further filtered to *intentional rotations* by removing ambient drift (sweeps with low cumulative displacement, short duration, or no visible reference target). Rotations enter this paper only as corroborating evidence of active engagement within a future validation study; they are not central to the interface claim.

4.7 Loop-Subtraction for Periodic-Device Confounds

When the target device emits its own periodic acoustic pattern (a drum-machine loop, a metronome, any stereotyped pacing stimulus), raw audio onsets conflate user-initiated events with device-initiated events. The detection framework addresses this by a phase-locked partition that yields a strict lower bound on user-initiated activity.

Given a session with an active periodic source at dominant inter-onset interval IOI (estimated from librosa beat tracking or from autocorrelation of the binned onset event series), a single scalar global phase offset $t_0 \in [0, \mathrm{IOI})$ is fit per session by minimizing the sum of squared distances from in-contact onset times to the nearest grid point $t_0 + k \cdot \mathrm{IOI}$. An in-contact onset is then labeled *on-grid* (attributed to the loop) if it falls within a small tolerance (typically ±30 ms, chosen to cover librosa spectral-flux onset-localization jitter for percussive onsets) of a grid point, and *off-grid* (attributed to the user) otherwise.

The partition is deliberately conservative in attributing to the loop: An on-grid onset might in truth be a user press that landed coincidentally on the grid, while an off-grid onset cannot be generated by a clean k-subdivision grid loop. The off-grid count is therefore a strict *lower bound* on user-initiated in-contact onsets, and the corresponding user-only ratio (off-grid onsets) / (contact intervals) is a strict lower bound on R_{SC}^* in that session.

Three caveats apply. First, the phase fit assumes a single time-invariant loop phase over the session; real loops can be restarted or tempo-nudged, which would degrade the on-grid count and inflate the off-grid count. A phase-stability cross-validation (refit phase on the first half of the session, score on the second half, and vice versa) quantifies this drift. Second, the IOI grid is the dominant period; multi-event loop patterns (e.g., kick-snare-hat) do not fire on every grid subdivision, so on-grid is an upper bound on loop contribution under the strongest assumption (loop fires at every grid point). Third, the tolerance window (typically ±30 ms) should be reported together with a sensitivity sweep across reasonable values (±15 to ±50 ms) so that the loop-subtracted lower bound is traceable to the chosen tolerance. Within any future validation study, the RMS distance from in-contact onset time to the nearest grid point can itself be compared against the librosa onset-localization uncertainty: RMS distances substantially exceeding the detector uncertainty are themselves evidence that the in-contact onset stream is dominated by non-periodic user-driven events rather than by the loop.

4.8 Diagnostic Protocol for Confound Separation

Three distinct confound sources can inflate or distort the onset and press series in a deployed session: environmental periodic noise (ventilators, HVAC cycles, infusion pumps, dishwashers, mechanical devices producing regular acoustic events), involuntary user motor output (clonus, physiological tremor, fasciculation, dystonic overflow, spasticity-driven twitches), and intentional user engagement. Each has a distinct temporal signature and can be isolated by a structured baseline-and-contrast recording protocol, described here so that a validation site can apply the same diagnostic flow without improvising per-session.

Environmental periodic noise. Characterized by constant inter-onset interval, mechanical frequency unrelated to user physiology, and presence during no-user baseline recordings. Isolated by a 2–3 minute baseline segment with the microphone active and the pad untouched: Autocorrelation of the resulting onset series shows sharp peaks at the loop IOI. The phase-locked partition of Section 4.7 then formalizes the correction on subsequent user-present segments.

Involuntary user motor output. Characterized by user-physiological frequencies (clonic seizure 2–3 Hz, myoclonic jerks 1–5 Hz, clonus 5–8 Hz, physiological tremor 4–12 Hz), presence during user-present but unprompted windows, and absence from the no-user environmental baseline. Isolated by a user-present-but-unprompted segment (caregiver silent, user seated at the pad without directed task): Onset and press rates in this segment characterize the involuntary-motor baseline for that user and session, and the power spectral density of the force signal across 1–12 Hz characterizes the per-user involuntary spectrum. Intentional engagement distinguishes itself from this baseline by the task-contingent contrast of Section 3 and by residing in the < 8 Hz volitional band with a frequency profile distinct from the involuntary spectrum captured in this segment. For epilepsy and rhythmic-involuntary-motor populations the frequency-domain separation is developed at full deployment depth in Section 5.8.8.

Intentional user engagement. Characterized by task-contingent upregulation above the involuntary-motor baseline, temporal clustering near caregiver-prompt onset, and the within-contact modulation structure that R_{SC} is defined to capture. Isolated by explicit caregiver-prompted segments contrasted against the two baselines above: The ratio of prompted-window onset rate to unprompted-window onset rate provides a per-session measure of intentional-output signal above the combined environmental-plus-involuntary floor.

A recommended deployment-session structure therefore interleaves three short segments—no-user baseline (2–3 min), user-present unprompted (2–3 min), user-prompted engagement (remainder of session)—before any R_{SC} or command-alphabet statistics are reported. The three baselines are reported alongside the primary statistics so that a reviewer can verify each confound has been characterized rather than assumed absent. The MediaPipe hand-detection rate of Section 4.3 is reported for all three segments for the same reason.

5. Interface Design

5.1 Requirements Derived from the Motor Signature

Four minimum information-capture requirements follow from the defining properties of the two in-scope continuous channels of Section 3—sustained-contact-with-modulation and rhythmic-tap:

1. **Contact onset time.** Resolution ≤ 30 ms to preserve alignment with external rhythmic stimuli.
2. **Contact duration.** Encoded as a continuous scalar, not thresholded to a binary “pressed / not pressed.” Capture range: A lower bound at or below the briefest volitional contact observed in the target population (≤ 50 ms, providing headroom below the < 300 ms tap ceiling of Section 3) and an upper bound above the longest sustained hold of clinical interest (≥ 10 s). Classification of captured durations into the Section 3 taxonomy classes (pure tap, silent hold, sustained-contact-with-modulation, rhythmic tap) is a downstream interpretation step, not a sensor requirement.
3. **Within-contact force modulation.** While contact is held, force must be sampled repeatedly over time with two resolution requirements. Temporal sampling rate: ≥ 25 Hz (Nyquist ceiling of 12.5 Hz), sufficient to resolve voluntary modulation up to the ~ 8 Hz volitional ceiling of Section 3 and to characterize involuntary-motor contaminants in the 1–12 Hz band (clonus, physiological tremor, clonic-seizure and myoclonic activity) for the frequency-domain screening step of Section 5.8.8. Amplitude resolution: ~8-bit (256 discriminable force levels), sufficient to resolve voluntary modulation patterns above pad noise. The 100 Hz hardware rate specified in Section 5.2 satisfies both requirements with substantial headroom.
4. **Inter-press interval resolution.** Press-series timestamps resolved to ≤ 30 ms, sufficient to distinguish Morse-ratio boundaries (≥ 3× intra-burst interval) at volitional tapping rates. This requirement is specific to the rhythmic-tap class of Section 3; the detection pipeline derived from it is deferred to future work, but the present hardware specification satisfies the timing resolution automatically by virtue of the 100 Hz sampling of Section 5.2.

Any interface that discards any of these four dimensions returns to the failure mode of Section 1.2.

5.2 Hardware Specification

A force-sensing pad built from either (a) a capacitive sensing array with analog front-end readout (e.g., Infineon PSoC (formerly Cypress) + capacitive matrix) or (b) a piezoresistive force-sensing resistor (FSR) array with per-cell ADC conversion. Either approach meets the four requirements with commodity components; Dahiya and Valle [8] survey tactile-sensing technologies whose spatial resolution and dynamic range are well above the requirements specified here.

- Active area: ≥ 100 cm^2, sized to a child-to-adult palm
- Spatial resolution: 4×4 to 8×8 cells, sufficient to detect finger-rolling within a hold
- Force range: 0.1–20 N per cell (light contact to intentional press)
- Sampling rate: ≥ 100 Hz per cell (10× oversampling against the minimum modulation rate)
- Output: serial stream of (cell, force, timestamp) tuples. *No in-device thresholding to binary.* This is the critical architectural choice.

5.3 Signal-Processing Pipeline

Raw force-time streams pass through:

1. Contact segmentation (threshold at ~0.1 N, hysteresis to suppress chatter), yielding contact intervals $[t_a, t_b]$ of duration $T_c = t_b - t_a$

2. Per-interval feature extraction, producing a feature vector per contact. Let $F(t)$ denote the aggregate force-time signal on $[t_a, t_b]$: For fingertip contact spanning multiple cells (the standard case on a 4×4 or denser grid), per-cell forces are summed into the scalar $F(t)$ for the kinematic and spectral features below and independently combined into a force-weighted 2D centroid for within-contact position modulation, with the aggregation performed in signal-processing software rather than pad firmware to preserve the non-thresholded architectural choice of Section 5.2. The features include:

   > $$T_c = t_b - t_a \quad \text{(contact duration)}$$
   > $$F_{peak} = \sup_{t \in [t_a,\, t_b]} F(t)$$
   > $$F_{mean} = (1/T_c) \int_{t_a}^{t_b} F(t)\, dt$$
   > $$CV_F = \sigma_F / F_{mean} \quad \text{(modulation intensity, coefficient of variation)}$$
   > $$f_c = \int_0^\infty f \cdot |\hat{F}(f)|^2\, df \;/\; \int_0^\infty |\hat{F}(f)|^2\, df \quad \text{(modulation rate, force-spectral centroid)}$$

   where $\hat{F}(f)$ is the Fourier transform of $F(t)$ on $[t_a, t_b]$. The spectral centroid f_c is the power-weighted mean modulation frequency: A steady hold has $f_c \to 0$; a rapidly-modulated hold has f_c in the 1–10 Hz range. Both CV_F and f_c are parameter-free given $F(t)$ and are well-suited as user-independent features.
3. Classification: feature vector → small AAC command alphabet (Section 5.4)

5.4 Mapping from Signature to AAC Command

The mapping layer is user-specific and must be calibrated. A minimally-viable command alphabet for a user whose repertoire spans the taxonomy of Section 3:

- **A.** Short contact: $T_c < 300$ ms, CV_F low
- **B.** Sustained hold, low modulation: $T_c \geq 300$ ms, $CV_F < 0.2$
- **C.** Sustained hold, high modulation: $T_c \geq 300$ ms, $CV_F > 0.4$
- **D.** Rhythmic taps: ≥ 3 contacts with $T_c < 300$ ms each; inter-contact interval within a per-user calibrated intra-burst window (≥ 200 ms floor; upper bound and Morse-ratio boundary pause set per the timing scheme of Section 3). The 500 ms fixed default used in early deployments is retained only as a fallback where per-user calibration has not yet been performed.
- **U.** Unclassified (intermediate-modulation sustained holds with $0.2 \leq CV_F \leq 0.4$): The interface should prompt for repetition rather than infer a command.

