BEING REASONABLE IS NOT ENOUGH: SELECTIVE RELATIONSHIP ADAPTATION IN FULL-DUPLEX SPEECH AGENTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Full-duplex speech agents must decide not only whether a conversational moment permits a response, but whether the same moment warrants a different action for a friend, supervisor, or task owner. Ordinary action accuracy cannot isolate this capability because a locally acceptable default may ignore the relationship condition. We introduce the Relationship Permission Boundary (RPB), a paired counterfactual benchmark that holds decision-preceding audio fixed while changing only the interlocutor relationship. Directional Correct Adaptation (DCA) requires both sides to be acceptable and any action change to follow the reviewed direction. On 79 human-reviewed strong pairs, three native audio systems obtain only 0–3.8% DCA although condition-level acceptable accuracy reaches 31.6–75.3%; a constant `WAIT` policy reaches 88.0% Acceptable with DCA 0. A concise training analysis identifies three recurring shortcuts: local fit without selective transfer, direction without invariant safety, and safety through zero-change collapse. We then train a detachable relationship FiLM router, speech-activity feature encoder, and paired action head using frozen MiniCPM-o-4.5 text scores. On a participant-isolated AMI Train-only endpoint, all three seeds obtain DCA 1, balanced safe utility 1, coverage 0.444444, exact 1, and zero harm or invalid change. The frozen base has DCA 0 and utility 0.638889; an equal-capacity relation-blind control has DCA 0 and utility 0.666667. Both nine-meeting recording-blocked tests give $p = 0.0020999$. Training and confirmation share a deterministic four-state/action mapping, so this is a policy-contract result, not evidence of learned human permission boundaries or audio-conditioned adaptation. A post-hoc exploratory audit supports dependence on supplied state and side binding; acoustic-feature nulling is unchanged ($p = 1.0$). The contribution is an auditable benchmark specification, a shortcut diagnosis, and bounded evidence of structured-state use. Independent human-policy validation and deployment remain outside this study.

1 INTRODUCTION

A real-time speech agent continually chooses among waiting, backchanneling, taking the turn, and interrupting. Acoustic timing constrains these actions, but it does not completely determine them. The same pause can invite a contribution from a close collaborator while still calling for restraint from a subordinate who lacks decision authority. A socially conditioned agent therefore needs a selective policy: change when relationship information changes the permission boundary, and remain stable when it does not.

Existing full-duplex evaluation has made overlap, latency, interruption, and instruction following measurable (Lin et al., 2025; Tang et al., 2026; Mazumdar et al., 2026). Those capabilities are necessary, but they do not establish whether a model uses social context in the correct direction. The distinction matters because condition-level accuracy is easy to inflate with action priors. If waiting is usually permissible, a constant waiting policy can appear strong even though it never reads the relationship prompt.

RPB turns this ambiguity into a paired behavioral test. Each episode supplies one fixed decision-preceding audio context and two relationship conditions. The Gold annotation is a permission profile

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

rather than a single mandatory action. We then distinguish three questions: Is each selected action locally acceptable? Does the model change across the pair? If it changes, does it move in the reviewed direction without becoming invalid? Invariant relationship pairs provide the complementary negative control: prompt reactivity is harmful when the reviewed policy should remain unchanged.

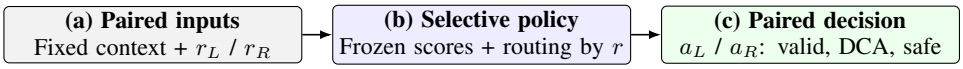


Figure 1: Fixed context + paired relationships; frozen-score relationship routing; local validity, DCA, and invariant safety.

Our contributions are:

- a paired counterfactual benchmark with decision-preceding inputs, reviewed acceptable and preferred distributions, directional adaptation, and invariant-safety metrics;
- a controlled evaluation showing that selective relationship use is a distinct capability that neither high condition-level validity nor generic prompt sensitivity can capture;
- a concise synthesis of failed-training lessons that distinguishes local fit, directional transfer, coverage, and invariant safety; and
- a sealed policy-contract result on AMI-derived inputs in which a hot-plug adapter outperforms frozen MiniCPM-o-4.5 text scoring and an equal-capacity relation-blind control under a shared fixed state/action mapping, together with a post-hoc exploratory intervention audit.

The scope of the claim is precise. We diagnose a shared failure across the tested native systems and demonstrate structured-state use under a separate, policy-generated contract. The latter does not establish a remedy for the native-audio benchmark failure. Architecture ranking, a human ceiling, and a general impossibility result for multimodal training are outside this study.

2 Related Work

Human turn organization. Turn-taking is locally managed and recipient-sensitive, rather than a sequence of context-free silence thresholds (Sacks et al., 1974; Duncan, 1972). Gaps, overlap, and turn timing vary systematically across interactional settings (Schegloff, 2000; Heldner & Edlund, 2010; Stivers et al., 2009; Levinson & Torreira, 2015). Listener responses are also active contributions: backchannels can preserve another speaker's floor and shape the unfolding account (Yngve, 1970; Bavelas et al., 2000). RPB adopts this distinction operationally. Audio defines an opportunity at a fixed cutoff; the relationship condition can alter what response is permissible there.

