Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Quantifying Inside-Out acting: facial action units and silence duration as markers of interior performance

Shunsuke Tanaka
Kyoto Institute of Technology
Aiko Hirai
Kyoto University of the Arts
Shogo Kajimura
kajimura.shogo.1204@gmail.com

Kyoto Institute of Technology

Article

Keywords: acting, Inside-Out, Stanislavski, Facial Action Coding System, Action Units, performance analysis, affective computing, performing arts pedagogy

Posted Date: August 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9837550/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Stage acting distinguishes between “Inside-Out” (IO) approaches, in which performers generate genuine inner experience to drive expression, and techniques that shape behaviour without such inner work. Although this distinction is central to Stanislavski-derived pedagogy, its evaluation has relied on instructors’ subjective judgement, and the sole prior attempt to ground it in facial-expression measurement produced null results. Here we ask whether IO leaves quantitatively detectable traces on the performer’s face and temporal use of silence. Twelve female drama students performed the same Chekhov monologue under counter-balanced IO and control (noIO) conditions. Seventeen facial Action Units (AUs) were extracted frame-by-frame with OpenFace, speech and silence intervals were obtained with pyannote-audio, and performances were segmented into six script-driven units via change-point detection. Generalised linear mixed models, corrected with stringent and exploratory Benjamini-Hochberg criteria, revealed that (i) AU12 (lip-corner puller) and AU14 (dimpler) were consistently more probable under IO across segments, (ii) silence duration was significantly longer under IO whilst speaking time did not differ, and (iii) context-specific markers emerged in scenes of inner-state/text mismatch and in short repeated exclamations. The majority of segment-by-AU cells showed no condition difference. The findings constitute an existence proof that IO can be operationalised quantitatively; they do not yet establish a universal marker across performers, scripts, or genders. The alternative explanation that the effects reflect demand characteristics rather than inner work is discussed and tested where possible.

1. Introduction

Theatre is one of the few art forms in which the human body is both the instrument and the medium of expression (cf. Grotowski 1968; Kerber 2017). Across history, drama has articulated social questions, moral frameworks, and political dissent by inviting audiences to recognize themselves in the lives staged before them (Gulkhara and Farzaliyeva 2025). Its cultural force, however, rests on a prerequisite that practitioners call “truthfulness”: the audience’s sense that what the actor does on stage is not merely shown but lived (Morgan 2025). When that impression of authenticity is secured, spectators form emotional connections with characters; when it is not, even a technically correct performance can feel empty.

The question of how to reliably produce truthful acting motivated the Russian director Konstantin Stanislavski to develop, in the early twentieth century, the first systematic actor-training method. Stanislavski rejected a view of acting as surface imitation and insisted instead that the actor must “live” the role (Stanislavski 1936). To that end, he articulated concrete techniques, including the famous “emotional memory”: the deliberate recall of a real, lived feeling that can be transposed onto the role, so that the emotional display the audience perceives is not invented but retrieved. Because an audience is unusually sensitive to genuine affect, emotional memory is hypothesised to yield expressions that are harder to fake and more convincing to watch (Stanislavski 1936; Morgan 2025).

Techniques that prioritise the actor's interior life in this way are collectively called "Inside-Out" (IO) approaches (Morgan 2025). It is useful, before proceeding, to note that "IO" is not a single technique. It is a family that includes at least three internally contested traditions: Stanislavski's original "system", with its emphasis on *active analysis* and on emotional memory as only one tool among many (Stanislavski 1936); Strasberg's "Method", which elevated emotional memory to the central exercise and which is the variant most often associated with reports of psychological risk (Shirley 2018); and contemporary affective-recall exercises in conservatoire training, which vary substantially across schools (Palanimurugan 2024; Morgan 2025). What unites these traditions, and what distinguishes them from purely external approaches (vocal phrasing, posture, and facial choices built from the outside in) is the insistence that expression be *driven* by an internally generated affective state, rather than performed as a display. Our experimental contrast in this paper is between IO in this generic sense and a control condition in which participants are instructed not to do the inner work; it is *not* a contrast between IO and a rival trained technique such as Meisner's partner-focus (Meisner and Longwell 1987) or Suzuki's body-based training (Allain 2009), and we do not claim to adjudicate between IO traditions.

The distinction is pedagogically fundamental: most contemporary conservatoire programmes teach some form of IO as the gold standard for dramatic realism, while cautioning that sustained emotional immersion can impose a psychological burden on performers (Morgan 2025; Palanimurugan 2024; Shirley 2018). Whether a student is actually *doing* IO, or only mimicking its outward signs, is a question that today is answered almost exclusively by the subjective impression of senior instructors. The absence of an objective marker is problematic on two fronts. First, it makes training outcomes opaque and non-reproducible. Second, and more seriously, it means that a student who is mistakenly told they are "not really inside" may be pushed toward deeper immersion, precisely the regime in which psychological risk accumulates (Palanimurugan 2024).

Facial expression is a natural candidate for an objective marker of the inner state of a performer, while acknowledging at the outset that performance studies has long argued that truthfulness on stage involves more than the face: presence, breath, and spatial relationship to the audience all enter the concept (Kerber 2017). Our operationalisation is therefore deliberately reductive; it chooses measurability at the cost of coverage. Within that constraint, facial behaviour is tightly coupled to affective state and, in conveying emotion, carries information that often exceeds what is in the voice or the words (Song 2021). Prior work on acting has used facial cues to distinguish character portrayals: Berry and Brown (2022) showed that actors playing assertive versus agreeable characters differed reliably in the mobility of the jaw and the periocular region. The Facial Action Coding System (FACS) of Ekman and Friesen (1978), and its implementation as the set of Action Units (AUs), provide the anatomically grounded vocabulary for such analyses. Specific AUs map onto concrete muscular movements, e.g. AU06 (cheek raiser) and AU12 (lip-corner puller), which co-occur in genuine smiles of enjoyment (Gosselin et al. 1995).

The most direct precedent for the present study is Gosselin et al. (1995), who asked whether IO acting would produce AU patterns closer to those of spontaneous emotion than an external ("noIO") style. They

found no statistically significant differences and attributed the null result to their small sample (n = 6). We argue that the limitations of that study went beyond sample size. First, participants in Gosselin et al. performed *different* scripts, which confounds script content with individual differences. Second, the analysis averaged over sets of AUs rather than examining each AU separately, which can dilute effects that are confined to a single muscle group. Third, only silent intervals were analysed, to avoid contamination from articulation, thereby discarding any effect that IO might exert through the interaction of affect and speech. A properly controlled replication, analysing individual AUs and treating speech and silence as distinct phases, might therefore reveal effects that earlier work missed.

The present study adopts that strategy, at a deliberately tractable scale. We asked twelve female drama students to perform the *same* Chekhov monologue under an IO and a noIO condition, order counter-balanced. From the video we extracted seventeen AUs frame-by-frame with OpenFace 2.0, separated speech from silence with pyannote-audio 2.1, and used change-point detection to divide the performance into script-relevant segments rather than imposing arbitrary time bins. Our primary question was whether, under these stricter methodological conditions, IO and noIO performances would diverge in ways that are (a) stable across segments (and therefore candidates for a general within-script marker of IO), and (b) context-specific (and therefore informative about *how* IO reshapes expression in particular dramatic situations). A secondary question was whether IO would leave a signature not only on the face but on the temporal structure of the performance itself, specifically on the use of silence.