Commands A–D are deliberately agnostic to semantic mapping. Semantic assignment ("yes," "no," "more," "stop") is a downstream decision negotiated with caregivers and adjusted based on the user's observed contextual responses, in the tradition of AAC vocabulary bootstrapping.

The thresholds above are specified at calibration time and assumed quasi-stationary over a deployment window. For users whose motor envelope drifts with disease time, fatigue, arousal, or medication state, the thresholds should be refit on a schedule matched to the expected rate of drift; see Section 5.8.2 for the progressive-disease case and Section 5.8.3 for the case in which distal upper-extremity force generation itself is the rate-limiting variable. Operationally, calibration is a single caregiver-prompted session of 10–15 minutes on the pad: T_c and CV_F distributions are fit from the resulting contact-interval record, and thresholds are placed at empirical valleys between classes, defaulting to the values above when the distribution is unimodal and flagging the user for the single-class fallback path of Section 5.5.

5.5 Bootstrap Protocol for a Binary Channel

A user who has never had a reliable yes/no channel cannot be asked to perform one (cf. Sigafoos et al. [7] on the prerequisites for teaching communicative rejecting in individuals without a pre-trained binary response). The bootstrap inverts the usual assumption: Rather than asking the user to produce a pre-trained binary response, the system observes the user's *existing* motor signature and defines the binary cut within it.

Theoretical precondition: two distinguishable classes of hold. The bootstrap requires that the user spontaneously produces two distinguishable classes of sustained hold—modulated (force-deflection events occur within the contact) and silent (the contact is maintained but no within-hold modulation above noise is emitted). The two classes are subclasses of sustained contact and therefore confluent under binary closure (both register as hold), but are separable on a force-sensing pad because in-hold modulation is directly measurable on the pressure axis regardless of whether it is audible on the desk. The precondition that both classes are spontaneously emitted by a given user is an empirical one and must be verified on a per-user basis before the bootstrap is deployed; a user who produces only one class cannot be bootstrapped into a binary channel by this protocol.

Proposed operational rule. Define a hold as sustained when contact duration $T \geq 1$ s. Compute N, the count of within-hold force-deflection events exceeding a calibrated noise floor. Classify:

- *High-modulation class* (tentative semantic pole α)—$N \geq 2$ during T.
- *Low-modulation class* (tentative semantic pole β)—$N \leq 1$ during T.

The $N = 2$ threshold is motivated theoretically by the intended sparseness of the $N = 1$ region of the within-hold modulation-count distribution: A single force deflection within a one-second sustained hold is more parsimoniously attributed to incidental pressure fluctuation (finger re-seating, breath-coupled postural drift, sensor noise) than to intentional modulation, while $N \geq 2$ events within the same hold require either deliberate re-pressing or persistent within-hold oscillation. The threshold therefore lies in the region of lowest expected density between “at most one possibly-incidental modulation” and “clear multi-event modulation,” which is the condition under which a bootstrapped binary class boundary is expected to be stable. Empirical validation of the threshold location on a per-user basis—including verification that the user’s own modulation-count distribution admits a sparse separating region around $N = 1$—is a required calibration step before deployment.

Semantic assignment is downstream and caregiver-mediated. The cut above delivers only two reliably-distinguishable motor classes; pinning α to “yes” or “no” is a negotiated decision made over repeated exposures with consistent consequence, in the tradition of AAC vocabulary bootstrapping. The interface provides the channel; the dyad provides the code-book. Operant conditioning of existing motor behavior with consistent environmental consequence is well-established in the applied-behavior-analysis and AAC literatures; the novelty here is only the specific choice of discriminative dimension—within-contact modulation density rather than presence or absence of contact.

Applied to progressive neuromuscular disease. A user transitioning off a failing tap-based switch typically retains an intact pre-trained binary code-book—the yes/no semantic mapping already exists in the user's prior AAC practice and in the dyad's conversational history [11, 12, 13]. The bootstrap in that case functions not as a code-book invention but as a *transfer operation*: The high-modulation and low-modulation classes are assigned the user's existing semantic poles rather than negotiated from zero. The transfer is expected to be faster and more reliable than from-scratch bootstrap, but is subject to the same precondition (the user must spontaneously produce both classes of hold) and should be re-verified before each transfer because progressive motor degradation can collapse the discrimination between classes over time. Periodic re-calibration of the class boundary as disease progresses is a standing deployment requirement for progressive-disease users and is not required for static-etiology users (Section 5.8).

Fallback for users who produce only one class. Some users will spontaneously produce sustained holds of only one class—either only modulated holds or only silent holds—in which case the two-class bootstrap above is not applicable. The interface still captures onset, duration, and modulation for these users and can sustain a graded single-dimensional channel: Presence-of-hold over fixed temporal windows is a valid communicative signal even without class separation, and modulation intensity within a single class can be calibrated to a graded attention/intensity dimension (e.g., “attending” vs. “not attending”) rather than a binary yes/no. A hand-off to a temporal-pattern bootstrap (hold-present vs. hold-absent across fixed windows, rather than modulated vs. silent within a hold) is a fallback path for users who never converge on class separation. These accommodations are meaningful: For a user with no prior communicative channel at all, establishing even a reliable single-dimensional signal is a substantial clinical outcome relative to the binary-null baseline.

The bootstrap protocol is specified as a design; empirical validation on a per-user basis is future work.

5.6 Channel Capacity—the Structural Argument

The design claim generalizes to any user with the sustained-engagement-with-modulation signature, regardless of how many such users exist and regardless of which transducer class (Section 2.4) instantiates the interface. The argument is information-theoretic.

A binary-closure switch transmits a single bit per contact event: The output is a symbol drawn from {activated, not-activated}.

> $$H_{binary} = 1 \text{ bit per contact.}$$

A force-sensing pad sampling at rate f_s with b-bit per-sample resolution encodes a vector of $f_s \cdot T$ samples of b bits each during a contact of duration T. Its per-contact information capacity is

> $$H_{force} = b \cdot f_s \cdot T \quad \text{bits per contact.}$$

For the specification in Section 5.2 ($b = 8$, $f_s = 100$ Hz, $T = 1$ s):

> $$H_{force} / H_{binary} = (8)(100)(1) / 1 = 800{:}1.$$

The nominal capacity ratio is nearly three orders of magnitude (800:1). A gesture whose information content lies in within-contact modulation cannot be recovered from a 1-bit channel; the projection is lossy by construction of the channel. *Information-destructive* is therefore a literal description, not a figurative one.

In progressive disease the argument has additional force. The tap channel's reliability is declining over disease time as the motor system loses the coordinated agonist–antagonist timing that clean taps require; the force-modulation channel's reliability is also declining, but from a much higher ceiling and along a slower trajectory (sustained contact requires only sustained agonist tone, not rapid release). The capacity advantage therefore widens rather than narrows across the clinical window in which sustained contact with modulation is still available—which is precisely the window in which tap-based switch assessment has already been declared non-viable. For static-etiology users the argument is even starker: Binary-switch assessments administered across the user's lifespan have projected the full motor signature onto a one-bit channel in each instance, so the capacity loss integrated over the user's lifetime is the ratio above multiplied by the total count of assessment sessions, scanning trials, and attempted selections.

5.6.1 Realistic Entropy-Rate Estimate (Bandwidth- and Noise-Corrected)

The 800:1 figure is a *raw channel capacity* expressed as the product of nominal sampling parameters. It is the ceiling; the achievable rate over a physical sensor will be lower for two well-understood reasons: (a) additive noise reduces the effective per-sample bit depth, and (b) the motor bandwidth of a finger pressing on a pad is on the order of 10 Hz, so oversampling at 100 Hz does not produce 100 independent samples per second—it produces a bandwidth-limited signal whose information content is bounded by Shannon–Hartley capacity over that bandwidth [6].

Using the force-sensor parameters of Section 5.2 and conservative values for the sources of loss:

> Nyquist motor bandwidth $B \approx 10$ Hz
> Per-sample signal-to-noise $SNR \approx 10^2$ (20 dB—typical resistive/capacitive pad)
> Shannon–Hartley per-sample capacity $C_{sample} = ½ \log_2(1 + SNR) \approx$
> 3.33 bits/sample
> Independent-sample rate $\approx 2B = 20$ samples/s
> Per-second channel capacity $H_{force,eff} \approx B \log_2(1 + SNR) \approx 66.6$ bits/s
> Per-contact capacity for $T = 1$ s hold: ≈ 66.6 bits per contact.