Computational turn management. Modern predictors use linguistic context or projected future voice activity to anticipate turn shifts and backchannels (Ekstedt & Skantze, 2020; 2022). Embodied agents have likewise combined timing, dialogue history, and task goals to resolve overlap (Gervits & Scheutz, 2018). These methods estimate when an action is possible or likely. RPB asks a complementary question: whether the same moment supports a different action for a different interlocutor relationship.

Full-duplex speech models and evaluation. End-to-end systems such as Moshi, Freeze-Omni, MiniCPM-o 4.5, and PersonaPlex make overlapping, low-latency speech interaction technically accessible (Défossez et al., 2024; Wang et al., 2024; Cui et al., 2026; Roy et al., 2026). Recent benchmarks measure turn-taking phenomena, explicit turn-taking instructions, and vision–speech interaction (Lin et al., 2025; Tang et al., 2026; Mazumdar et al., 2026). RPB holds the acoustic evidence fixed and changes a relationship condition that does not name the target action, thereby testing selective social conditioning rather than general duplex fluency or direct instruction following.

Social context. Computational studies show that linguistic choices correlate with social power and politeness (Prabhakaran et al., 2012; Danescu-Niculescu-Mizil et al., 2013). Those results motivate treating authority, familiarity, and responsibility as potentially decision-relevant, but they do not define a universal conversational norm. RPB therefore evaluates reviewed local permission profiles and avoids mapping demographic groups to fixed actions.

Paired behavioral evaluation. Behavioral test suites and contrast sets expose failures hidden by aggregate accuracy, while counterfactually augmented data isolates which changes should alter a

prediction (Ribeiro et al., 2020; Gardner et al., 2020; Kaushik et al., 2020). RPB applies the same minimal-change logic to a real-time speech action: preserve observable audio, alter one relationship description, and score both correct change and required invariance.

Detachable conditioning and provenance. FiLM supplies a compact affine mechanism for injecting an explicit condition, and adapter tuning shows how task-specific capacity can be added while keeping a backbone frozen (Perez et al., 2018; Houlsby et al., 2019). Our hot-plug policy module uses these ideas as engineering primitives, not as evidence of social understanding. Because the claim depends on narrowly constructed training and confirmation sets, we also follow dataset-documentation principles that require explicit provenance, intended use, and limitations (Gebru et al., 2021; Bender & Friedman, 2018).

3 Benchmark Construction

3.1 Episode and counterfactual pair

An episode contains decision-preceding audio, aligned transcript, speaker identities, the current floor holder, and an absolute decision cutoff. Audio after the cutoff is excluded. Two conditions reuse the same episode identity, audio, speakers, and cutoff, but describe different interlocutor relationships. The relationship record includes relation type, social distance, familiarity, power asymmetry, decision authority, and task responsibility. It never includes an action label or an instruction such as "interrupt now."

The frozen core contains 511 paired episodes: 79 strong relationship pairs, 244 boundary pairs, and 188 invariant controls. The main model table uses the 79 strong and 188 invariant pairs. A separately licensed SmoothConv extension contains 300 audited pairs and contributes 67 additional strong pairs, 114 boundary pairs, and 119 controls. It is not merged into the core model results.

Table 4 in the appendix tabulates the frozen core composition.

3.2 Action and Gold representation

The action space is $\{\texttt{WAIT}, \texttt{BACKCHANNEL}, \texttt{TAKE_TURN}, \texttt{INTERRUPT}\}$. For condition (e, r), Gold contains an acceptable set $A_{e,r}$, a preferred set $P_{e,r} \subseteq A_{e,r}$ when a reliable preference exists, and a probability vector $q_{e,r}$ over the four actions. Set Gold prevents legitimate conversational alternatives from being treated as errors; ranking and probability Gold retain consensus strength.

Two audio-capable model families broaden candidate coverage, surface disagreements, and estimate distributions; dataset authors then review every final condition and can override any model proposal. The released labels are human-reviewed Gold: model families propose, and human review ratifies. This provenance distinguishes author-reviewed labels from independent annotation. A train-only 96-item subset has a user-attested, model-assisted four-person review: model answers were visible and raw platform exports are unavailable. Its derived agreement ledger is not independently auditable reviewer-level evidence (Section 9). Benchmark-wide independent multi-rater IAA and a human ceiling are unavailable; they are outside the present study.

3.3 Leakage and release controls

Model input and Gold are physically separated. Inputs contain only the decision-preceding window, relationship description, identities, transcript, and cutoff. Post-decision behavior, acceptable/preferred sets, probabilities, and target actions are absent. Release manifests lock counts and SHA-256 hashes. The SmoothConv extension remains separate because it retains `CC-BY-NC-4.0` terms.