This is a demonstration study, not a generalisation study. The sample is small (n = 12), single-gender (women), single-script (Sonya's soliloquy from Act II of *Uncle Vanya*), and single-institution. Our aim is to establish that IO is in principle quantifiable from video in a well-controlled within-participant comparison, and to identify candidate markers worth pursuing in larger, multi-script, multi-gender work. We are explicit throughout the paper about the alternative interpretation that the observed effects reflect participants' compliance with the instruction labels rather than any inner psychological work, and § 4.8 is devoted to taking that alternative seriously.

Beyond its immediate empirical contribution, the study is intended as a step toward an evidence-based pedagogy of acting. If IO can be quantified from video without specialised instrumentation, instructors gain an external criterion that is independent of their own impression and that is, in principle, comparable across students and across years; students gain a form of feedback that does not require them to trust, or distrust, a teacher's reading of their inner life; and the field gains a basis on which to ask when pushing a performer toward greater immersion is genuinely warranted.

2. Methods

2.1 Participants

Twenty-one undergraduate students and staff (17 women, 4 men), aged 18–34, from the Department of Performing Arts, [institution anonymised for review], were initially recruited. All data collection was

conducted in the presence of a co-author [anonymised for review], professor of performing arts at the host institution, who confirmed the adequacy of each performance attempt.

Exclusions. the co-author watched each performance in real time and judged whether the IO condition had been authentically attempted. When the co-author judged an attempt inadequate, the participant was asked to perform the IO condition again; the retaken performance was recorded but the participant was excluded from the final analysis so that no performance contaminated by the instructor's corrective feedback would enter the dataset. By this criterion, five women and two men were excluded. Of the five excluded women, three had their video data fully processed through the OpenFace pipeline before the decision to exclude was finalised, which makes an intention-to-treat (ITT) re-analysis possible for them, reported in Supplementary Information. The remaining two excluded women required a mid-performance retake and their data were not processed further; their raw video is archived but no AU data are available. We note that this is a non-blind, expert-mediated exclusion on a variable correlated with the dependent measure. Because the instructor's judgment is itself part of what a quantitative marker is intended to replace, this constitutes a form of selection on the dependent variable and is discussed at length in § 4.8. Separately, two additional men (who were not flagged by the co-author) were excluded because their script differed from the women's (see § 2.2) and because n = 2 does not permit meaningful statistical inference. In total, nine of the twenty-one participants were excluded: five women and two men for IO inadequacy, and two men for script and sample-size reasons. The final confirmatory sample comprised twelve women; eleven were native Japanese speakers, one performed in English.

The study was approved by the Research Ethics Committee for Human Subjects Research of the corresponding author's institution (approval/receipt number 2024-93); the institution's name is withheld here for double-blind review and is stated in full in the accompanying cover letter. The research was conducted in Japan, and ethical approval was obtained locally, from an approval body located in Japan. All participants gave written informed consent for participation and for the use of their anonymised video data in scientific publication.

2.2 Script

IO practice requires the recall of personal affective memories and the sustained allocation of attentional resources. A novel script encountered at the time of the experiment would divert cognitive resources toward line memorisation and risk making IO impossible for participants who would otherwise be capable of it. We therefore used a monologue that forms part of the first-year performing arts curriculum at [institution anonymised for review] and is already internalised by all participants. This avoided a memorisation load and allowed the same script to be used across participants, holding content constant.

The monologue used for the female cast is Sonya's Act II soliloquy from Chekhov's *Uncle Vanya* (Chekhov 1897/2006), in which Sonya moves from a private reverie of happiness after a conversation with Astrov, whom she loves, to a despairing acknowledgement of her own plainness, culminating in the repeated self-accusation "plain, plain, plain." The source texts are Chekhov (2006; Jinzai translation, Act

II, pp. 223–224 of the cited edition) for the Japanese version and Chekhov (2002; Carson translation, Act II) for the English version. The English text was used by the one English-speaking participant and the Japanese text by the remaining eleven.

2.3 Experimental conditions

Each participant performed the same monologue twice, under two instructions:

- **IO.** “Draw on your own past experiences and the emotional memories associated with them, project those memories onto the role and the situation, and let the emotion arise from inside you.”
- **noIO.** “Do not engage in any inner emotional generation. Produce the performance using only external means: the movement of your facial muscles and the tone of your voice.”

To mitigate carry-over, all participants in a given session first completed one condition before anyone began the other, so that each performance was followed by a meaningful interval before its counterpart. The two orders (IO → noIO and noIO → IO) were counter-balanced across groups. Ample rehearsal time was provided, and recording began only when the participant indicated readiness.

2.4 Recording environment

Performances were recorded in a studio at the host institution [name anonymised for review], in Japan, using a single camera (Panasonic HC-VX992MS) placed directly in front of the performer at approximately eye level, at 3840 × 2160 resolution and 30 fps. The room was illuminated with standard indoor fluorescent lighting. Participants stood at a fixed distance from the camera, yielding a head-and-shoulders framing. Groups of approximately ten participants were recorded per session. OpenFace confidence scores were inspected for all processed videos; frames with confidence below 0.80 or with unsuccessful face detection were excluded from AU computation, and no participant had more than 5% of frames excluded on this basis.

2.5 Data extraction

Facial behaviour. Action Unit time series were extracted from each video with OpenFace 2.0 (Baltrušaitis et al., 2018; Baltrušaitis, 2018). OpenFace returns, per frame, an intensity score in [0, 5] for seventeen AUs (Table 1). Following the OpenFace documentation, intensities ≥ 1 were treated as “AU present” and intensities < 1 as “AU absent”, yielding a per-frame binary occurrence vector from which occurrence probabilities were computed per segment. We use occurrence probability rather than continuous intensity as the dependent measure because (i) FACS itself is a binary presence/absence coding system (Ekman and Friesen 1978) and occurrence probability is therefore the quantity with the clearest theoretical interpretation, and (ii) occurrence probability is the quantity addressed by the direct precedent (Gosselin et al. 1995). To guard against over-reliance on OpenFace’s nominal 1-unit threshold, a sensitivity analysis at alternative thresholds (0.5, 1.0, 1.5) is reported in § 3.6 for the two cross-segment markers (AU12, AU14). The seventeen AUs are listed in Table 1.

Table 1
The seventeen Action Units analysed in this study and the facial movements they describe.

AU	Label	AU	Label
AU01	Inner brow raiser	AU14	Dimpler
AU02	Outer brow raiser	AU15	Lip-corner depressor
AU04	Brow lowerer	AU17	Chin raiser
AU05	Upper lid raiser	AU20	Lip stretcher
AU06	Cheek raiser	AU23	Lip tightener
AU07	Lid tightener	AU25	Lips part
AU09	Nose wrinkler	AU26	Jaw drop
AU10	Upper lip raiser	AU45	Blink
AU12	Lip-corner puller	AU14	Dimpler

Speech versus silence. Audio was extracted from each video and processed with pyannote-audio 2.1 (Bredin 2023). This yielded, for each performance, the timestamps of speech and silence intervals; silences included the natural pauses of the monologue but also any hesitation or cognitive pause produced by the performer.