A 1-bit binary switch, regardless of sampling rate, transmits at most 1 bit per contact transition (the observation of "switch-closed" once is redundant with all subsequent "still-closed" samples while held). The conservative, bandwidth- and noise-corrected capacity ratio is therefore still

> $$H_{force,eff} / H_{binary} \approx 66 / 1 \approx 66{:}1.$$

Even under aggressively pessimistic noise assumptions (SNR = 10 dB, B = 5 Hz—a badly calibrated pad with sluggish mechanics), the per-second effective capacity falls to $B \cdot \log_2(1 + 10) \approx 17$ bits/s, which at $T = 1$ s is still a 17:1 ratio against binary. The order-of-magnitude-or-greater gap (ranging from ≈ 17:1 under pessimistic assumptions to ≈ 66:1 under realistic ones) is structural to the comparison; it is not an artifact of the 8-bit choice in Section 5.2.

Table 1. Shannon–Hartley effective capacity $H_{force,eff} = B \cdot \log_2(1 + SNR)$ in bits/s, across a 4×4 grid of motor bandwidth B (rows) and per-sample signal-to-noise SNR (columns, dB). Each cell also equals per-contact capacity at T = 1 s (hold length), and therefore the ratio against a 1-bit binary closure. The middle of the grid (B = 10 Hz, SNR = 20 dB; bold) is the value cited above and in the abstract; across the corners most defensible for a finger-on-pad interface the envelope is ∼ 17:1 (worst plausible, 5 Hz / 10 dB) to ∼ 266:1 (fast-finger, 20 Hz / 40 dB), extending to ∼ 664:1 only at the 50 Hz instrumental-bandwidth corner. The 800:1 raw-capacity figure of Section 5.6 (nominal 8-bit × 100 Hz × 1 s product) sits just above the Shannon–Hartley ceiling of this envelope (800 / 664 ≈ 1.2×); the two diverge because the nominal product assumes 100 *independent* bits/s of channel capacity where Shannon–Hartley at 50 Hz bandwidth and 40 dB SNR caps the achievable rate at 664 bits/s.

B \ SNR	10 DB	20 DB	30 DB	40 DB
5 Hz (slow)	17.3	33.3	49.8	66.4
10 Hz (nominal)	34.6	**66.6**	99.7	132.9
20 Hz (fast finger)	69.2	133.2	199.3	265.8
50 Hz (instrument-like)	173.0	332.9	498.4	664.4

The roughly 16-fold span across the grid (17–266 bits/s at the corners most defensible for a finger-on-pad interface, extending to 664 only at the instrumental-bandwidth upper corner) is dominated by the B-axis: At any fixed SNR, doubling bandwidth doubles capacity, whereas at any fixed bandwidth capacity scales as $\log_2(1 + SNR)$ and is close to linear in dB. Bandwidth calibration is therefore the primary design lever in Section 5.2, not bit depth; a 4-bit ADC at 10 Hz with 20 dB SNR already delivers ∼ 66 bits/s of effective capacity without reaching the 8-bit, 100 Hz nominal envelope. The 66:1 headline sits at mid-grid SNR and nominal motor bandwidth—neither optimistic nor at the ceiling. Ratios above 100:1 occupy eight of the sixteen cells: One in the 10 Hz row and three in the 20 Hz row fall within the physiologically plausible envelope for finger-on-pad motor bandwidth; the four cells of the 50 Hz row correspond to a sensor-bandwidth-limited instrumental ceiling rather than a typical finger-motor operating point (cf. Section 2.2).

5.7 Translation to AAC Throughput

A 40-item core AAC vocabulary requires $\log_2(40) \approx 5.32$ bits per selection. At the conservative per-contact capacity $H_{\text{force,eff}} \approx 66$ bits (1 s hold, 20 dB SNR, 10 Hz bandwidth), a single sustained contact carries capacity sufficient to address any realistic core-vocabulary size with substantial headroom, so the *bottleneck is not the channel*; it is the mapping design in Section 5.4. A binary switch at the same rate addresses $2^1 = 2$ items per contact, so reaching a 40-item vocabulary requires a 6-deep scanning tree and correspondingly more contacts; the $\sim 6\times$ throughput loss in a scanning interface is a direct consequence of binary projection and is recovered by force-sensing without any change to the underlying motor signal.

In terms of effective selections per minute, the relevant rate for the continuous-output interface specified in Section 5.3 is the engagement-interval rate (the unit the proposed signal-processing pipeline segments on), not the velocity-reversal rate. At any given contact-interval rate r, binary-switch scanning of an N-item set at scanning depth $d = \lceil \log_2 N \rceil$ requires on the order of d switch activations per selection, giving a binary throughput of $\sim r/d$ selections/min; direct continuous-output mapping is bounded by the engagement rate itself at $\sim r$ selections/min. For a 40-item vocabulary ($d \approx 6$), the AAC-relevant throughput ratio is therefore a factor of ~ 6 improvement, on the same order as the scanning-depth saving, not the 800:1 raw-capacity ratio—because engagement events are the rate-limiting resource, not bits-per-engagement. The engagement-interval rate r for users with the sustained-engagement signature is an open empirical question (Section 6).

This argument applies to any user whose structured motor output has nontrivial distribution along the modulation dimension. It does not require prevalence data. It follows from the interface specification.

5.8 Clinical Deployment Across Target Populations

The interface specification above is population-agnostic: The channel-capacity argument holds for any user whose motor output carries information along the modulation dimension. Deployment, however, differs substantially across candidate populations, and the differences are worth stating explicitly. Eight scenarios are developed below: static-etiology users for whom the interface is a first channel (Section 5.8.1); progressive-disease users for whom it is a bridge channel (Section 5.8.2); users whose primary constraint is distal upper-extremity force generation or atypical wrist kinematics (Section 5.8.3); integration with existing clinical assessment batteries (Section 5.8.4); a non-contact transducer variant, developed from the engagement-state generalization of Section 2.4, for users for whom physical contact is contraindicated by tactile hypersensitivity, skin compromise, or pressure-sore risk (Section 5.8.5); prosthesis-mediated and direct myoelectric deployments for users with upper-extremity amputation or orthotic limb use (Section 5.8.6), for whom the multi-channel structure of the EMG record expands the symbol alphabet by one axis per electrode; combat blast-injury survivors (Section 5.8.7), a cross-cutting clinical population in whom burn scarring, upper-extremity amputation, and blast-induced communication disorder frequently co-occur in the same patient and for whom preserved visuomotor coordination makes the non-contact variant's distance-dimension vocabulary gain immediately usable; and epilepsy and rhythmic-involuntary-motor populations (Section 5.8.8), including the clinically severe case of continuous subclinical seizure activity, for whom frequency-domain separation of involuntary rhythmic contamination from intentional modulation is the operational key. A ninth subsection (Section 5.8.9) notes related populations that share the motor-signature structure but carry population-specific confounds beyond the scope of this paper.

5.8.1 Static-Etiology Users—the Sole-Channel Case

For users whose sustained-contact-with-modulation signature has been the structure of their motor output since congenital or early-acquired injury—severe hypoxic-anoxic brain injury (HAI) is the canonical case, alongside severe cerebral palsy with preserved tactile response and post-status-epilepticus syndromes in which tap fidelity was never re-established—the interface is not a bridge to any prior channel. It is, as far as the clinical record goes, the first channel with sufficient bandwidth to register the motor signature at all. Every binary-switch assessment administered across such a user's lifespan has projected that signature to a single bit and recorded the result as absence of reliable voluntary motor control. Absence of tap-based evidence was, and was recorded as, evidence of absence.

Three consequences follow for deployment in this population. First, the bootstrap protocol of Section 5.5 is code-book *invention* rather than transfer: The user has no pre-existing binary semantic mapping in their communication history (or, if any was taught, it was attempted through the tap-based channel and is unlikely to have stabilized), so the high-modulation / low-modulation cut established by the interface is not inheriting an existing yes/no distinction. Semantic assignment is negotiated from zero with caregivers across repeated exposures and consistent environmental consequences, in the tradition of AAC vocabulary bootstrapping. Second, clinical-record history is actively misleading here: A user's file will contain years or decades of failed binary-switch assessments, coded as negative evidence for voluntary motor control. That history should be treated as *assessment-channel-limited data*, not as a prior against the hypothesis the force-sensing interface is designed to test. Third, the failure cost is asymmetric: A static-etiology user wrongly classified as having no voluntary motor output bears the cost of that classification for the remainder of their life. The deployment recommendation that follows is to treat a force-sensing-pad session as a standing, repeatable adjunct to clinical assessment—not a one-time diagnostic—for any candidate whose prior assessment record consists of binary-switch negatives.

The deployment presumes at least minimal conscious awareness sufficient to direct voluntary motor output toward a target. A force-sensing session is proposed here as a diagnostic adjunct capable of upgrading an ambiguous or MCS-minus CRS-R classification to MCS-plus through detection of sustained-contact modulation under caregiver prompting; it is not a substitute for awareness assessment in users with no documented behavioral responsiveness, and users consistently classified in the unresponsive-wakefulness range without subsequent upgrade fall outside the framework's scope.

The signal-processing thresholds for this population (Section 5.4) are assumed quasi-stationary across a deployment window, subject only to the user's arousal, fatigue, and medication state. Re-calibration on a monthly to quarterly schedule is typically sufficient; the thresholds are not expected to drift on a timescale shorter than that in the absence of concurrent disease.

5.8.2 Progressive Neuromuscular Disease—the Bridge-Channel Case

In progressive neuromuscular disease, most prominently late-stage amyotrophic lateral sclerosis (ALS) [10], the trajectory of upper-extremity motor decline is well documented [11], and the clinical sequence implies a specific prediction for the motor-signature framework of this paper: Tap reliability should degrade before sustained-contact-with-modulation output is lost. The mechanistic basis for this prediction is that clean tap execution requires coordinated agonist–antagonist timing for rapid release, while sustained contact requires only sustained agonist tone; the former is expected to fail earlier in upper-motor-neuron-predominant and mixed ALS phenotypes because the neural circuitry for coordinated release fatigues or deteriorates ahead of the circuitry for simple sustained contraction. This dissociation is a mechanistic prediction from known motor-unit physiology, not a directly cited empirical finding; its validation is flagged as an open empirical question in Section 6.2. The interface's role for these users is therefore a *bridge channel*: It maintains direct-selection AAC usability across the months to year or longer [11, 12, 13] between the disease stage at which tap-based switch scanning becomes unreliable and the disease stage at which all hand motor control is lost and the user transitions to eye-gaze AAC or later modalities. The channel-capacity argument in this population is not about sole-channel rescue; it is about preserving the working AAC channel through a clinical window that existing switch technology treats as a cliff.