A preregistered transfer run froze 6,000 target-free predictions before Gold access. On this cross-corpus endpoint the primary semantic adapter reached DCA 0.0504 and utility 0.4706, close to the frozen base (0.0581 and 0.4728) with comparable invariant harm (0.3209 versus 0.3131), while the relation-blind control reached utility 0.6631. This preregistered check did not establish a cross-corpus benefit for the primary semantic adapter. The separate positive policy-contract result does

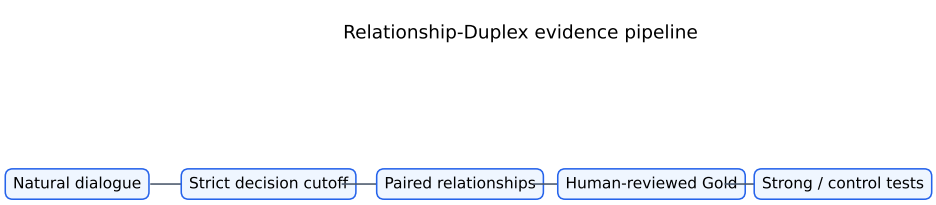


Figure 2: RPB evidence pipeline. Pair members share decision-preceding audio and differ only in the relationship condition; scoring combines local validity, paired direction, and invariant controls.

not overturn this negative transfer evidence; cross-corpus generalization remains unestablished (Section 8).

4 METRICS

4.1 CONDITION-LEVEL VALIDITY

For predicted action $\hat{a}_{e,r}$, Acceptable Hit and Preferred Hit are

$$\mathrm{AH}_{e,r} = \mathbf{1}[\hat{a}_{e,r} \in A_{e,r}], \tag{1}$$

$$\mathrm{PH}_{e,r} = \mathbf{1}[\hat{a}_{e,r} \in P_{e,r}]. \tag{2}$$

Probability mass is $q_{e,r}(\hat{a}_{e,r})$. These metrics answer whether one action is locally defensible; the paired metrics below add the orthogonal question of whether the relationship condition drove the choice.

4.2 PAIRED SENSITIVITY AND DIRECTION

For a strong pair with conditions r_0, r_1, correct sensitivity first requires both actions to be acceptable and different:

$$\mathrm{CS}_e = \mathbf{1}[\hat{a}_0 \in A_0]\mathbf{1}[\hat{a}_1 \in A_1]\mathbf{1}[\hat{a}_0 \neq \hat{a}_1]. \tag{3}$$

Directional Correct Adaptation is the strict subset whose change follows a Gold-supported boundary or preferred-set reversal:

$$\begin{aligned}\mathrm{DCA}_e = \mathrm{CS}_e\, \mathbf{1}[&(\hat{a}_0 \notin A_1) \vee (\hat{a}_1 \notin A_0) \\ &\vee ((\hat{a}_0 \in P_0) \wedge (\hat{a}_1 \in P_1) \wedge ((\hat{a}_0 \notin P_1) \vee (\hat{a}_1 \notin P_0)))].\end{aligned} \tag{4}$$

Thus DCA requires more than changing between two jointly allowed actions.

We also partition every strong pair into four exhaustive outcomes: both sides acceptable and changed; changed with at least one invalid side; unchanged with both sides acceptable; and unchanged with at least one invalid side. DCA is a strict subset of the first cell.

4.3 INVARIANT CONTROLS

On an invariant pair, any action change is oversensitivity. A change is harmful oversensitivity when at least one predicted side falls outside its acceptable set; a change with both sides acceptable is benign drift. This negative control prevents generic prompt reactivity from being mistaken for social adaptation.

4.4 STATISTICAL PROTOCOL

Intervals use an episode cluster bootstrap so the two conditions sharing one audio context are resampled together. Pair metrics are always computed within episode. Source and event-type strata are reported as error analysis rather than causal effects or architecture rankings. The SRPT diagnostics use source-held-out folds and preregistered within-source/recording permutations.

Table 1: Strong-set results (percent). Probability is mean Gold mass on the selected action; insensitive is the fraction of unchanged pairs.

Model	Accept.	Prefer.	Prob.	CS	DCA	Insens.
MiniCPM-o-4.5	75.3	31.6	30.8	11.4	3.8	79.7
PersonaPlex 7B	51.3	20.3	20.7	11.4	1.3	51.9
Freeze-Omni	31.6	19.0	15.4	3.8	0.0	89.9

5 Experimental Setup

We evaluate MiniCPM-o-4.5, PersonaPlex 7B, and Freeze-Omni through native audio interfaces. Every system receives the same decision-preceding content and relationship fields, then maps its native response into one of four actions. The model pool is intentionally heterogeneous: each system is evaluated as a behavioral black box under one common protocol, which is precisely the setting in which a relationship-blind shortcut would otherwise pass undetected. The protocol reports behavior under identical inputs; attributing differences to specific architectural choices is a separate controlled study enabled by, but distinct from, this benchmark.

The main PersonaPlex run uses an episode-derived seed shared across pair sides. Two additional complete seeds reuse the same frozen 158 strong conditions, Gold, and scorer. MiniCPM-o and Freeze-Omni are deterministic or near- deterministic under their frozen protocols. All three systems complete all strong and invariant-control conditions with zero failed records.