2.6 Pre-processing

Temporal normalisation. Because the duration of the performance varied across participants and conditions, each AU time series was resampled by linear interpolation to the length of the shortest recording in frames. Each AU was then standardised (*z*-scored) within performance to reduce individual differences in baseline facial mobility.

Segmentation. Holistic, whole-performance comparisons conflate distinct dramatic moments. We therefore partitioned each performance into script-aligned segments. For each participant and condition, structural change points in the multivariate AU time series were detected using the Python package ruptures (Truong 2020; Truong et al. 2020), with dynamic programming as the search algorithm (for global optimality) and a radial-basis-function kernel cost (to capture both mean and distributional change). Because the change points of individual performances did not align exactly, we used kernel density estimation (KDE) over all detected change points to identify the modal locations, and defined these KDE peaks as the *representative segment boundaries* of the script. Across choices of the expected number of change points, five dominant KDE peaks recurred, yielding a robust partition into six segments that was used uniformly for all subsequent analyses.

Guarding against circularity in segmentation. Because the representative boundaries are derived from the same AU time series that are subsequently tested for condition effects, the boundaries could in

principle be shaped by condition rather than by script structure. We checked this in two ways. First, we re-ran the KDE over change points *separately* for IO and noIO performances; the five dominant peaks coincided to within a small tolerance (§ 3.1), indicating that the boundaries reflect script-driven structure rather than IO-specific structure. Second, the boundaries of the six segments map onto recognisable dramatic beats of the monologue, which is the strongest qualitative evidence that what the change-point detector is tracking is the text itself.

2.7 Statistical analysis

For each AU in each segment, the probability of occurrence was modelled with a generalised linear mixed model (GLMM, logistic link) with acting condition (IO vs. noIO) as the fixed effect of interest. Two additional fixed effects were included: a binary indicator of performance errors (line-forgetting or stumbling; 1 if present) and a binary indicator of script language (1 for English). Modelling errors as a covariate, rather than excluding them, allowed us to retain all usable data while controlling for their effect on expression. Participant was included as a random intercept, permitting within-participant correlation across the two performances and across segments.

The same GLMM structure was used for the following analyses, presented here in the order in which they appear in the Results:

1. **Descriptive pooled analysis (§ 3.2).** A GLMM per AU per segment with condition as the only fixed effect of interest (collapsed over speech state). Reported as a descriptive summary of the overall condition effect.
2. **Duration analysis (§ 3.3).** A GLMM on speech and silence duration.
3. **Descriptive split analysis (§ 3.4).** The same model structure applied separately within speech and silence intervals. Reported as descriptive support for the primary interaction analysis.
4. **Primary analysis of speech × condition (§ 3.5).** A single GLMM per AU per segment including acting condition, speech state (Speech/Silence), and their interaction. This is the formal test of the speech-dependent modulation of expression and is the principal inferential model of the paper.

Although § 3.2 and § 3.4 are presented first to follow the order of the Results, readers should treat them as descriptive evidence and the § 3.5 interaction model as the inferential test, consistent with the principle that a single interaction test is the statistically appropriate way to ask whether IO differentially modulates speech and silence.

Effect sizes and uncertainty. All contrasts are reported with 95% confidence intervals (profile-likelihood where available, Wald otherwise) and converted to odds ratios. We summarise each contrast as an odds ratio with 95% CI and, for the cross-segment markers, a meta-analytic summary pooled across segments.

Multiple-comparison correction. Because the analysis crossed seventeen AUs with six segments, and in some analyses with two speech states, the number of tests was large (102 to 204 depending on

analysis). We used the Benjamini–Hochberg (BH) procedure to control the false discovery rate, and applied two correction schemes in parallel to balance robustness against exploratory sensitivity:

- **Stringent criterion.** BH applied jointly across *all* tests within the analysis. Effects surviving this criterion are interpreted as the confirmatory findings of the paper.
- **Exploratory criterion.** BH applied *within each AU* independently (across its segments and, where applicable, speech states). Effects surviving only this criterion are interpreted as *exploratory* and treated as hypotheses for future confirmatory replication, not as findings.

All analyses were run in R (version 4.5.1; R Core Team 2025) using glmmTMB (version 1.1.11; Brooks et al. 2017) for the primary beta-family models and lme4 (version 1.1.37; Bates et al. 2015) for auxiliary linear mixed models. Significance markers are reported as * $p < .05$, ** $p < .01$, *** $p < .001$ with respect to the relevant BH-corrected threshold.

3. Results

3.1 Representative segment boundaries

The KDE of change points across participants and conditions exhibited five dominant peaks that persisted across the pre-specified number of change points. Re-running the KDE separately within IO and within noIO recovered the same five peaks with only small shifts in location, indicating that the segmentation is driven by the script rather than by condition. These peaks were adopted as the representative segment boundaries, yielding six segments whose interpretation maps onto recognisable dramatic beats of the monologue: (S1) opening reverie about the beloved; (S2) private recollection of his voice; (S3) self-doubt about her speech; (S4) transition to the mirror-like recognition of her appearance; (S5) recollection of the overheard conversation; (S6) repeated “plain, plain, plain.” (Fig. 1)

3.2 Condition effect on AU occurrence (descriptive, pooled over speech state)

We first summarise the overall condition effect descriptively, pooling across speech state. Out of the 102 AU × segment cells tested, 96 showed no statistically detectable condition effect under the stringent BH criterion; the interpretation that follows concerns the six cells in which effects were found, concentrated in Segment 1 (AU01, AU12, AU23, AU45) and Segment 6 (AU14, AU15). The pattern of the four significant S1 effects indicates a sharply divergent opening: relative to noIO, IO showed lower AU01 and AU45 and higher AU12 and AU23. Under the exploratory criterion, AU12 was elevated under IO in five of the six segments, and in every case the direction of the effect favoured IO, providing the basis for treating AU12 as the paper’s candidate cross-segment marker (§ 4.2). Marginal probabilities and 95% CIs for AU12 and AU14 across the six segments are shown in Fig. 2 (see also Fig. 4A,B for the Segment 1 pattern).

3.3 Speech and silence duration

There was no significant condition effect on total speech duration (estimate = 0.008, SE = 0.009, p = .354), but total silence duration was significantly greater under IO (estimate = 0.547, SE = 0.049, $p < .001$). Thus IO did not make participants speak more, or more slowly in articulation; it made them *pause longer* (Fig. 3).

3.4 Condition effect within speech and silence intervals

Within speech intervals. Under the stringent criterion, four contrasts were significant, concentrated in S1 (AU06, AU12, AU45) and S4 (AU12). Under the exploratory criterion, AU12 was significantly elevated under IO in five segments, and AU06 (absent from the pooled analysis) was significantly elevated under IO in S1 and S2.

Within silence intervals. Under the stringent criterion, only one contrast reached significance: AU26 (jaw drop) was significantly *higher* under noIO in S3. Under the exploratory criterion, AU12 was significantly higher under IO in two segments.

The contrast between speech- and silence-interval findings suggests that several IO-specific markers (notably AU06 and AU12) are most strongly coupled to the act of speaking itself.

