

FORTHwrite / FORTHspoken: A Position Paper on Defined Command Semantics for Voice Interfaces

Earl H. Moreo III, SFO — Independent Researcher, Dayton, Ohio
Radio amateur (N6FXB); FCC Commercial General Class licensee; retired electrician;
veteran, Army Security Agency and Navy signals intelligence

Draft of 11 July 2026 — prepared for the W3C Voice Interaction Community Group

Summary. Current voice assistants fail their users not primarily through poor speech recognition but through *undefined command semantics*: the words users speak have no rigorously specified meaning, so the machine feels free to reinterpret them. This paper proposes design laws for a spoken command language with defined semantics, motivated by a blind user's daily experience, and grounded in two engineering traditions that solved analogous problems decades ago: radiotelephony procedure and early programming-language design.

The Motivating Failure

A blind reader uses a commercial voice assistant as her primary reading interface. When she says “stop,” the device goes silent, waits a short interval, and then resumes playing on its own initiative — sometimes the very material it had been reading — destroying her place in the book. Recovering that place, without sight, is an agony. The failure is not acoustic; the device heard the word correctly. The failure is semantic: “stop” is not a defined term in the device's language. It is a suggestion the machine reinterprets, and silence is treated as a timer rather than a state. For sighted users this is an annoyance; for a blind user, for whom the spoken channel is the entire interface, it is loss of control of the instrument itself.

Proposed Design Laws

1. The Silence Triad. Every suspension of machine activity must have a defined type, and the type vocabulary must exhaust the space. Three commands suffice, inherited (deliberately) from FORTRAN's statement repertoire: **PAUSE** — suspend, preserve all state, *no clock*; silence is stable indefinitely and ends only on an affirmative human word (CONTINUE). **WAIT n** — suspend with a counter; the only command in which time is a legal actor, because the human set the value. **STOP** — terminate the turn; silence is final until a human opens a new turn. The taxonomy is complete: no clock, human-set clock, no resumption.

2. The No-Self-Resumption Prescription. The machine never self-resumes. Silence after PAUSE or STOP is a stable state, never a countdown. The commercial failure described above is precisely the violation of this rule: a timer attached to a command that never asked for one.

3. The Self-Command Prohibition. The machine must never obey its own output. A device reading a novel aloud will speak the words “stop,” “pause,” and “wait” on nearly every page; none of these may carry command force. Command authority rides on the human channel only.

Reserved words therefore require a command channel — candidate mechanisms include a vocative address prefix and radio-style procedure words legal only in command position.

4. The Minimal-Pair Prohibition. No two spoken command words may differ by a single phoneme. The eye tolerates minimal pairs (per-/pre-); the ear, under noise, fatigue, accent, echo, or an inexpensive microphone, does not. One phoneme of separation is zero margin.

5. The Reference Dialect and the Robustness Standard. A spoken command language must be defined against a reference pronunciation (proposed: Standard American, in the maximally articulated register of a national spelling bee final), and its vocabulary must meet the standard of the NATO phonetic alphabet, whose words were selected by testing under noise and distortion: *even when mangled, the command must still be parseable*. Speakers may deviate; the language is measured at the reference.

6. Proportionate Overhead. Speech runs at a few words per second and no hardware advance will ever accelerate the human larynx; unlike the packet layers beneath it, the spoken layer can never outrun its own protocol ceremony. Every syllable of addressing, confirmation, or disambiguation is therefore perceptible cost, and must be proportionate to the cost of the error it prevents. A two-word envelope on a one-word payload (“Scottie, stop”) is justified where a false trigger costs a blind reader her place; a full confirmation ceremony on every routine command builds the opposite failure, a language too heavy to speak.

7. The Lexicon as Codebook. A slow channel is redeemed by encoding density: one predefined syllable can carry a paragraph of specified meaning, exactly as radiotelegraphy’s Q-codes and prosigns compress full sentences into two or three characters over channels of a few words per minute. The compression works only because both stations share the codebook in advance. A published lexicon with fixed semantics is therefore not documentation of the language — it is the language, and the reason a one-syllable command can be trusted.

Two Carriers, One Grammar

The proposal specifies one command grammar in two cognate carriers: **FORTHwrite (FW)** for keyboarded interaction and **FORTHspoken (FS)** for voice. The distinction is a protocol-layering discipline: the command semantics live at the application layer and do not care about the carrier, while each carrier imposes its own physical-layer constraints — FS carries the phonetic-distance and reference-dialect requirements above; FW may relax them (the eye disambiguates what the ear cannot) but shares the reserved-word and state-semantics rules. Naming note: FORTH supplies the name (and a pun on *forthright* and *outspoken*); FORTRAN supplies the form — statement labels for addressable dialogue history, a WRITE(unit, format) filing idiom, and an END terminator.

Relation to Existing Work

The February 2026 W3C Workshop on Smart Voice Agents identified unresolved cross-cutting issues in voice-agent design, including interaction-design failures that break conversational flow — the family of failure this paper addresses at the command-semantics root. The author is aware of

ETSI ES 202 076 (generic spoken command vocabulary for ICT devices) and of W3C WAI's Natural Language Interface Accessibility User Requirements, and offers the present design laws as complementary: prior vocabulary work standardizes *which words*; this proposal standardizes *what the words are guaranteed to mean*, with state semantics (the Silence Triad), safety prohibitions (no self-resumption; no self-command), and channel-robustness requirements (minimal-pair prohibition; reference dialect; mangled-but-parseable) stated as testable rules.

Method and Status

This work is developed by an independent researcher in structured, documented human–AI working sessions (the author's “AI colloquy” method), with a maintained lexicon, design-law register, and verbatim session records. The design laws above are drafts; open questions (unit and format assignments, disambiguation conventions for near-homophones, counter-cancellation rules, and the reserved-word channel mechanism) are docketed for further specification. The author welcomes correspondence, criticism, and collaboration, and offers the core design laws for consideration in any current or future W3C voice-interaction standardization effort.

END