Four constant-action policies expose label-frequency shortcuts. Because they never read the relationship, their DCA is zero by construction, which is exactly the diagnostic point: their Acceptable and Preferred scores stay high, so any metric that rewards them alone would reward relationship-blind behavior.

6 Results

6.1 Local reasonableness does not imply selective adaptation

Table 1 reports the frozen 79-pair strong set. MiniCPM-o obtains the highest local scores, yet only 3 of 79 pairs satisfy DCA. PersonaPlex changes more often but frequently makes at least one side invalid. Freeze-Omni is nearly relationship-invariant and heavily favors `INTERRUPT`. Across all three, the best DCA is 3.8%.

6.2 Constant policies expose the accuracy shortcut

The constant `WAIT` baseline reaches 88.0% Acceptable and 45.6% Preferred, exceeding every tested model on Acceptable while using no relationship information. Constant `TAKE_TURN` reaches 49.4% Preferred. These baselines make the central confound visible: a high local score can be an action prior rather than selective conditioning.

The full constant-policy results appear in Table 5 in the appendix.

6.3 Failure decomposition

Table 6 explains why sensitivity alone is insufficient. PersonaPlex changes on 38 pairs, but 29 changes include an invalid side. MiniCPM-o remains unchanged on 63 pairs, and Freeze-Omni remains unchanged on 71. Each system thus fails selectivity for a distinct, directly observable reason: invalid reactivity, near-invariance with high local fit, and near-invariance with a dominant action prior.

6.4 Invariant controls complete the selectivity test

The 188 invariant pairs ask whether a system stays stable when relationship wording should not alter the policy. MiniCPM-o and Freeze-Omni change infrequently, but they also have very low

Table 2: Invariant-control results (percent).

Model	Accept.	Prefer.	Prob.	Any over.	Harmful over.
MiniCPM-o-4.5	81.9	38.6	38.3	13.8	6.9
PersonaPlex 7B	54.3	35.1	30.1	37.2	32.4
Freeze-Omni	41.2	19.4	20.2	5.9	5.3

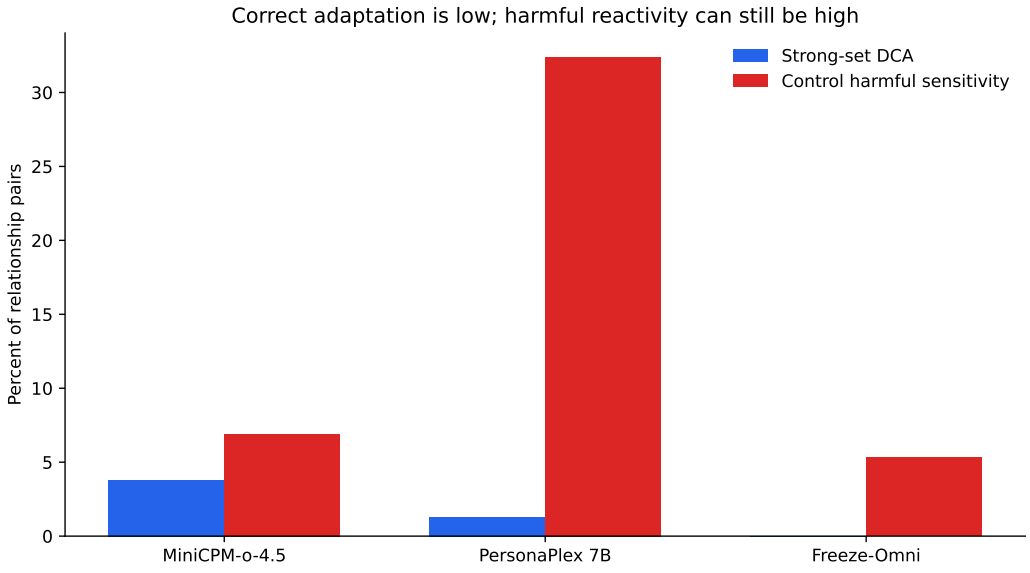


Figure 3: Strong-set DCA and harmful invariant-control sensitivity, shown as grouped bars. A selective policy should have high DCA and low harmful sensitivity.

strong-set DCA. PersonaPlex reacts more strongly to relationship text and reaches 32.4% harmful oversensitivity. Prompt sensitivity therefore cannot substitute for correct direction.

6.5 Sampling changes action marginals, not the diagnosis

Across three complete PersonaPlex seeds, Acceptable ranges from 50.6% to 53.2%, Preferred from 20.3% to 22.8%, and probability mass from 20.7% to 23.2%. Action marginals vary, including 58–68 `INTERRUPT` and 49–60 `WAIT` predictions. DCA nevertheless stays between 1.3% and 2.5%, only 1–2 of 79 pairs. The low-DCA diagnosis is stable across these three runs; this does not establish invariance across other seeds, models, or datasets.

7 From Diagnostic Lessons to a Sealed Improvement

7.1 Three design principles distilled from the training lineage

The development lineage yields three transferable design principles, each grounded in a measured failure mode. First, *selective transfer must be tested separately from local fit.* Small action heads can fit training distributions and substantially raise condition-level Acceptable or exact accuracy while retaining zero recording-held-out DCA; a recording-isolated paired metric is therefore the correct gate, because even perfect all-train accuracy leaves the relationship–boundary link unproven.