3.5 Primary analysis: interaction of acting condition and speech state

Main effects. Under the stringent criterion one main effect of condition survived: in S4, AU14 was significantly higher under IO. Under the exploratory criterion, AU14 was elevated under IO in five of six segments, identifying AU14 as a second candidate cross-segment marker of IO (Fig. 2B).

Interaction. No condition × speech-state interactions survived the stringent correction. Under the exploratory criterion, two interactions emerged, i.e., AU06 in S2 and AU07 in S6, and should be treated as hypotheses. Simple-effect decomposition of the S6 AU07 interaction showed that in noIO only, AU07 occurred more frequently during speech than during silence, whereas under IO the occurrence rate of AU07 did not differ between speech and silence. In other words, lid-tightening was intermittent in noIO, switched on when a word was spoken and off when the performer fell silent, whereas IO sustained it throughout the scene (see § 4.5 for interpretation) (Fig. 4C).

3.6 Sensitivity to the AU occurrence threshold

Across thresholds 0.5, 1.0 (the primary analysis), and 1.5, the direction of every reported IO > noIO difference for AU12 and AU14 was preserved (Supplementary Table S1). The qualitative pattern reported in § 3.2 / § 3.5 is therefore not an artefact of the choice of presence threshold.

4. Discussion

4.1 Summary of findings

We asked whether Inside-Out acting, a technique that is central to modern actor training but that has resisted quantification, can be distinguished from external ("noIO") performance on the basis of objective video-derived measures. Using matched performances of the same Chekhov monologue and individual-AU GLMMs separated by speech and silence, we identified three classes of marker. First, *cross-segment* markers robust to dramatic context: AU12 (lip-corner puller, the central component of a smile) and AU14 (dimpler, a small contraction at the corner of the mouth often associated with restrained or wry expression) were consistently more probable under IO, and silence duration was consistently longer under IO while speaking time did not change. Second, *context-specific* markers tied to particular dramatic moments: in the opening segment, where inner state and spoken content diverge, IO produced a distinctive combination of higher more lip-corner pulling and lip tightening (AU12, AU23) but less inner brow raising and less blinking (AU01, AU45) than under noIO, a pattern consistent with a performer who is privately savouring an emotion while outwardly composing a different face. Third, a *temporal-coupling* marker: in the segment of short repeated exclamations, AU07 under noIO was intermittent and tied to speech, whereas under IO it was sustained across silence. Together, these findings show that IO is not experientially private but behaviourally legible (Figs. 1–4).

4.2 AU12: hypotheses for the cross-segment effect

The most robust single marker was AU12 (lip-corner puller). We considered three candidate explanations for an AU12 increase under IO, and treat the discussion that follows as generating hypotheses for targeted follow-up work rather than adjudicating between them from the present data (Fig. 2A).

The first candidate is speaking time: AU12 accompanies articulation, and more speech could mechanically produce more AU12 (Shah et al. 2013). The second is affective congruence: AU12 is a marker of happiness (Gosselin et al. 1995), and IO might recruit positive affect from the performer's emotional memory. The third is a change in articulatory patterning: Newby et al. (2025) report that AU12 co-occurs with specific phonemes, so an IO-driven change in how the same script is phonated (in timing, in vowel colouring, in how consonants are released), could alter the frequency of AU12-linked articulatory gestures without any change in the words spoken.

Two of these candidates sit uncomfortably with the present data. Speaking time did not differ between conditions (§ 3.3), which would be expected if the mechanical-speech hypothesis were the whole story, though it does not rule out more subtle differences in articulation rate or phoneme-level distribution. And the script moves from happiness in the first half to despair in the second; if AU12 tracked the emotional content of the script, it should decline in the later segments, yet it remained elevated throughout in IO, which is hard to reconcile with a simple happiness-tracking account. We regard the articulatory hypothesis as the best-motivated remaining candidate, but we emphasise that the present study has no phoneme-level audio analysis and therefore cannot test it. A complementary possibility we cannot exclude is that IO recruits a facial morphology during emotional recall that OpenFace's AU12 regressor is especially sensitive to, which would inflate the estimate without any change in lip behaviour proper.

Discriminating between these hypotheses requires a forced-alignment phoneme-level analysis of the audio, a comparison of OpenFace AU12 against manually FACS-coded AU12, and ideally a replication with a script that does not contain strongly valenced text.

4.3 Concentration and the temporal structure of performance

A second robust marker, AU14 (dimpler), and the longer silences under IO, converge on a single interpretation: IO is a state of sustained concentration. AU14 has been reported to accompany effortful thought (Grafsgaard et al. 2013); its persistent elevation under IO is consistent with the moment-to-moment cognitive work of retrieving, holding, and projecting an emotional memory. Longer silences, in turn, are what that work requires. Speaking competes for attentional resources with the inner operations specific to IO (retrieval, maintenance, reconciliation with the script), and the performer appears to reclaim those resources in pauses. The result is an objective correlate of something that Stanislavski (1936) insisted on without being able to measure: “Living the role” is expensive, and in IO that expense is visible in the distribution of silence (Figs. 2B and 3).

4.4 Segment 1 and the subtext

Segment 1, a scene in which the woman speaks of her confusion about the man’s true feelings while inwardly basking in her affection for him, was the most discriminating of all the segments, and the pattern there is theoretically informative. Under IO, AU06 and AU12 (both associated with happiness) were significantly elevated during speech, as if the inner state had leaked through the delivery of words whose surface meaning runs the other way. Under noIO, by contrast, AU01 (inner brow raiser), which has been reported as a marker of confusion and frustration (Bosch et al. 2014; Lee et al. 2021), was significantly higher. In other words, the noIO performer’s face was responding to the *text*, whereas the IO performer’s face was responding to the *subtext*. The Stanislavski tradition attaches great importance to subtext (the unspoken inner utterance that a line carries; Stanislavski 1936), and our data give that tradition a directly measurable handle: under IO, subtext dominates the face; under noIO, the semantic content of the line does (Fig. 4A,B).

The same segment also showed, under IO, higher AU23 (lip tightener) and lower AU45 (blink). AU23 rises under psychological stress (Blasberg et al. 2023), and blink rate drops under high cognitive load (Cornelis et al. 2025). Both are consistent with a cost: sustaining a subtext that contradicts the line one is currently speaking is hard work. A training curriculum that values IO therefore needs to value the conditions under which that work can be done safely, a point we return to in § 4.7.

4.5 Segment 6 and the continuity of affect

Segment 6, in which the character spits out the word “plain” three times, places a different demand on the performer: there are no long phrases to hide behind. Here we found a condition × speech-state interaction for AU07 (lid tightener), an index of anger (Gosselin et al. 1995). In noIO, AU07 rose only during the utterance of “plain” and fell between utterances; in IO it remained present across silences. The

interpretation we propose is that noIO, lacking a sustaining inner state, attempts to meet the demand by varying the outward delivery ("three times differently"), producing a performance in which affective expression is welded to the act of speech and collapses between words. IO, having built the anger from emotional memory, does not need the word to keep the feeling alive. Taken together with § 4.4, this suggests that a general signature of IO is *continuity* of expressive state across the speech–silence boundary, whereas noIO expression is more strongly event-locked (Fig. 4C).