Three deployment consequences distinguish this case from Section 5.8.1. First, the bootstrap protocol's role inverts. A user transitioning off a failing tap-based switch typically retains an intact pre-trained binary code-book: The yes/no semantic mapping already exists in the user's prior AAC practice and in the dyad's conversational history. The bootstrap functions as a *transfer operation* rather than an invention, assigning the high-modulation and low-modulation classes to the user's existing semantic poles. The transfer is expected to be faster and more reliable than from-scratch bootstrap, but is subject to the same precondition (the user must spontaneously produce both classes of hold). Second, signal-processing thresholds are non-stationary on a timescale of weeks to months: As motor force generation declines, the T_c and CV_F thresholds of Section 5.4 drift, and the interface requires a scheduled re-calibration protocol tracking disease time. Third, the interface should exit gracefully: When sustained-contact-with-modulation itself degrades below threshold, the clinical workflow should transition the user to the next available modality (eye-gaze, typically) rather than require the user to fail silently inside the continuous-output interface. A hand-off protocol, specified as a deployment requirement, is an open design question for this population.

The ALS population is also more logistically tractable as a validation target than the static-etiology population: Users can consent to validation studies, are well-networked through patient-advocacy organizations, and the research infrastructure around ALS AAC is mature. These considerations are not scientific but they bear on which population a validation study should recruit from first (Section 6.4).

5.8.3 Distal Upper-Extremity Weakness and Atypical Wrist Kinematics

A third population is defined not by a central-nervous-system etiology but by the mechanics of the user's wrist and hand. Users with reduced distal upper-extremity force generation—whether from myopathy, rheumatological disease, post-stroke flexor synergy, degenerative joint disease, or advanced age—may lack the wrist-extensor strength required to execute a clean tap release. Clean taps require the wrist to actively lift the fingertip off the pad between activations; when extension force generation is insufficient for that lift, the fingertip remains in contact and the motor output defaults to sustained contact as a compensatory pattern, even when the user's neurological motor intent is a discrete tap. The resulting motor-signature profile is indistinguishable, from the pad's point of view, from the sustained-contact-with-modulation signature this paper is designed to register.

Users with atypical wrist kinematics—contractures (post-traumatic, spastic, or post-burn), joint hypermobility or hypomobility, congenital variants (radial club hand, Madelung deformity, post-surgical alignment), or rheumatoid deformities—face a related but distinct problem: The standard pad-positioning geometry (horizontal pad at the user's midline, pressed from above with a neutrally-positioned wrist) is not reachable, and the interface is excluded at the level of access mechanics regardless of any neurological consideration. These users are often served by one-off caregiver-fabricated mounts; the force-sensing interface described here is compatible with such mounts, but the hardware specification in Section 5.2 should be read as describing pad electronics and sampling parameters, not pad mounting or access geometry.

Two deployment consequences apply. First, a candidate-selection protocol that routes users with distal weakness or atypical wrist geometry into force-sensing assessment (in preference to a tap-based assessment the user will fail for mechanical rather than neurological reasons) is a clinical workflow change the interface enables and the prior art does not. Second, the pad mounting geometry should be assessed and fitted per user before any signal-processing threshold is calibrated: Section 5.4 thresholds are meaningful only with respect to a mounting geometry that permits the user's actual range of motion to reach the pad, and a calibration session conducted on a non-reachable mount will report spurious modulation statistics. Consistent with this, the active-area and cell-layout parameters in Section 5.2 should be read as a floor rather than a target: A larger active area (up to the user's full reachable pad footprint) and a finer cell grid (to accommodate oblique or rotational approach patterns) are welcomed.

5.8.4 Integration with Standard Assessment Batteries

The detection framework of Section 4 is designed to run as an adjunct to existing clinical assessment rather than as a replacement. A candidate user who scores ambiguously on the Coma Recovery Scale–Revised (CRS-R) motor subscale [2] or whose Motor Behavior Tool [3] record contains uncategorized responses is a prime target for a force-sensing-pad session: R_{SC}^{*} and CV_F statistics quantify whether the user's pad behavior is consistent with sustained-contact-with-modulation and can either upgrade the categorical motor score to “localization” or establish an upper bound on voluntary motor output that the standard assessment could not supply. The interface therefore *enriches* rather than replaces the existing workflow, which materially reduces deployment friction: Rehab centers already administer CRS-R, already have trained raters, and already have scheduling infrastructure for motor assessment. Framing the force-sensing session as a CRS-R adjunct rather than a new standalone battery is the deployment path that minimizes adoption cost.

For the progressive-disease population (Section 5.8.2), the analogous integration is with AAC assessment batteries standard in the ALS clinical literature [11, 12, 13]—most notably switch-access timing protocols—where a force-sensing-pad session at regular disease-progression intervals quantifies the declining tap channel alongside the holding force-modulation channel, and provides the objective evidence needed to time the transition from scanning AAC to direct-selection force-sensing AAC.

5.8.5 Non-Contact Transduction and Users with Tactile Contraindications

A non-trivial fraction of candidate users for the interface cannot tolerate the sustained physical contact that Section 5.8.1–Section 5.8.4 assume. Tactile hypersensitivity is a well-documented sensory feature in autism spectrum disorder, sensory processing differences, and in complex-regional and allodynic pain conditions; localized skin compromise is the standard state for burn survivors, patients with epidermolysis bullosa, and users recovering from grafts or flaps; pressure-sore risk is a continuous clinical constraint for spinal-cord-injury users, the chronically bedbound, and users whose peripheral sensation is diminished and who therefore cannot themselves report early tissue breakdown at a contact site; and immunocompromised users (bone-marrow transplant recipients, severe combined immunodeficiency, ongoing cytotoxic chemotherapy) require minimized shared-surface contact for reasons orthogonal to the motor signature. For each of these populations, the motor signature the interface is designed to register may be present and the contact-instantiation of the transducer nonetheless unavailable. The engagement-state generalization of Section 2.4 exists in large part to serve this set of users.

The non-contact transducer classes most directly applicable are capacitive proximity arrays (open-air electrode panels measuring displacement of the hand through a near-field capacitive volume), infrared and time-of-flight optical rangers (single-beam or low-resolution depth arrays), and millimeter-wave radar front ends (Soli-class small-aperture radars designed for hand gesture recognition at centimeter range). All three classes deliver an adequately-sampled continuous scalar or low-dimensional vector per frame; all three admit the onset-duration-modulation decomposition that the force-pad case in Section 4 and Section 5.4 is built on. The detection framework of Section 4 transfers under units substitution: Where Section 4 writes of the pad's force-contact-state estimate, the capacitive case substitutes a normalized field-occupancy estimate, the optical case substitutes a range-to-target estimate, and the radar case substitutes a return-amplitude or Doppler estimate. R_{SC} as defined in Section 4.5 remains well-formed, with “onset” redefined as an above-threshold excursion of the continuous engagement signal and “press” redefined as a full engagement-disengagement cycle, and the loop-subtraction technique of Section 4.7 applies unchanged to any instantiation in which the target device emits its own periodic audio during operation.

Four points of divergence from the contact case deserve explicit statement. First, the noise floor is different: Non-contact transducers are subject to drift (temperature, humidity, ambient RF for capacitive; ambient IR, surface reflectivity for optical; clutter and multipath for radar) that is absent or much reduced in a force pad. The Shannon–Hartley numerics of Section 5.6 therefore shift—generally toward lower effective SNR—but the form of the argument is unchanged and the 800:1 nominal ratio against a binary switch remains the correct structural comparison. Threshold re-calibration, on the schedule of Section 5.4, should run more frequently for non-contact deployments than for force-pad deployments; a session-start calibration pass is the conservative recommendation. Second, there is no hard mechanical “zero”: The pad registers 0 N when no finger contacts it, while a proximity sensor registers an ambient background that varies. Engagement-onset detection in the non-contact case therefore requires an adaptive baseline rather than a fixed threshold, and the implementation notes of Section 5.3 expand accordingly. Third, sole-channel rescue, the clinical argument of Section 5.8.1, does *not* in general extend: A user with no visuomotor guidance cannot reliably locate the active volume of a contactless sensor, and the tactile anchor that a force pad provides is in fact doing load-bearing work for static-etiology users whose cortical visual processing is impaired. Non-contact transducers are therefore not a substitute for contact transducers in the sole-channel case; they are a parallel option for the subset of candidate users who have sufficient intact visuomotor coordination to localize to the active volume and who cannot tolerate the contact itself. Fourth, and most importantly for interface-design purposes, a non-contact transducer exposes a spatial *distance* dimension that the force pad does not. For a pressure pad, engagement-state is a scalar function of time (applied force). For a proximity sensor, engagement-state is an ordered pair of (time-of-engagement, distance-at-engagement), and the distance axis is orthogonal to the within-engagement modulation axis of the force-pad case.