Second, *direction and safety must be learned jointly.* Relationship-only signals can recover a nonzero directional change while over-reacting on invariant pairs, while the tested conservative selectors eliminated invalid changes by suppressing all change into zero DCA and zero coverage. The policy-contract result combines directional adaptation with stability between invariant pair sides.

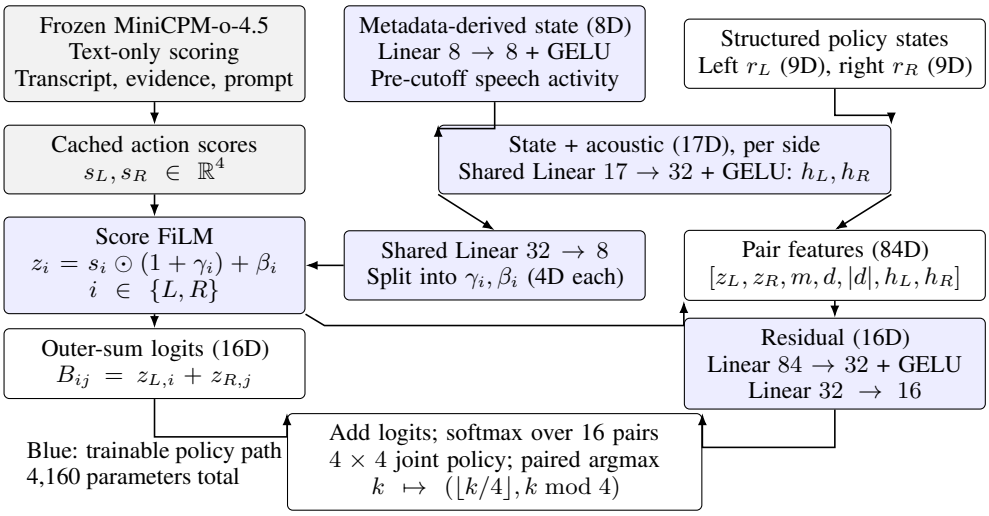


Figure 4: Implemented score-level hot-plug policy module. The frozen backbone scores transcript, selected evidence, and policy prompt with no waveform input. s_L, s_R are cached four-action scores, not hidden-layer activations. The 8D input comprises metadata-derived pre-cutoff speech-activity features with a transcript-based fallback; its embedding is shared across sides. $m = (z_L + z_R)/2$; $d = z_R - z_L$. Blue blocks share 4,160 trainable parameters. The relation-blind control zeros both 9D states; detach restores the base. The acoustic-null audit is unchanged ($p = 1.0$), not evidence of acoustic grounding.

This is not preservation of the base model's actions; behavior on unsupported pairs is not established by these metrics.

Third, *the supervision and evaluation contracts define the claim*. Because repeated model judgments can be operationally complete yet directionally collapsed, and output-constrained scorers can be executable without carrying a transferable permission delta, the successful design balances directional and invariant policy views before training, isolates participants across evaluation, and freezes all target-free predictions before the sealed target ledger is loaded.

Table 7 in the appendix summarizes these lessons.

7.2 Frozen-backbone relationship hot-plug adapter

The resulting method trains only a detachable policy module on top of a frozen MiniCPM-o-4.5 backbone. The module contains a 32-dimensional relationship FiLM router, an 8-dimensional acoustic-state encoder, and a paired 16-action residual head. Attach/detach replay restores the exact base identity. An equal-capacity relation-blind control uses the same architecture and optimization but receives a fixed null relationship vector at the policy head; the cached prompt-conditioned text scores are shared, not relationship-blinded.

Figure 4 shows the implemented data flow.

Appendix A retains the implementation size and policy-only latency measurements.

Training uses 65 Train-only source pairs expanded through 16 fixed policy patterns, producing 1,040 views: 520 directional and 520 invariant. Each of the four actions appears 260 times on each side. The supervision is a project-accepted, policy-generated contract, not independently ratified human policy. The four relation states deterministically map to the four actions; they encode authority, familiarity, urgency, and floor state, not relationship alone. Training and confirmation use the same state/action mapping. Freezing predictions before target-file access therefore does not remove the policy information already encoded in the supplied states. This result does not establish audio-conditioned adaptation or generalization to unseen permission rules.

The confirmation endpoint is frozen from nine participant-isolated AMI Train meetings (Carletta et al., 2006), with four pre-cutoff windows per meeting. Its 36-row ledger contains 16 directional and 20 invariant cases, and each action occurs nine times on each side. Thus an ideal selective policy has coverage $16/36 = 0.444444$, inside the preregistered $[0.20, 0.70]$ interval. Three relationship-conditioned seeds and three matched-control seeds are trained for 200 steps. Base and all six adapter groups then produce $7 \times 36 = 252$ target-free predictions, which are atomically frozen and SHA-verified before the confirmation target file is loaded. The input states are constructed from the fixed policy patterns, rather than independently annotated from the recordings.