4.6 Contribution and methodological novelty

The substantive contribution of this study is a multi-layered, reproducible objective description of IO acting: a cross-segment signature (AU12, AU14, longer silences), a context-sensitive signature (subtext-driven facial pattern in scenes of inner–outer mismatch), and a temporal-coupling signature (continuity of affect across silence in repeated-word scenes). Where Gosselin et al. (1995) found no statistically significant IO/noIO difference with n = 6, a different script per participant and a pooled-AU analysis, we find robust and interpretable differences with n = 12, a common script, per-AU modelling, and an explicit treatment of speech state. Two methodological choices appear to matter: (i) examining AUs individually rather than averaging them, which preserves effects that are muscularly specific; and (ii) treating performance as a sequence of script-aligned segments identified by change-point detection rather than as a block, which makes it possible to separate cross-segment from context-specific effects. Both are transferable to other questions about acting and, more broadly, to any analysis of affective video in which the event of interest is structured in time.

4.7 Implications for performing-arts pedagogy and beyond

The practical motivation of this work was to give acting teachers an external criterion for IO that does not depend on their own impression and that does not force students into deeper immersion on trust alone. Our findings suggest three immediately usable handles. First, AU12 and AU14 elevation and longer silences are candidate *screening* markers; a student producing all three is likely engaging in IO even without a subjective report. Second, the segment-specific markers in scenes of inner–outer mismatch give instructors a *diagnostic*: a student who is not producing the "subtext signature" in such a scene can be coached on that scene specifically, rather than being told, globally, to "get more inside." Third, the elevated AU23 and suppressed AU45 under IO in subtext-heavy scenes are a *warning*: IO in these moments costs something, and teachers should be able to see the cost objectively rather than discovering it only when a student becomes distressed. A pedagogy that observes these signals can preserve the artistic value of IO while reducing the psychological risk that its unmonitored use entails (Palanimurugan 2024; Shirley 2018).

More broadly, an objective vocabulary for IO has applications beyond the theatre. "Truthful" behaviour during speech is a criterion of evaluation in customer-service training, in interview coaching, and in the assessment of therapeutic interaction; the analytical pipeline presented here (change-point-based segmentation, speech-resolved AU analysis, GLMM with participant random effects) is immediately portable to those domains. In the longer term, it provides a basis on which generative models of

expressive behaviour (e.g. digital actors, affective agents) can be evaluated against an IO-grounded ground truth rather than against surface mimicry alone.

4.8 Limitations and alternative interpretations

Demand characteristics. The most serious alternative to our interpretation is that the observed differences reflect participants' compliance with the instruction labels, not any inner psychological work. All twelve participants were students in an institution where IO is taught as the gold standard, and all knew that the co-author, their instructor, was present during recording. A cooperative student who is asked to do "noIO" after months of being taught that IO is the desirable approach has every incentive to signal that she is *not* doing IO (by muting expression, by reducing the very facial behaviours that IO is associated with), and an equal incentive, on switching to IO, to display the behaviours she believes the instructor will read as successful IO. Under this compliance account, AU12 and AU14 elevation under IO, and longer silences under IO, are not traces of emotional memory; they are traces of a performer trying to look like a person who is doing IO.

We cannot rule this account out from the present data, and we take seriously the possibility that it is partly or even wholly responsible for the findings. Three considerations, however, make us treat it as an incomplete explanation rather than a decisive one. First, the script-context findings in § 4.4 (in particular the subtext-aligned pattern of AU12, AU23, and suppressed AU45 specifically in Segment 1, where text and inner state diverge) would require a remarkably sophisticated compliance strategy: the participant would need to know in advance which specific facial behaviours, in which specific dramatic beat, are associated with IO, and to produce them in exactly those places. Simple "look inward" compliance does not predict a Segment-1-specific subtext signature. Second, the AU07 interaction in Segment 6 (§ 4.5) shows IO sustaining affect across speech and silence while noIO clips affect to spoken words, a pattern that again does not follow from "muting expression in noIO" in any straightforward way. Third, an intention-to-treat re-analysis (Supplementary Information) provides a direct empirical handle: AU14 survived the inclusion of excluded participants with direction preserved in all six segments, while AU12 was attenuated but preserved direction in five of six segments. If the effects were entirely driven by the instructor's selection, they should have collapsed under ITT; they did not, though the attenuation of AU12 means pre-selection cannot be ruled out as a contributing factor.

That said, the cleanest test of the compliance alternative is not within the present dataset. It requires (i) a blinded instructor who judges IO authenticity from audio only or from a masked video, (ii) a structured post-experiment debrief on how participants interpreted the "noIO" instruction and how effortfully they attempted each condition, and (iii) a replication in a population of participants who are *not* students of the instructor judging their performance. We did not administer a structured debrief in the present study, and we regard that as a methodological debt that future work must repay before the causal claim that IO *itself* produces the observed markers can be considered established.

Instructor-mediated exclusion. Three women with processed OpenFace data were excluded from the primary analysis because the co-author judged in real time that their IO was not authentically attempted.

An intention-to-treat re-analysis, reported in the Supplementary Information, addresses this directly. The results are informative but ambiguous. AU14 survived ITT with direction and approximate significance preserved in all six segments, which is evidence that it is not solely an artefact of selection. AU12 survived in five of six segments but was substantially attenuated and lost significance in several segments; moreover, the three excluded participants showed an AU12 pattern that was categorically *reversed* (noIO > IO in all six segments). This raises two possibilities. If the instructor's exclusion was valid, then the excluded participants genuinely did not do IO, their reversed AU12 is expected, and the ITT analysis over-dilutes the true effect. If the instructor's presence created a compliance effect, then the excluded participants are the ones who did not comply with the demand to "look like IO," and the ITT analysis is the more trustworthy estimate. We cannot distinguish these possibilities with the present design. A fully objective pipeline would replace this exclusion step with a blinded manual FACS spot-check or, in subsequent work, with a pre-specified computational criterion for "IO attempt." Supplementary Figure S1 visualises the AU12/AU14 pattern under this ITT sample.

"noIO" is a control, not a rival pedagogy. The noIO instruction ("use only external means") does not correspond to any trained acting technique. Meisner's partner-focus, Suzuki's body-based training, Laban, and Lecoq are all, in their own terms, "external" techniques but none of them would recognise itself in our noIO instruction. We therefore do not claim to have compared IO with an externally-driven pedagogy. noIO is an experimental control for "IO minus the inner work", and the comparison is of theoretical interest for what it reveals about IO, not as an evaluation of an alternative approach.

Generality across performers, scripts, and genders. The sample comprised twelve women, one script, one institution. The commitment to a single script was a deliberate trade-off against the script-confounding of Gosselin et al. (1995), but it means our design cannot separate "IO effects that generalise" from "IO × script interactions specific to *Uncle Vanya*." Some segment-specific findings (for example the noIO-higher AU26 in S3 during silence) have no direct precedent in the AU literature and may reflect features peculiar to this scene. Generalising to men, to experienced professional actors, and to other repertoire requires larger multi-script multi-gender replication.