The command alphabet of Section 5.4 therefore admits a strictly larger symbol set for a non-contact transducer deployed to a user with intact visuomotor coordination. In addition to the class A (brief engagement), class B (sustained engagement with low modulation), class C (sustained engagement with high modulation), class D (rhythmic brief engagements—the non-contact analogue of rhythmic taps), and class U (unclassified) symbols inherited from the force-pad case, the non-contact variant supports distance-discriminated analogues—sustained-hover-at-near-range, sustained-hover-at-far-range, and lateral-sweep-at-a-held-height—that have no force-pad counterpart. The achievable channel capacity per engagement event, by the Shannon–Hartley argument of Section 5.6, rises accordingly. Quantitative bounds for the specific non-contact hardware variants are deferred to per-platform validation work; the point established here is structural. For users whose motor constraint is primarily the sustained-engagement-with-modulation signature and who retain intact visuomotor coordination, moving from a force pad to a proximity sensor widens the achievable vocabulary by an orthogonal axis, not merely by a multiplicative factor. This is a genuine population-specific gain for a candidate user who can see their fingers land, choose to sustain at a selected distance, rhythmically break engagement, or sweep laterally at a held height, and whose cognitive and visual capacities support the interaction design that exploits the added dimension. Users for whom any of those capacities are not reliably present should be deployed on the contact instantiation of the interface, for the reasons established in Section 5.8.1.

A final implementation note: The non-contact and contact instantiations of the interface are not mutually exclusive. A single deployment can present both—a force pad in the user's hand or on the armrest of their seating system, plus an overhead or bedside proximity sensor—and the detection framework will register engagement events from either transducer using the same R_{SC}-family statistics. For candidate users with fluctuating tolerance (intermittent skin irritation, fluctuating hypersensitivity, post-operative healing windows), a dual-transducer deployment is clinically preferable to forcing a single-instantiation choice at intake.

5.8.6 Prosthesis Users and Direct Myoelectric Deployment

A further set of candidate users have intact or surgically reconstructed neuromuscular signaling in residual upper-extremity musculature but lack a biological end-effector through which the force-pad interface of Section 5.2 could be driven. The population comprises acquired upper-extremity amputees (traumatic, vascular, or oncological etiology), congenital upper-limb difference (transverse or longitudinal reduction), and users of externally powered orthoses for high cervical spinal cord injury or brachial plexus injury. When any of these presentations co-occurs with a communication disorder—ventilator-dependent dysarthria in C1–C4 SCI, post-traumatic apraxia or dysarthria after polytrauma, co-morbid developmental communication disorder in congenital limb difference, or bulbar onset in ALS users who still retain meaningful forearm EMG—direct-selection AAC is clinically indicated and neither the force-pad deployments of Section 5.8.1–Section 5.8.4 nor the non-contact variants of Section 5.8.5 directly fit, because the effector assumed by those deployments is absent or mechanically inaccessible.

Two architectural options follow from the engagement-state generalization of Section 2.4. The first is *prosthesis-mediated* deployment: A myoelectric or body-powered prosthesis user interacts with an ordinary force pad through the terminal device of the prosthesis. The pad records a valid engagement-state signal—applied force as a function of time—and R_{SC}, the press operationalizations of Section 4, and the command alphabet of Section 5.4 are well-formed on the resulting record. Two important confounds are inherited from the prosthesis, however. The user's neuromuscular signal passes through the prosthesis's own control law—EMG thresholding, grasp classifier, deadband, hysteresis, proportional-control gain—before any force reaches the pad, so the within-engagement modulation measured at the pad is a composition of the user's intended modulation and the transfer characteristics of the prosthesis. For a well-characterized prosthesis with documented control parameters, the composition is tractable and the signature remains recoverable; for a poorly-characterized or user-tuned prosthesis, the pad record is best treated as a lower bound on the user's available information capacity. Cable-driven body-powered prostheses introduce a mechanically simpler but stiffness-limited transfer function and a different, typically larger, effective deadband. Prosthesis-mediated deployment is the clinically simpler option for users who already have a prosthesis integrated into their daily function and who do not want to add a parallel interface.

The second option is *direct myoelectric* deployment: Surface EMG electrodes placed over the residual musculature, or over targeted-muscle-reinnervation sites in users who have had TMR surgery, deliver the engagement-state signal to the AAC interface without a prosthesis in the path. Architecturally this is the cleaner match to the framework of this paper, because the composition with prosthesis control law is eliminated. The signal-processing pipeline of Section 5.3 extends by one standard preprocessing stage: bandpass filtering (typically 20–450 Hz for surface EMG), full-wave rectification, and envelope smoothing (root-mean-square or low-pass at 2–10 Hz) yield a per-electrode engagement-state scalar on which the detection framework of Section 4 operates directly. Bio-signal sampling rates typically run at 1000–2000 Hz at the electrode and are downsampled after envelope extraction to rates comparable to the ≥ 100 Hz assumed elsewhere in the paper; the Shannon–Hartley argument of Section 5.6 applies with the substituted SNR and bandwidth parameters (EMG envelope bandwidth is typically lower than force-pad bandwidth; effective SNR varies widely with electrode placement, skin preparation, and user-specific factors).

The structural gain for the direct myoelectric case is that EMG is intrinsically multi-channel. Clinical prosthesis control regularly uses 2–8 electrodes, and research pattern-recognition control uses as many as 16. Each electrode is an independent engagement-state axis. The command alphabet of Section 5.4, written for the scalar force-pad case, extends to a vector-valued engagement state under the same detection logic: The sustained-hold classes B (low modulation) and C (high modulation) from Section 5.4 are defined per electrode, and the full symbol per engagement event is an ordered tuple of per-electrode class labels rather than a single class label. Co-contraction patterns (two or more channels simultaneously active above threshold), sequential activation patterns (channel ordering within a single sustained engagement), and inter-channel amplitude ratios are established vocabulary in the myoelectric prosthesis literature and transfer directly to AAC symbol classes. The per-engagement capacity bound of Section 5.6 rises multiplicatively with electrode count, subject to well-studied inter-channel independence constraints that are not re-derived here. The structural implication for this paper is that multi-channel bio-signal deployment provides the largest vocabulary expansion of any transducer class discussed in Section 5.8, and that the channel-capacity argument against binary-closure switch access is proportionally stronger for these users because a binary switch collapses not one but k information-bearing axes simultaneously.

Three deployment constraints specific to the bio-signal class deserve explicit statement. First, EMG is non-stationary on a session-to-session timescale: Electrode impedance, skin condition, and muscle fatigue shift the amplitude distribution, and per-session calibration is the standard remedy. The refit schedule of Section 5.4 should be a per-session pass for bio-signal deployments, rather than monthly to quarterly as for static-etiology force-pad deployments. Second, intramuscular and implanted electrode deployments fall outside the scope of a non-invasive design paper of this kind; the framework remains valid for those signal sources but their surgical, regulatory, and long-term biocompatibility considerations are not treated here. Third, the sole-channel rescue argument of Section 5.8.1 does not in general apply to bio-signal deployments: This population has, by assumption, neuromuscular signaling the clinical workflow can already recruit, and the AAC-specific question is not whether a channel exists but whether the information capacity of the channel is being used at its full dimensionality. The framework's contribution here is to argue that standard myoelectric prosthesis control is already a force-sensing-analog deployment—a continuous, adequately sampled, sufficiently-high-resolution engagement-state signal—and that AAC architectures that convert this signal into binary switch events before routing it into the communication application discard information the user has already made available at the signal source.

5.8.7 Combat Blast Injury and Veteran Populations

Combat blast injury produces a clinically coherent constellation of impairments that draws on three preceding deployment subsections simultaneously and that the present framework addresses through a single coordinated interface rather than three separate assistive-technology workstreams. The population of interest comprises survivors of improvised explosive device (IED), explosively formed penetrator (EFP), and vehicle-borne IED blast exposure; vehicle and building fire from munition strikes; rocket-propelled grenade and shrapnel injury; thermal and chemical burn injury (including white-phosphorus burns); and penetrating head injury. This presentation is characteristic of veterans of Operation Enduring Freedom, Operation Iraqi Freedom, and Operation New Dawn (OEF/OIF/OND), and of more recent conflicts, and the same constellation occurs in active-duty and civilian blast-injury survivors (industrial explosion, mass-casualty terrorist attack, munitions-factory incident). The paper's framing here is deliberately extendable to civilian cases; the clinical-infrastructure discussion of the Veterans Health Administration polytrauma system of care below reflects where the AAC-engineering community will most often encounter the presentation, not an exclusion of civilian patients.

Three impairment categories co-occur in this population at rates far above those in the general AAC-candidate population, each mapping to one of the preceding subsections. First, burn scarring, skin grafts, ongoing contracture releases, and pressure-garment regimens make sustained physical contact with a force pad intolerable or medically contraindicated on the affected body regions—the non-contact deployment of Section 5.8.5 is indicated for interaction surfaces that would otherwise contact those regions. Second, upper-extremity amputation from primary blast or subsequent surgical revision places many blast-injury survivors in the myoelectric-deployment population of Section 5.8.6, with residual forearm musculature (for transradial amputation) or residual shoulder musculature supplemented by targeted muscle reinnervation (for transhumeral and glenohumeral cases) providing the bio-signal source. Third, blast-induced traumatic brain injury, whether from primary blast overpressure or from tertiary-blast head impact, commonly produces dysarthria, apraxia of speech, or non-fluent aphasia sufficient to make direct-selection AAC clinically indicated even for survivors whose physical recovery is otherwise advanced. The structural point is that a single user frequently presents with all three at once, and that the engagement-state generalization of Section 2.4 permits a single interface architecture—detection framework, command alphabet, and bootstrap protocol held constant across transducer class—to serve the user across all three impairments rather than requiring three separate assistive systems with three separate learning curves.

The distinguishing clinical feature of blast-injury survivors relative to the static-etiology HAI population of Section 5.8.1 is that cortical visual processing and visuomotor coordination are typically preserved, even in cases with severe extremity injury and substantial TBI-related speech and cognitive sequelae. This is not a universal claim—penetrating head injury to occipital or parietal regions, and severe diffuse axonal injury, can compromise visuomotor function—but in the modal presentation, the user can see where a fingertip, prosthetic terminal device, or residual limb lands relative to a target, and can intentionally modulate within an engagement once established. This preservation is the clinical key that makes the distance-dimension vocabulary gain of the non-contact deployment of Section 5.8.5 fully available to this population. The symbol alphabet expansion described in that subsection—sustained hover at selected distance, lateral sweep at a held height, discrete near-far transitions as additional symbol classes—is not a theoretical possibility for blast-injury survivors; it is an immediately usable capacity that scanning-switch AAC deployments systematically fail to capture.