7.3 SEALED RESULT

Table 3: Sealed AMI Train-only policy-contract result with a shared fixed state/action mapping. The relationship row aggregates three identical-metric seeds.

System	Acceptable	Preferred	Exact	DCA	Utility	Coverage	Harm	Invalid
Frozen MiniCPM-o-4.5 base	0.194444	0.194444	0.194444	0	0.638889	0.083333	0	0.083333
Relation-blind matched control	0.291667	0.291667	0.291667	0	0.666667	0	0	0
Relationship hot-plug adapter	**1**	**1**	**1**	**1**	**1**	**0.444444**	**0**	**0**

Every relationship-conditioned seed obtains DCA 1, balanced safe utility 1, coverage 0.444444, exact accuracy 1, and zero invalid change or invariant harm. The frozen base and relation-blind control both have DCA 0. Seed-9105, 20,000-draw, nine-meeting recording-blocked comparisons against both baselines give $p = 0.0020999$. All preregistered efficacy, coverage, exactness, safety, seed-consistency, identity, and target-access gates pass.

The confirmation contract assigns one policy action per condition, so its singleton Acceptable and Preferred sets coincide with condition Exact. We report all three columns explicitly to keep the metric definition transparent and consistent with the set-valued metrics used elsewhere in the paper.

The positive result establishes use of supplied structured state under this fixed policy contract, relative to frozen text scoring and a matched control. The state already identifies the prescribed action, so recording isolation and blocked significance do not establish learning beyond that mapping. This is distinct from correcting the native-audio benchmark failure, independently ratified human policy, public-test transfer, or online interaction.

7.4 EXPLORATORY MECHANISM STRESS TEST

After the sealed result was complete, the frozen adapters were evaluated under five deterministic input interventions without retraining or backbone forward calls. All intervention predictions were again frozen before target join. Cross-recording relationship mismatch reduces DCA from 1 to 0.125 and utility from 1 to 0.428704. Relationship nulling reduces them to 0 and 0.666667, while left/right relationship-state swapping reduces them to 0 and 0.518519. Recording-blocked comparisons of the correct input against each of these relationship interventions give $p = 0.0019999$ for DCA and utility.

Acoustic-feature nulling leaves every seed and metric unchanged: DCA and utility remain 1 and the blocked comparison gives $p = 1.0$. These interventions support dependence on supplied state and side binding but provide no evidence of an acoustic-feature contribution on this endpoint. This is a post-hoc exploratory analysis using revealed confirmation targets, not a second confirmatory result or a general causal claim about audio-conditioned behavior.

8 DISCUSSION

What the benchmark measures. The benchmark isolates one capability: whether a duplex policy changes its action in the reviewed direction at a fixed acoustic decision point, and holds steady when the relationship should not change the permission profile. It is a behavioral selectivity test, not relationship classification or politeness scoring. Multiple acceptable actions represent legitimate conversational variation, and DCA identifies whether paired behavior follows the condition rather than merely landing somewhere legal.

Why ordinary accuracy is insufficient. A constant `WAIT` policy outperforms native systems on condition-level Acceptable while scoring DCA 0, and reactive systems change often but land on an invalid action. Selective adaptation is therefore a conjunction that only the paired metrics capture: local validity, correct directional change, nontrivial but bounded coverage, and invariant safety. The policy-contract result satisfies these metrics on its own constructed targets; it does not resolve their failure on the native-audio benchmark.

What the positive result establishes. The matched control holds policy-module capacity and training setup fixed but zeros the explicit structured-state input to the policy head. Its DCA remains 0 while the conditioned adapter passes the frozen gates. This comparison and the post-hoc interventions support use of the supplied policy code and its side assignment. They do not distinguish relationship reasoning from learning the fixed lookup, nor establish an acoustic contribution or preservation of the base policy.

Limitations and future directions. Independent model-blind permission judgments, a human ceiling, perceptual validation of the action extractor, and transfer to new corpora and permission rules remain open questions. They would require additional evidence, not a reinterpretation of the present fixed-contract result. No additional experiments or human annotation are included in this revision. These are possible future directions, not promised validation for this submission. The supplementary lineage retains the completed experiments, including failed gates and access limits.

9 Scope of Validity and Ethics

We state the scope of validity precisely so that every claim can be reused without over-reading. *Construct.* Relationship descriptions are deliberately compact local-role summaries, and counterfactual prompts hold audio fixed while varying that role, which yields a controlled but stylized audio–relationship combination. *Coverage.* The strong core is dominated by EDACC and by particular preferred transitions; the aggregate diagnosis does not establish fine-grained rankings across relationship attributes or event types. *Annotation provenance.* Gold is human-reviewed and release-audited, with model families proposing and human review ratifying; a train-only review subset adds a user-attested, model-assisted four-person agreement summary. The reviewers saw model answers; the deterministic ledger reflects attested acceptance, not retained raw platform exports or model-blind ratings. It does not supply benchmark-wide raw-export evidence or independent multi-rater IAA. A separate model-first exercise—a frozen equal-weight panel of two Qwen Omni and two Gemini variants—proposed set-valued labels and is preserved purely as a descriptive audit in the supplement; it is independent of the Gold and of adapter selection. *Protocol.* The three systems expose different native interfaces and are evaluated as a common behavioral protocol, which is the design that exposes relationship-blind shortcuts across heterogeneous architectures.