Exploratory versus confirmatory status. The cross-segment AU12 effect survives the stringent BH correction in five of six segments and is the closest thing in this paper to a confirmatory finding. Every other segment-level effect we report survives only the exploratory (within-AU) correction and should be treated as a hypothesis for confirmatory replication, not as an established result. The paper's main contribution is therefore an *existence proof*, namely that IO acting can in principle be quantified from video, together with a prioritised list of markers to test in larger, preregistered follow-up work.

Facial AUs are not the whole of truthfulness. Finally, and most fundamentally, performance studies has long argued that truthful acting is a matter of presence, breath, and relationship as well as face (Kerber 2017). Our operationalisation chooses facial AUs because they are measurable; it is not, and should not be read as, a complete account of what makes acting truthful.

5. Conclusion

Using video and standard open-source tools, we have shown that a pedagogical distinction that has resisted objective description for a century, Inside-Out versus external acting, is quantitatively visible in the face and in the timing of the performance. A small set of Action Units, together with the duration of silence, yields a cross-segment signature of IO; finer analysis segmented by dramatic beat reveals context-specific signatures that align with the theatrical concept of subtext; and a speech-by-condition analysis shows that IO sustains affect across silence in a way that external performance does not. These results bring three pay-offs into view: a more objective and safer pedagogy of acting, a transferable toolkit for the quantitative study of “truthfulness” in spoken behaviour more generally, and a new empirical handle on the relationship between inner experience and its bodily expression. Broader samples, more scripts, and direct measurement of articulatory and cognitive load will be needed to turn these findings into a universal evaluative standard, but the pathway is now open.

Declarations

Acknowledgements

This work was supported by the Acquisition, Technology & Logistics Agency (ATLA), Japan, under the Innovative Science and Technology Initiative for Security (Security Technology Research Promotion Program, Type D) (grant number JPJ013268), and by an internal young-researcher support programme at the corresponding author’s institution [name anonymised for review]. The authors thank the participating drama students for their willingness to perform under experimental conditions, and the colleagues at the host institution who assisted with recording logistics.

Use of generative AI

During the preparation of this manuscript, the authors used a large-language-model writing assistant (Anthropic Claude, Opus 4.6 family, accessed April 2026) to draft and to reorganise the text for the structure and length expectations of the journal. The assistant was not used to design the study, to run statistical analyses, to generate or interpret data, or to produce citations without human verification. All claims, references, data interpretations, and scientific content were reviewed and verified by the authors, who accept full responsibility for the manuscript.

Author contributions

[Author 1]: Conceptualisation, Methodology, Software, Formal analysis, Investigation, Data curation, Visualisation, Writing – original draft.
[Author 2]: Conceptualisation, Investigation, Resources, Supervision (performing arts), Writing – review and editing.

[Author 3]: Conceptualisation, Methodology, Supervision, Project administration, Funding acquisition, Writing – review and editing.

Competing interests

The authors declare no competing interests.

Ethics declaration

This study was approved by the Research Ethics Committee for Human Subjects Research of the corresponding author's institution (approval/receipt number 2024-93); the institution's name is withheld here for double-blind review and is stated in full in the accompanying cover letter. The research was conducted in Japan. Ethical approval was obtained from an approval body located in Japan, the same country in which the research was conducted, so the study was approved within the local area. Ethical review was carried out by the corresponding author's home institution because the corresponding author was the principal investigator; the performances were recorded at a separate host institution in the same city, which provided the participants and facilities and did not require separate review. All research was performed in accordance with the Declaration of Helsinki and with the ethical guidelines of the approving institution. All participants provided written informed consent for participation and for the publication of anonymised data.

Data availability

The anonymised Action Unit time series, segment boundaries, and statistical analysis scripts that support the findings of this study will be made available via the Open Science Framework upon publication. Raw video is not shared because the consent obtained from participants does not extend to public release of recognisable facial video.

References

1. Allain P (2009) *The art of stillness: the theatre practice of Tadashi Suzuki*. Methuen Drama, London
2. Baltrušaitis T (2018) OpenFace. https://github.com/TadasBaltrusaitis/OpenFace. Accessed 31 Dec 2025
3. Baltrušaitis T, Zadeh A, Lim YC, Morency L-P (2018) OpenFace 2.0: facial behaviour analysis toolkit. In: *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pp 59–66. https://doi.org/10.1109/FG.2018.00019
4. Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48. https://doi.org/10.18637/jss.v067.i01

5. Berry M, Brown S (2022) The dynamic mask: facial correlates of character portrayal in professional actors. *Quarterly Journal of Experimental Psychology*, 75(5):936–953.
6. Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, Skaug HJ, Mächler M, Bolker BM (2017) glmmTMB balances speed and flexibility among packages for zero-inflated generalised linear mixed modeling. *The R Journal*, 9(2):378–400. https://doi.org/10.32614/RJ-2017-066
7. Blasberg JU, Gallistl M, Degering M, Baierlein F, Engert V (2023) You look stressed: a pilot study on facial action unit activity in the context of psychosocial stress. *Comprehensive Psychoneuroendocrinology*, 15:100187. https://doi.org/10.1016/j.cpnec.2023.100187
8. Bosch N, Chen Y, D'Mello S (2014) It's written on your face: detecting affective states from facial expressions while learning computer programming. In: *International Conference on Intelligent Tutoring Systems*. Springer, pp 39–44.
9. Bredin H (2023) pyannote.audio 2.1 speaker diarization pipeline: principle, benchmark, and recipe. In: *Proc. INTERSPEECH 2023*, pp 1983–1987.
10. Chekhov A (2006) *Kamome / Wānya Ojisan (trans: Jinzai K)* [*The Seagull and Uncle Vanya*]. Shinchōsha, Tokyo
11. Chekhov A (2002) *The Seagull and Uncle Vanya* (trans: Carson P). Penguin Classics, London
12. Cornelis X, Dirix N, Bogaerts L (2025) The timing of spontaneous eye blinks in text reading suggests cognitive role. *Scientific Reports*, 15:19849. https://doi.org/10.1038/s41598-025-04839-y
13. Ekman P, Friesen WV (1978) *Facial Action Coding System (FACS): a technique for the measurement of facial action*. Consulting Psychologists Press, Palo Alto, CA
14. Gosselin P, Kirouac G, Doré FY (1995) Components and recognition of facial expression in the communication of emotion by actors. *Journal of Personality and Social Psychology*, 68:83–96.
15. Grotowski J (1968) *Towards a Poor Theatre*. Odin Teatrets Forlag, Holstebro
16. Grafsgaard JF, Wiggins JB, Boyer KE, Wiebe EN, Lester JC (2013) Automatically recognizing facial indicators of frustration: a learning-centric analysis. In: *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pp 159–165. https://doi.org/10.1109/ACII.2013.33
17. Gulkhara A, Farzaliyeva E (2025) Theatre as a reflection of social change: how dramatic arts capture cultural shifts and historical transformations. *Acta Globalis Humanitatis et Linguarum*, 2:254–261. https://doi.org/10.69760/aghel.02500133
18. Kerber TJ (2017) *The vocabulary of acting: a study of the Stanislavski 'system' in modern practice*. Dissertation, University of Birmingham
19. Lee J, So H-J, Ha S, Kim E, Park K (2021) Unpacking academic emotions in asynchronous video-based learning: focusing on Korean learners' affective experiences. *The Asia-Pacific Education Researcher*, 30(3):247–261.
20. Meisner S, Longwell D (1987) *Sanford Meisner on acting*. Vintage Books, New York