Three deployment considerations specific to this population warrant explicit statement. First, the clinical course after combat blast injury typically involves multiple staged procedures—sequential debridement, split-thickness skin grafts, full-thickness reconstruction, contracture releases, prosthetic fittings and refittings—extending over months to years from injury. Skin tolerance at any given site varies substantially over this period, and the dual-transducer deployment noted at the close of Section 5.8.5—a force pad plus a co-located non-contact transducer, both feeding the same detection framework—is therefore strongly indicated for this population. The user's preferred transducer can migrate between the two as healing phase, scar maturation, and pressure-garment compliance permit, without any change to the command alphabet or to the AAC application downstream. Second, many blast-injury survivors have received extensive assistive-technology workups in the acute and subacute phases and carry standing prescriptions for eye-gaze trackers, speech-generating devices, and adapted switch arrays; the framework of this paper is intended to integrate with, not displace, that infrastructure, and the CRS-R and AAC assessment integration points of Section 5.8.4 apply with the obvious adaptation that blast-injury survivors typically score well above the CRS-R floor and the assessment question concerns throughput rather than residual awareness. Third, a meaningful fraction of this population experiences pain syndromes (phantom limb pain, complex regional pain syndrome, burn-related allodynia) that interact with contact tolerance on a day-to-day basis independent of underlying tissue healing; the dual-transducer recommendation above covers this fluctuation as well.

A further deployment advantage is specific to this population: A nontrivial subset of blast-injury veterans carries pre-existing Morse-code fluency from military signals training (Navy radiomen, Army Signal Corps, and the broader cryptologic-technician and radioman rating communities) or from amateur-radio service before or after military separation. For these users, the rhythmic-tap class of Section 3 constitutes a vocabulary already encoded at motor-cognitive level, and the timing conventions—dot, dash, intra-character gap, inter-character gap, inter-word gap—are already internalized. The rhythmic-tap interface is therefore effectively a no-training-cost vocabulary on day one for this subset, eliminating the training overhead that typically gates novel AAC adoption and permitting the user to transmit linguistically complete content from the first session on the pad.

The natural clinical deployment contexts for this population are the Veterans Health Administration polytrauma system of care (the five Polytrauma Rehabilitation Centers and the broader Polytrauma Network Sites), the Department of Defense (operating as Department of War) rehabilitation research pipeline (the U.S. Army Institute of Surgical Research, the Walter Reed National Military Medical Center, and the Center for the Intrepid at Brooke Army Medical Center), and the DARPA-adjacent research programs on upper-limb prosthetics and neural interfaces (RE-NET, HAPTIX, and their successors). The framework of this paper is offered to those communities as a design contribution. Civilian burn unit networks, and the American Burn Association verified burn centers more broadly, are the natural civilian-side deployment context and encounter the same impairment constellation from non-combat etiology. The channel-capacity argument, the detection framework, and the deployment specifications are identical across military and civilian cases; only the clinical-pathway description differs.

5.8.8 Epilepsy and Rhythmic-Involuntary-Motor Populations

A large fraction of the candidate AAC population for this interface carries a co-occurring epileptic or rhythmic-involuntary-motor diagnosis. Severe hypoxic-anoxic brain injury (Section 5.8.1), post-status-epilepticus syndromes, severe cerebral palsy, and the progressive neuromuscular etiologies of Section 5.8.2 are each associated with elevated rates of clonic seizure activity (rhythmic 2–3 Hz jerking), myoclonic jerks (brief 1–5 Hz discrete contractions, either generalized or multifocal), interictal and post-ictal myoclonus, and—in post-cardiac-arrest and post-near-drowning survivors specifically—Lance-Adams (post-anoxic) myoclonus, often refractory and triggered by voluntary movement. Progressive myoclonic epilepsies (Unverricht-Lundborg disease, Lafora body disease, the neuronal ceroid lipofuscinoses, Ramsay Hunt syndrome) and the generalized epilepsy syndromes add a further population whose motor signature includes action myoclonus at frequencies overlapping voluntary modulation. For these users the epilepsy is not a side condition but a continuous, session-to-session modulator of what the pad records. An interface that cannot separate involuntary rhythmic contamination from intentional modulation will misattribute seizure activity to user engagement, and—symmetrically—suppress seizure-obscured intentional modulation as noise.

A clinically severe and under-discussed subpopulation within this group deserves explicit treatment: users whose electroencephalographic record shows continuous or near-continuous subclinical seizure activity—many discrete mini-seizures per second, rather than discrete ictal events separated by interictal quiet. These users present clinically without the dramatic generalized motor features of a grand mal event but with a background of superimposed involuntary micro-twitches that contaminate every motor output. The operational consequence for this interface is distinctive: Sustained contact is interrupted by many brief velocity reversals per second rather than the occasional involuntary twitch the bridging rule of Section 4.5.1 is written for, and a single intentional press event can itself contain multiple overlapping involuntary contractions that corrupt the within-contact force profile and the onset count. This failure mode is rarely discussed in the switch-access literature because binary-closure assessment collapses to a null result for these users without ever surfacing the structure that produced the null: Every attempted press is read as either a noisy closure or no closure at all, and the user is recorded as lacking reliable voluntary motor control. The continuous-output framework of this paper is the first interface architecture under which the distinction between intentional modulation and continuous involuntary contamination is even representable as a measurement question. Accommodations for this subpopulation include: longer Section 4.8 unprompted baseline windows (5–10 minutes rather than 2–3) to characterize the dense involuntary spectrum; a prolonged per-session calibration phase to fit the within-contact force-modulation classifier against a much higher involuntary-power floor; per-user adjustment of the Section 4.5.1 bridging rule to tolerate multiple sub-300 ms velocity reversals within a single sustained-contact interval rather than flagging each reversal as an event boundary; and, critically, statistical rather than deterministic attribution of intentional modulation—the intentional signal is established as a task-contingent shift in the distribution of within-contact force statistics relative to the unprompted baseline, not as a set of clean isolated press events. For users at this extreme, the interface is less a direct input channel than a continuous estimator of intentional-motor probability, and downstream AAC applications must be specified accordingly.

The frequency-domain structure of the relevant phenomena is the operational key. Clonic seizure activity concentrates at 2–3 Hz; generalized and multifocal myoclonic jerks span 1–5 Hz; clonus occupies 5–8 Hz; physiological tremor occupies 4–12 Hz; action myoclonus can extend higher. Voluntary force modulation, by contrast, is bandlimited to approximately 0–8 Hz with most content below 5 Hz (Section 3). The involuntary and voluntary bands therefore overlap substantially in the 1–8 Hz range, and separation cannot be achieved by a fixed global frequency cut. It is achieved instead by per-user characterization: The power spectral density of the force signal during the user-present-but-unprompted segment of the Section 4.8 diagnostic protocol captures the individual's involuntary spectrum, and subsequent intentional-engagement windows are compared against that per-user baseline rather than against a population-level filter. This is why the hardware sampling floor is specified at ≥ 25 Hz (Section 5.1, item 3): The Nyquist ceiling of 12.5 Hz must cover the full 1–12 Hz involuntary band, not merely the voluntary band, because characterization of the contaminant is as essential as capture of the signal.

Four design accommodations follow for deployment in this population. First, the Section 4.8 diagnostic protocol is mandatory rather than recommended: Every session begins with a no-user baseline and a user-present-unprompted segment, with the per-user involuntary power spectral density reported alongside R_{SC}. Second, the within-contact force-modulation classifier of Section 5.3 is tuned per user to reject energy concentrated at the peaks identified in that baseline, treating spectral energy at the user's individual clonic or myoclonic frequencies as confound rather than signal. Third, where the user carries concurrent electroencephalographic monitoring—routine in long-term video-EEG epilepsy monitoring units, and continuous for patients with implanted responsive neurostimulation systems (NeuroPace RNS) or chronic ambulatory EEG—the EEG annotation stream is used to flag ictal and immediate post-ictal windows, during which R_{SC} and command-alphabet outputs are suppressed or separately tagged rather than interpreted as user intent. Fourth, the rhythmic-tap class of Section 3 requires particular care in this population: The per-user intra-burst interval calibration must be established during a seizure-free baseline window, and candidate rhythmic-tap sequences whose inter-tap intervals align with the user's documented clonic or myoclonic frequency are flagged for review rather than automatically accepted as intentional bursts. The bridging-and-flagging rule of Section 4.5.1—treating sub-300 ms velocity reversals bracketed by sustained contact as involuntary-twitch candidates—applies with particular force here and provides a second, independent check on the same phenomenon.

The natural clinical deployment contexts for this population are National Association of Epilepsy Centers Level 3 and Level 4 epilepsy centers, pediatric and adult neurology services managing drug-resistant epilepsy, and post-anoxic rehabilitation services receiving post-cardiac-arrest and post-near-drowning survivors (with substantial overlap with the hypoxic-anoxic-injury cohort of Section 5.8.1). For patients with implanted responsive neurostimulation or deep brain stimulation systems, the continuous intracranial EEG stream already captured by those devices constitutes an ictal-detection channel that the interface can consume directly where clinical arrangements permit. The framework does not require this coupling but benefits from it where available. The structural claim of this subsection is that epilepsy is not a peripheral confound to be footnoted but a first-class deployment scenario for a substantial fraction of the target population, and that the interface’s frequency-domain accommodations and the Section 4.8 diagnostic protocol are what make a usable continuous-output channel available to a cohort for whom binary-switch assessment has been confounded by involuntary rhythmic activity across decades of clinical record.