10 Conclusion

RPB exposes a duplex-evaluation blind spot: selecting an action that is often reasonable does not, by itself, show that a system used the interlocutor relationship, and a constant policy can dominate ordinary accuracy while real systems either stay fixed, change in an invalid direction, or react harmfully when the relationship should not matter. Paired direction and invariant controls turn those failures into measurable quantities. The resulting contribution is an auditable benchmark specification, a shortcut diagnosis, and a bounded policy-contract demonstration on AMI-derived inputs. The frozen-backbone relationship hot-plug adapter reaches DCA and utility 1.0, coverage 0.4444, and zero harm or invalid change for every seed, while the base model and the equal-capacity relation-blind control both hold DCA 0, with both recording-blocked comparisons at $p = 0.00210$. This shows use of explicit state under the shared deterministic state/action mapping, not learned human permission boundaries or audio-conditioned generalization. The post-hoc stress audit supports state and side dependence, with acoustic-feature nulling unchanged ($p = 1.0$). Independent human validation, perceptual grounding and cross-corpus transfer remain outside the evidence provided by this study.

ETHICS STATEMENT

Ethically, relationship fields describe local interaction roles rather than stable personal traits, and must not be used for individual profiling or to bind demographic groups to normative actions. The dataset card documents intended use, and the separately distributed SmoothConv extension preserves its non-commercial license and challenge path.

REPRODUCIBILITY STATEMENT

The submission bundle supports inspection of the specification, selected frozen results, figure data, and included integrity manifests. It does not include the core dataset, the three native adapter sources, or the runtime-frozen scorer source; therefore, exact model-evaluation rerun from this bundle alone is unavailable. The tracked SmoothConv extension does not substitute for the core. The supplementary `srpt_diagnostic_lineage.md` records result and access boundaries; unavailable inputs are not represented as publicly released. Repository-only tests are not included in the reviewer bundle. This manuscript introduces no new Gold, dev evaluation, MaaS call, full-parameter backbone update, policy deployment, or test opening.

AI USE STATEMENT

AI tools were used for manuscript polishing, experimental design, and results analysis.

REFERENCES

Janet B. Bavelas, Linda Coates, and Trudy Johnson. Listeners as co-narrators. *Journal of Personality and Social Psychology*, 79(6):941–952, 2000. doi: 10.1037/0022-3514.79.6.941.

Emily M. Bender and Batya Friedman. Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics*, 6:587–604, 2018. doi: 10.1162/tacl_a_00041.

Jean Carletta, Simone Ashby, Sebastien Bourban, Mike Flynn, Mael Guillemot, Thomas Hain, Jaroslav Kadlec, Vasilis Karaiskos, Wessel Kraaij, Melissa Kronenthal, et al. The ami meeting corpus: A pre-announcement. In *Machine Learning for Multimodal Interaction*, pp. 28–39. Springer, 2006. doi: 10.1007/11677482_3.

Junbo Cui, Bokai Xu, Chongyi Wang, Tianyu Yu, Weiyue Sun, Yingjing Xu, Tianran Wang, Zhihui He, Wenshuo Ma, Tianchi Cai, et al. Minicpm-o 4.5: Towards real-time full-duplex omni-modal interaction, 2026.

Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. A computational approach to politeness with application to social factors. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pp. 250–259, 2013.

Alexandre Défossez, Laurent Mazaré, Manu Orsini, Amélie Royer, Patrick Pérez, Hervé Jégou, Edouard Grave, and Neil Zeghidour. Moshi: A speech-text foundation model for real-time dialogue, 2024.

Starkey Duncan. Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, 23(2):283–292, 1972. doi: 10.1037/h0033031.

Erik Ekstedt and Gabriel Skantze. Turngpt: A transformer-based language model for predicting turn-taking in spoken dialog. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pp. 2981–2990, 2020. doi: 10.18653/v1/2020.findings-emnlp.268.

Erik Ekstedt and Gabriel Skantze. Voice activity projection: Self-supervised learning of turn-taking events. In *Interspeech*, 2022.

Matt Gardner, Yoav Artzi, Victoria Basmov, Jonathan Berant, Ben Bogin, Sihao Chen, Pradeep Dasigi, Dheeru Dua, Yanai Elazar, Ananth Gottumukkala, et al. Evaluating models' local decision boundaries via contrast sets. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pp. 1307–1323, 2020. doi: 10.18653/v1/2020.findings-emnlp.117.

Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. Datasheets for datasets. *Communications of the ACM*, 64 (12):86–92, 2021. doi: 10.1145/3458723.

Felix Gervits and Matthias Scheutz. Pardon the interruption: Managing turn-taking through overlap resolution in embodied artificial agents. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, pp. 99–109, 2018. doi: 10.18653/v1/W18-5011.