21. Morgan L (2025) How do listeners perceive emotional genuineness in musical theatre performers, and how might this help to inform the pedagogy of acting through song? *Theatre, Dance and Performance Training*, pp 1–20. https://doi.org/10.1080/19443927.2025.2570811
22. Newby A, Bhatta A, Kirkland C, Arnold N, Thompson L (2025) The role of facial action units in investigating facial movements during speech. *Electronics*, 14:2066. https://doi.org/10.3390/electronics14102066
23. Palanimurugan P (2024) The art of method acting: a comprehensive review of theory and practice. *Iconic Research and Engineering Journals*, 7(8):156–165
24. R Core Team (2025) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/
25. Shah M, Cooper DG, Cao H, Gur RC, Nenkova A, Verma R (2013) Action unit models of facial expression of emotion in the presence of speech. In: *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pp 49–54. https://doi.org/10.1109/ACII.2013.15
26. Shirley D (2018) His dream of passion: reflections on the work of Lee Strasberg and his influence on British actor training: part two. *Stanislavski Studies*, 6(2):165–182. https://doi.org/10.1080/20567790.2018.1496555
27. Song Z (2021) Facial expression emotion recognition model integrating philosophy and machine learning theory. *Frontiers in Psychology*, 12:759485.
28. Stanislavski C (1936) *An Actor Prepares*. Theatre Arts, New York
29. Truong C (2020) *ruptures: change point detection in Python*. https://github.com/deepcharles/ruptures. Accessed 31 Dec 2025.
30. Truong C, Oudre L, Vayatis N (2020) Selective review of offline change point detection methods. *Signal Processing*, 167:107299. https://doi.org/10.1016/j.sigpro.2019.107299

Figures

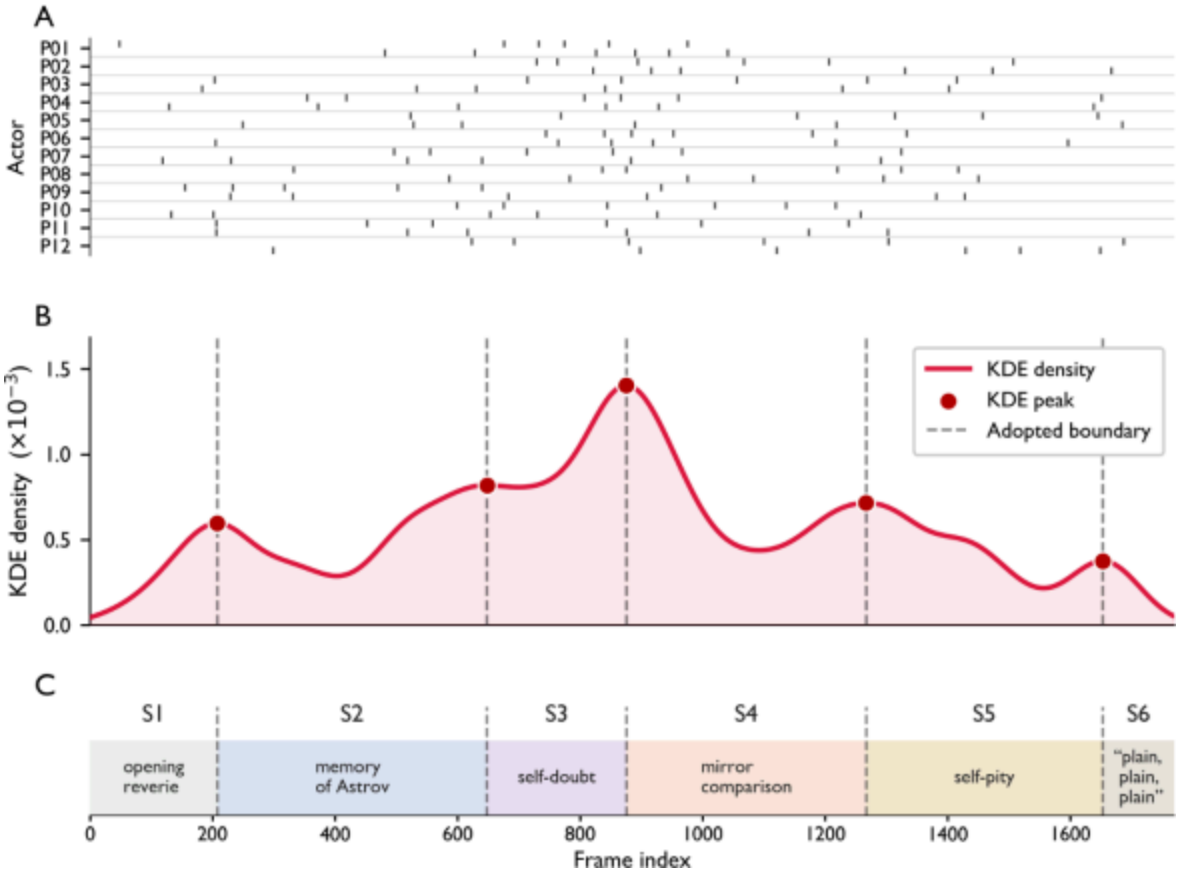


Figure 1

Five representative segment boundaries derived from the kernel-density estimate of per-recording change-points. (A) Per-recording change-point rug. Each row is one recording; the 12 actors (P01–P12) appear as adjacent pairs (IO row above, noIO below, separated by a thin horizontal divider). Each tick is one of the six change points detected by ruptures (Dynp + RBF kernel, n_bkps = 6) per recording, for 24 × 6 = 144 ticks total. (B) Pooled kernel-density estimate over all 144 change points (Gaussian kernel, bandwidth ≈ 53.1 = MIN_LEN / 100 × 3, matching the master-thesis setting). Five KDE peaks were identified by scipy.signal.find_peaks (filled red circles at frames 207, 648, 875, 1266, 1653) and adopted as the representative segment boundaries (208, 648, 876, 1267, 1653; black dashed verticals) within a 1-frame rounding tolerance, so peaks and boundaries coincide on the figure. (C) Resulting six-segment partition. The boundary structure passed a pre-specified circularity check: when the KDE was re-run separately for IO and noIO recordings the five adopted peaks coincided to within 22 frames (§3.1).