5.8.9 Related Populations

Additional candidate populations—late-stage Huntington's disease, severe cerebral palsy with dystonia or athetosis, rare progressive encephalopathies—share the underlying motor-signature structure but each introduce population-specific confounds (choreatic overlay on modulation measurement, involuntary movement contamination of the engagement-state series, etiological heterogeneity) that require dedicated treatment beyond the scope of this paper. The detection framework's sensitivity to each confound is an open empirical question for per-population validation work.

---

6. Design Limitations, Open Empirical Questions, and Proposed Validation Studies

The paper's claims are design and methods claims. Four distinct categories of limitation and future work are listed separately below: inherent limitations of the design itself (Section 6.1), open empirical questions that require data this paper does not supply (Section 6.2), related work on involuntary-motor bridging (Section 6.3), and concrete validation study designs that would address those questions (Section 6.4).

6.1 Design Limitations

The following are limitations of the approach specified in Section 4–Section 5 and are not resolvable by additional empirical work within this design:

- **Contact-transducer instantiation requires contact tolerance.** The force-pad instantiation requires the user to tolerate sustained contact with a pad surface. Users with severe tactile defensiveness, skin fragility, or medical contraindications to sustained hand positioning are excluded from the contact instantiation specifically; the non-contact transducer variant (Section 5.8.5) and bio-signal variant (Section 5.8.6) address these populations, but each introduces its own deployment constraints documented in those subsections.
- **Motor output must have modulation structure.** A user whose sole motor output is a uniform silent hold with no within-hold modulation above pad noise is not served by the full command alphabet of Section 5.4. The single-class fallback path in Section 5.5 supports a reduced channel for such users but cannot recover the full binary bootstrap.
- **Pad mounting geometry is population-specific.** The hardware specification (Section 5.2) describes pad electronics and sampling parameters, not mounting and access geometry. For users with atypical wrist kinematics (Section 5.8.3), pad mounting is a per-user fitting problem the design does not solve.

- **Calibration requires a baseline session.** Threshold fitting in Section 5.4 requires a calibration session with the user engaged; for users who cannot tolerate an initial calibration, the interface is excluded until a baseline is achievable. A passive-ambient calibration mode is plausible but not specified.
- **Progressive-disease users require re-calibration.** The thresholds in Section 5.4 are quasi-stationary; ALS and similar users need a re-calibration schedule that this paper mentions but does not specify in operational detail (Section 5.8.2).
- **The bootstrap requires two spontaneous classes.** Section 5.5 specifies a code-book transfer or invention contingent on the user spontaneously producing both modulated and silent sustained holds. Users who produce only one class are served by the fallback in Section 5.5 at reduced channel capacity.

6.2 Open Empirical Questions

The following are questions the paper does not resolve and that require data or user-facing studies to answer.

Not addressed by this paper.

- How prevalent is the sustained-contact-with-modulation signature in HAI, and in related diagnostic categories (advanced ALS, post-status-epilepticus, late-stage Huntington's, severe cerebral palsy with retained tactile response)? This requires an epidemiological study with systematic instrumented-interface assessment across a recruited cohort.
- Does tap fidelity empirically degrade before sustained-contact-with-modulation output in ALS hand function? The prediction (Section 5.8.2) is mechanistically grounded—coordinated agonist–antagonist timing for tap release should fail ahead of sustained agonist tone—but direct empirical documentation of this temporal dissociation in the specific modality of hand-pad interaction has not been published. A within-subject longitudinal comparison of single-switch-assessment pass rate against the sustained-contact ratio R_{SC} across disease progression would resolve this question.
- Does the proposed interface work when deployed on live users? The present paper is a specification; validation on instrumented pads across a recruited cohort is future work (Section 6.4).
- How should the interface handle a user whose signature drifts across the day due to fatigue, arousal, or seizure proximity? Longitudinal stability of the sustained-engagement-with-modulation signature over hours-to-weeks timescales is an open empirical question across all three transducer classes.
- For progressive-disease users, what is the expected rate of threshold drift in the Section 5.4 command alphabet, and what re-calibration schedule matches it? Section 5.8.2 flags this as a standing deployment requirement but does not specify the interval.
- What is the inter-rater reliability of the “press” label when applied to the same video by independent human coders rather than by the detection framework of Section 4? Agreement statistics between human coders and between human coders and the automated pipeline are a standard validation target for any empirical application of the framework.
- Does external cueing (auditory metronome, visual pacer, tactile buzz) increase the modulation rate or the consistency of modulation within a held contact? This bears directly on whether an interface should drive the user with a pacing stimulus or passively read ambient activity. A controlled comparison across cued and uncued conditions is required to resolve it.
- Do sustained-contact signatures in HAI co-vary with lesion localization on structural MRI? HAI injuries are known to affect hippocampus, basal ganglia, and cortex with selective vulnerability that varies by hypoxia severity, duration, and age at injury [9]. Sustained-contact motor output requires some degree of cortical-motor circuit sparing; the prediction that the signature is more commonly preserved when primary motor cortex and corticospinal tract are relatively spared, and more commonly absent when those regions are affected, is testable against any cohort with paired motor-behavior and MRI data.
- Hand-tracker sensitivity. All analyses in the detection framework use MediaPipe Hands at the configuration specified in Section 4.2. A sensitivity analysis that reprocesses a validation corpus with alternative trackers (Apple VisionKit, MediaPipe Holistic v2, YOLO-v9-hand) would bound the sensitivity of R_{SC} and R_{SC}^{*} to tracker choice. The qualitative finding is expected to survive because onsets-inside (the numerator) depends only on the union of detected contact intervals, not their boundary precision, but the magnitude of the estimate may drift.
- Audio-onset detector choice. The framework specifies librosa's default complex-domain onset strength function. A sensitivity analysis with Essentia's superflux, an energy-based detector, and a learned CNN-based detector (e.g., OnsetsAndFrames) would bound the dependence of R_{SC}^{*} and the within-hold modulation estimates on the onset operator. The tap-like null ($R_{SC} \approx 1$) is operator-independent by construction, but the absolute numerator may drift.

6.3 Related Work on Involuntary-Motor Bridging

The general principle of filtering involuntary motor output within a continuous input stream predates the specific implementation of Section 4.5.1. Operating-system accessibility layers (Windows FilterKeys, Apple Touch Accommodations) apply time-window filters to suppress repeated or tremor-driven input. Eye-gaze AAC dwell-click filters bridge saccadic and tremor micro-movements within fixations, and the adaptive-input HCI literature (Wobbrock, Hurst, and colleagues) treats motor variability through gap-and-error tolerance schemes. What is new in the present framework is threefold: (i) threshold symmetry, in which the tap-duration ceiling of Section 3 is reused as the bridging threshold, so that a single motor-physiology constant governs both “what is a tap” and “what micro-release may be involuntary within a hold”; (ii) the bridge-and-flag stance, which preserves involuntary-twitch counts as a session-level statistic feeding the diagnostic protocol of Section 4.8 rather than silently discarding them as prior accessibility filters do; and (iii) application to switch-access AAC specifically, where the dominant interface paradigm has historically collapsed to binary at the hardware level before any such rule could apply.

6.4 Proposed Validation Study Designs

Three compact study designs would address the central empirical questions raised above. Each is specified at the level of sample size, primary endpoint, and success criterion; full protocols are outside the scope of this paper.

Study A—Prevalence screening in static-etiology populations. Cross-sectional recruitment of candidate HAI users through rehabilitation centers, $n \approx 30$–50. Each participant completes a standardized 10–15 minute instrumented-pad session under an established caregiver-prompted interaction protocol. *Primary endpoint:* fraction of participants for whom $R_{SC}^* > 2$ with bootstrap 95% CI excluding 1. *Secondary endpoints:* inter-rater reliability of press coding between two independent raters (Cohen's $\kappa \geq 0.7$ target); agreement between manual coding and the automated pipeline on contact-interval count (intraclass correlation ≥ 0.8 target). *Success criterion:* a prevalence estimate with confidence bounds that informs the next-stage study sizing.

Study B—Progressive-disease longitudinal. ALS users recruited through AAC clinics or patient-advocacy channels at the disease stage at which single-switch scanning has been identified as approaching clinical threshold. $n \approx 15$–20, monthly instrumented-pad sessions over 12 months. *Primary endpoint:* within-subject trajectory of R_{SC}^* and CV_F across the 12-month window, indexed against a concurrent single-switch-assessment pass rate. *Secondary endpoints:* time in months from first study visit to force-modulation-channel drop-out; dyad-reported communicative utility of the force-sensing interface versus the legacy binary switch during the overlap window. *Success criterion:* evidence that $R_{SC}^* > 2$ is maintained for a clinically meaningful window (≥ 6 months) after single-switch-assessment pass rate falls below 0.8.

Study C—Bootstrap feasibility. Single-subject experimental design replicated across $n \approx 8$ participants drawn from Study A's $R_{SC}^* > 2$ subset. Over a 4-week dyad-mediated training phase, the caregiver delivers consistent environmental consequence for high-modulation versus low-modulation sustained holds according to the bootstrap protocol in Section 5.5. *Primary endpoint:* whether the dyad converges on a reliable binary code-book, operationalized as caregiver–user agreement ≥ 80% across 20 consecutive trials in week 4. *Secondary endpoint:* which semantic-pole assignment (high-modulation = yes vs. high-modulation = no) converges faster, informing deployment defaults. *Success criterion:* ≥ 50% of participants reach the agreement threshold within 4 weeks, with qualitative analysis of the non-converging subset to inform protocol refinement.