Mattias Heldner and Jens Edlund. Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38(4):555–568, 2010. doi: 10.1016/j.wocn.2010.08.002.

Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin de Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. Parameter-efficient transfer learning for nlp. In *Proceedings of the 36th International Conference on Machine Learning*, pp. 2790–2799, 2019.

Divyansh Kaushik, Eduard Hovy, and Zachary C. Lipton. Learning the difference that makes a difference with counterfactually-augmented data. In *International Conference on Learning Representations*, 2020.

Stephen C. Levinson and Francisco Torreira. Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, 6:731, 2015. doi: 10.3389/fpsyg.2015.00731.

Guan-Ting Lin, Jiachen Lian, Tingle Li, Qirui Wang, Gopala Anumanchipalli, Alexander H. Liu, and Hung-yi Lee. Full-duplex-bench: A benchmark to evaluate full-duplex spoken dialogue models on turn-taking capabilities, 2025.

Abhirup Mazumdar et al. Videofdb: Evaluating full-duplex vision-speech capabilities in conversational agents, 2026. Preprint.

Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. Film: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018. doi: 10.1609/aaai.v32i1.11671.

Vinodkumar Prabhakaran, Owen Rambow, and Mona Diab. Annotations for power relations on email threads. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, 2012.

Marco Tulio Ribeiro, Tongshuang Wu, Carlos Guestrin, and Sameer Singh. Beyond accuracy: Behavioral testing of nlp models with checklist. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 4902–4912, 2020. doi: 10.18653/v1/2020.acl-main.442.

Rajarshi Roy, Jonathan Raiman, Sang-gil Lee, Teodor-Dumitru Ene, Robert Kirby, Sungwon Kim, Jaehyeon Kim, and Bryan Catanzaro. Personaplex: Voice and role control for full duplex conversational speech models, 2026.

Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696–735, 1974. doi: 10.2307/412243.

Emanuel A. Schegloff. Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, 29(1):1–63, 2000.

Tanya Stivers, N. J. Enfield, Penelope Brown, Christina Englert, Makoto Hayashi, Trine Heinemann, Gertie Hoymann, Federico Rossano, Jan Peter de Ruiter, Kyung-Eun Yoon, and Stephen C. Levinson. Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, 106(26):10587–10592, 2009. doi: 10.1073/pnas.0903616106.

Yunhao Tang et al. Instruct-fd: Can your full-duplex speech system follow turn-taking instructions?, 2026. Preprint.

Xiong Wang, Yangze Li, Chaoyou Fu, Yunhang Shen, Lei Xie, Ke Li, Xing Sun, and Long Ma. Freeze-omni: A smart and low latency speech-to-speech dialogue model with frozen llm, 2024.

Victor H. Yngve. On getting a word in edgewise. In *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*, pp. 567–578, 1970.

A Implementation Details

The complete policy module has 4,160 trainable float32 parameters in ten tensors (16,640 raw parameter bytes). Its immutable checkpoint is 20,229 bytes and the checkpoint, manifest, and SHA ledger together occupy 22,335 bytes. In a preregistered single-example policy-only microbenchmark with 2,000 warm-up and 10,000 timed forwards, median/p95 latency was 172.03/212.67 μs on an x86-64 CPU and 397.001/467.291 μs on an NVIDIA RTX PRO 5000 72GB Blackwell. These figures exclude frozen-backbone inference, audio preprocessing, file I/O, and model or adapter loading; they quantify only the incremental policy module and are not end-to-end conversational latency.

B Auxiliary Tables

Table 4: Frozen core composition. Each pair has two conditions.

Split	Pairs	Conditions	Role in evaluation
Strong	79	158	Correct change when policy differs
Boundary	244	488	Weaker permission-profile variation
Invariant control	188	376	Stability when policy should not change
Total	511	1,022	Frozen explicit-relationship release

Table 5: Constant-policy strong-set baselines (percent).

Policy	Acceptable	Preferred	DCA
Always `WAIT`	88.0	45.6	0.0
Always `BACKCHANNEL`	75.3	13.3	0.0
Always `TAKE_TURN`	62.7	49.4	0.0
Always `INTERRUPT`	0.6	0.0	0.0

Table 6: Exhaustive strong-pair outcomes: valid change, invalid change, valid no-change, invalid no-change. DCA counts are 3, 1, and 0 respectively.

Model	Valid change	Invalid change	Valid same	Invalid same
MiniCPM-o-4.5	9	7	38	25
PersonaPlex 7B	9	29	18	23
Freeze-Omni	3	5	16	55

Table 7: Transferable lessons retained from the training lineage. Full run-by-run evidence, failed gates, and artifact identities remain in the supplementary lineage.

Observed shortcut	Why it is insufficient	Required corrective contract
High local action fit	Can preserve zero paired directional adaptation	Recording-isolated DCA and exactness
Safe no-change collapse	Removes harm by eliminating useful changes	Joint direction, coverage, and invariant-safety gates
Complete model-generated scores	May contain little action diversity or permission contrast	Balanced policy supervision fixed before training