---

7. Conclusion

The paper specifies a continuous-output switch-access interface—instantiated as a force-sensing pad, a non-contact proximity transducer, or a direct myoelectric front end—that preserves engagement onset, engagement duration, and within-engagement modulation, together with a detection framework for identifying sustained-engagement-with-modulation motor signatures from video, audio, and bio-signal data, and a proposed binary (yes/no) bootstrap protocol derived from modulated-vs-silent sustained holds. The structural argument—that a binary-closure switch is information-destructive by construction for any user whose motor output lives on the within-contact modulation dimension—is established from channel-capacity math: The nominal capacity ratio against a non-thresholded continuous channel is 800:1 for a 1-second contact at the specified sampling parameters, and under realistic pad electronics (20 dB SNR, 10 Hz motor bandwidth) the Shannon–Hartley ratio is ~ 66:1, falling to ~ 17:1 only under aggressively pessimistic assumptions (10 dB SNR, 5 Hz bandwidth). Deployment is carried through to eight scenarios (Section 5.8): static-etiology users for whom the interface is a first channel rather than a bridge; progressive-disease users for whom it is a bridge channel through the window between tap failure and eye-gaze transition; users whose primary constraint is distal upper-extremity force generation or atypical wrist kinematics; integration with standard clinical assessment batteries; a non-contact transducer variant (capacitive proximity, optical ranging, millimeter-wave radar) for users for whom physical contact is contraindicated by tactile hypersensitivity, skin compromise, or pressure-sore risk, which additionally exposes a distance dimension that widens the achievable symbol alphabet for users with intact visuomotor coordination; prosthesis-mediated and direct myoelectric deployments for users with upper-extremity amputation or orthotic limb use, in which multi-channel EMG expands the symbol alphabet by one axis per electrode and the argument against binary-closure switch access is structurally stronger in proportion to electrode count; combat blast-injury survivors, a cross-cutting population in whom burn scarring, upper-extremity amputation, and blast-induced communication disorder frequently co-occur and for whom preserved visuomotor coordination makes the non-contact variant's distance-dimension vocabulary gain immediately usable; and epilepsy and rhythmic-involuntary-motor populations, including the severe case of continuous subclinical seizure activity, for whom frequency-domain separation of involuntary rhythmic contamination from intentional modulation—supported by the mandatory per-session diagnostic protocol of Section 4.8 and by the ≥ 25 Hz sampling floor that covers the full 1–12 Hz involuntary band—is the operational key. The specification is offered as an engineering and methods contribution. Empirical validation across recruited cohorts is flagged as future work with three concrete study designs proposed (Section 6.4).

Acknowledgments

This work was performed at the Cara Voice Lab using commodity hardware. The author thanks the AAC, rehabilitation engineering, and information-theory research communities whose work is cited herein.

Declarations

Ethics Approval

Not applicable. This paper presents a design and methods contribution; no human-subjects data were collected, and no participant recruitment was performed. No institutional review board (IRB) or ethics committee approval was required.

Data Availability

This is a design paper; no empirical data are reported. Reference implementation of the detection framework described in Section 4 is available on request—contact `carajadecatalano@gmail.com`.

Code Availability

Reference implementations of the detection framework (MediaPipe contact-interval extraction, librosa onset detection, loop-subtraction phase fitting, bootstrap resampling utilities for a future validation study) are available on request under a permissive license (BSD-3-Clause) and are intended for public release alongside an empirical validation paper. Dependencies: `numpy`, `scipy`, `pandas`, `mediapipe`, `librosa`, `matplotlib`. Python 3.11+.

Conflicts of Interest

The author declares no financial conflicts of interest.

Funding

This work received no external funding. All computational work for this paper was performed on the author's personal laptop and on a self-funded local GPU workstation on the author's home network; no cloud compute was used for the WP16 analysis pipeline. No institution, foundation, or commercial entity contributed to the study design, data analysis, or writing.

Author Contributions

Single author. CRediT taxonomy: conceptualization, methodology, software, formal analysis, investigation, data curation, writing—original draft, writing—review & editing, visualization, supervision, project administration—all C.J.C.

Use of Generative AI

The analysis scripts, statistical methods, and interpretive text were authored by the author. A large language model (Claude Opus) was used as a drafting and editing assistant, including in the preparation of the manuscript prose and the rigor-upgrade review passes visible in the document's version history. The author reviewed and verified all numerical claims against regenerated source files before each commit; no numerical claim in this paper was accepted on the LLM's assertion alone. All computational components (MediaPipe landmark extraction, librosa onset detection, contact-interval segmentation, loop-subtraction phase fitting) specified in the detection framework are reference-implemented as named scripts on deterministic seeds; the numerical outputs of any empirical validation performed with this framework are traceable to those scripts and not to LLM output.

References

References cited in this paper.

1. Beukelman, D. R., & Mirenda, P. (2013). *Augmentative and Alternative Communication: Supporting Children and Adults with Complex Communication Needs* (4th ed.). Baltimore: Paul H. Brookes Publishing.
2. Giacino, J. T., Kalmar, K., & Whyte, J. (2004). The JFK Coma Recovery Scale-Revised: Measurement characteristics and diagnostic utility. *Archives of Physical Medicine and Rehabilitation*, 85(12), 2020–2029.
3. Pignat, J.-M., Mauron, E., Jöhr, J., Gilart de Keranflec'h, C., Van De Ville, D., Preti, M. G., Meskaldji, D. E., Hömberg, V., Laureys, S., Draganski, B., Frackowiak, R., & Diserens, K. (2016). Outcome prediction of consciousness disorders in the acute stage based on a complementary motor behavioural tool. *PLOS ONE*, 11(6), e0156882.
4. Zhang, F., Bazarevsky, V., Vakunov, A., Tkachenka, A., Sung, G., Chang, C.-L., & Grundmann, M. (2020). MediaPipe Hands: On-device Real-time Hand Tracking. *arXiv*:2006.10214.
5. McFee, B., Raffel, C., Liang, D., Ellis, D. P. W., McVicar, M., Battenberg, E., & Nieto, O. (2015). librosa: Audio and music signal analysis in Python. In *Proceedings of the 14th Python in Science Conference* (pp. 18–25).
6. Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423; 27(4), 623–656.
7. Sigafoos, J., Drasgow, E., Reichle, J., O'Reilly, M., & Tait, K. (2004). Tutorial: Teaching communicative rejecting to children with severe disabilities. *American Journal of Speech-Language Pathology*, 13(1), 31–42.
8. Dahiya, R. S., & Valle, M. (2013). *Robotic Tactile Sensing: Technologies and System*. Dordrecht: Springer.

9. Caine, D., & Watson, J. D. G. (2000). Neuropsychological and neuropathological sequelae of cerebral anoxia: A critical review. *Journal of the International Neuropsychological Society*, 6(1), 86–99.
10. Kiernan, M. C., Vucic, S., Cheah, B. C., Turner, M. R., Eisen, A., Hardiman, O., Burrell, J. R., & Zoing, M. C. (2011). Amyotrophic lateral sclerosis. *The Lancet*, 377(9769), 942–955.
11. Ball, L. J., Beukelman, D. R., & Pattee, G. L. (2002). Timing of speech deterioration in people with amyotrophic lateral sclerosis. *Journal of Medical Speech-Language Pathology*, 10(4), 231–235.
12. Beukelman, D., Fager, S., Ball, L., & Dietz, A. (2007). AAC for adults with acquired neurological conditions: A review. *Augmentative and Alternative Communication*, 23(3), 230–242.
13. Fried-Oken, M., Mooney, A., & Peters, B. (2015). Supporting communication for patients with neurodegenerative disease. *NeuroRehabilitation*, 37(1), 69–87.

How to Cite

> Catalano, C. J. (2026). *Design of a Continuous-Output Switch-Access Interface for Users Whose Structured Motor Signature Is Sustained Engagement with Modulation: Detection Framework and Channel-Capacity Argument.* Cara Voice Lab Working Paper Series, WP16, Draft v1.0.0 (2026-04-16).

BibTeX:

```
@techreport{catalano2026wp16,
  author = {Catalano, Cara Jade},
  title  = {Design of a Continuous-Output Switch-Access Interface for Users
            Whose Structured Motor Signature Is Sustained Engagement with
            Modulation: Detection Framework and Channel-Capacity Argument},
  institution = {Cara Voice Lab},
  type   = {Working Paper},
  number = {WP16},
  year   = {2026},
  note   = {Draft v1.0.0, 2026-04-16}
}
```

Appendix A. Reproducibility

A.1 Reference Implementation

Reference implementations of the detection framework components (contact-interval extraction, onset co-registration, R_{SC} and R_{SC}^* computation, bootstrap resampling with fixed PRNG seed) will accompany an empirical validation paper. All randomized operations in the reference implementation use fixed PRNG seeds so outputs regenerate byte-identically.

A.2 Justification of the 25th-Percentile $y_{contact}$ Threshold

The per-session contact threshold y_{contact} defined in Section 4.5.1 is specified as the 25th percentile of index-fingertip vertical coordinate across hand-present frames. The specification is motivated on geometric grounds: The 10th-percentile is too permissive (treats nearly all hand-present frames as in-contact, collapsing the contact-interval count toward the total hand-present duration and losing the distinction between "finger in pad region" and "hand resting off-pad"); the 50th-percentile (median) is interior-splitting by construction (half of hand-present frames fall on each side) and too strict for sessions where contact-time dominates the per-session y distribution. The 25th-percentile threshold is a principled compromise: permissive enough to capture clear contact events, strict enough to preserve a contact / non-contact split that discriminates pressed-and-held from hand-hovering. A validation study should sweep {10, 20, 25, 30, 40, 50}-th percentile thresholds, gap-tolerance values {1, 2, 3, 5} frames, and onset-membership windows {0.05, 0.10, 0.15, 0.20} s and report the sensitivity of R_{SC}^* to each. The 200 ms gap tolerance (at most two consecutive zero-frames at the 10-Hz sampling rate, inclusive) is specified against typical MediaPipe detection-dropout rates under normal lighting. The ±100 ms onset-membership window is set by the minimum of (a) librosa spectral-flux onset backtrack latency and (b) 10-Hz landmark-sampling granularity.