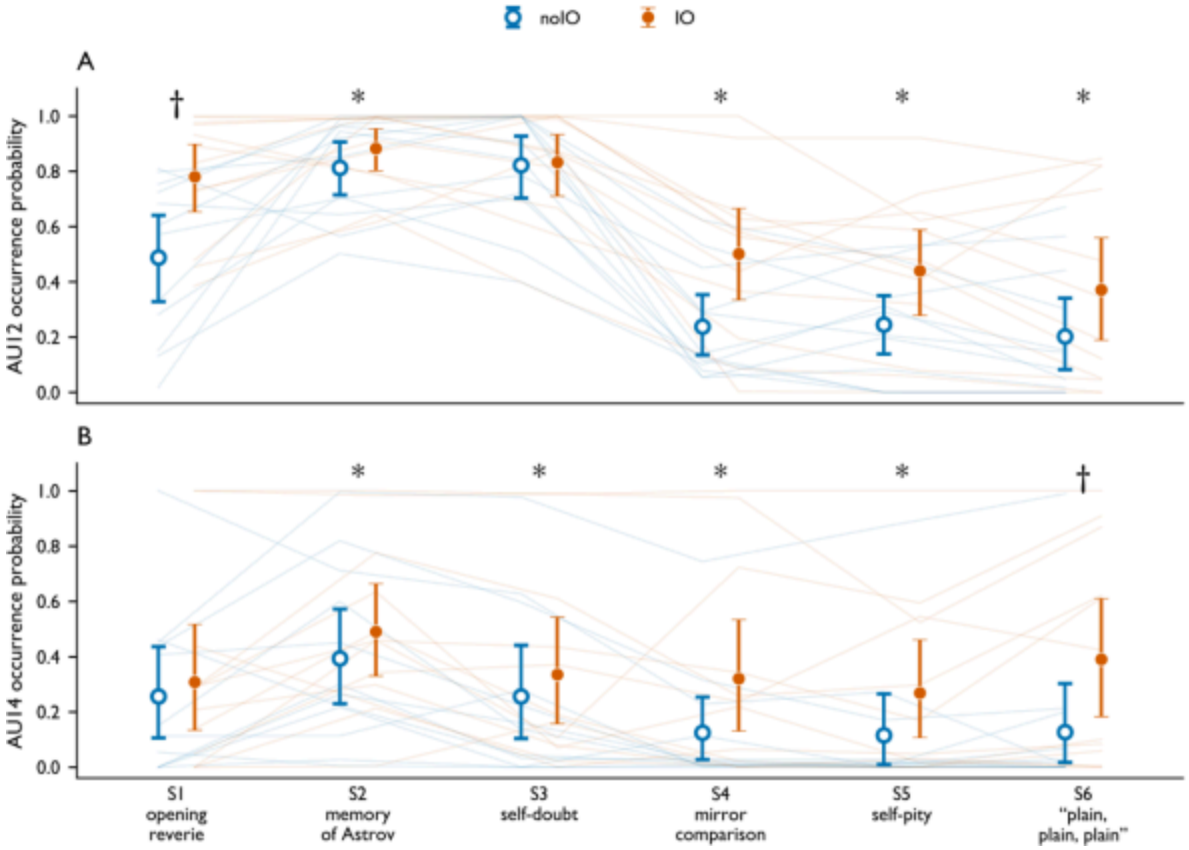


Figure 2

AU12 (lip-corner puller) and AU14 (dimpler), the two cross-segment markers that distinguish IO from noIO performances. Mean per-frame AU occurrence probability (proportion of frames with AU intensity ≥ 1) for each of the six monologue segments (S1 = opening reverie; S2 = memory of Astrov; S3 = self-doubt; S4 = mirror comparison; S5 = self-pity; S6 = "plain, plain, plain") for IO (filled orange circles) and noIO (open blue circles). Thin coloured lines join paired performances of each of the 12 retained actors. Vertical bars are 95 % percentile bootstrap confidence intervals over actors (5,000 resamples). (A) AU12. (B) AU14. Daggers (†) and asterisks (*) denote stringent (Benjamini–Hochberg q < .05) and exploratory (q < .10) significance from the manuscript's primary glmmTMB beta-family GLMMs (BH-corrected over the full 102 AU × segment cells; §3.2). Both AUs are elevated under IO in five of six segments under the exploratory criterion; AU12 reaches stringent significance in S1, AU14 in S6.

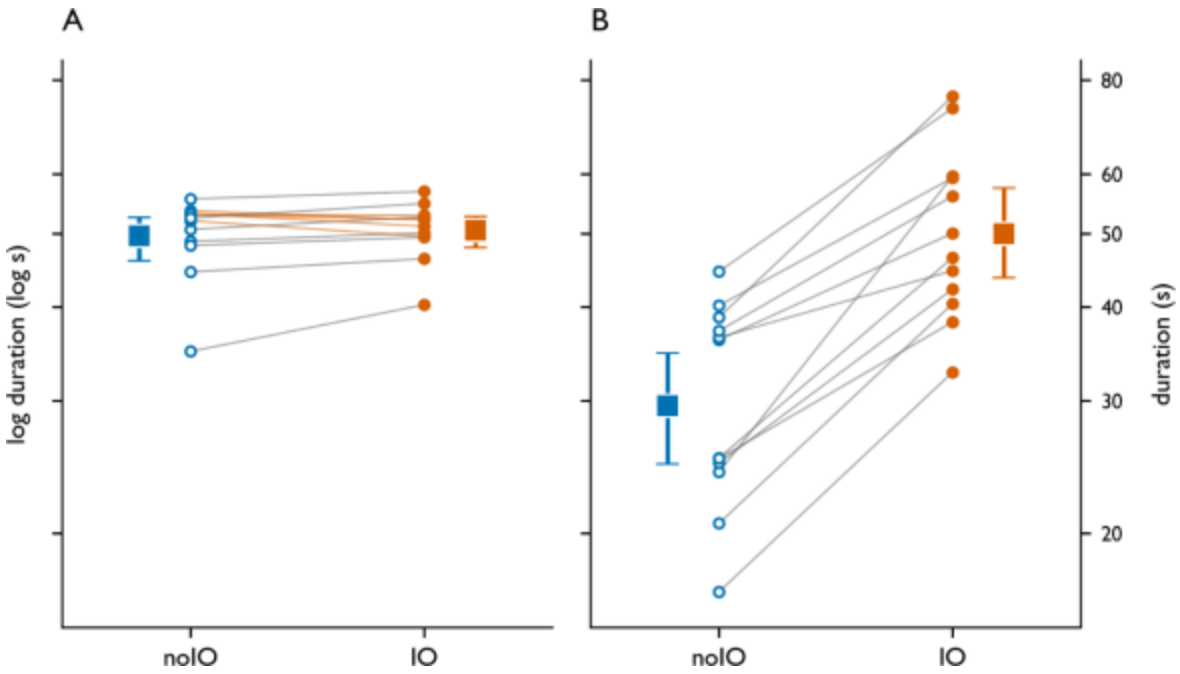


Figure 3

Inside-Out performances do not change total speaking time but extend total silence time by roughly 70 %. Total speech (A) and silence (B) duration per actor under the two conditions (n = 12). Each thin line connects the same actor's two performances: grey when IO duration exceeds noIO duration, orange when it is shorter. Open circles = noIO, filled circles = IO. Square markers offset to the left (noIO) and right (IO) are group means with 95 % percentile bootstrap CIs over actors. The y-axis is in log seconds (left); the right axis of panel B re-maps it to linear seconds. §3.3 reports a null effect on speech duration ($\beta = 0.008$, SE = 0.009, $p = .354$) and a robust positive effect on silence duration ($\beta = 0.547$, SE = 0.049, $p < .001$) from lme4 lmer on log seconds. Twelve of twelve actors increased their total silence under IO.

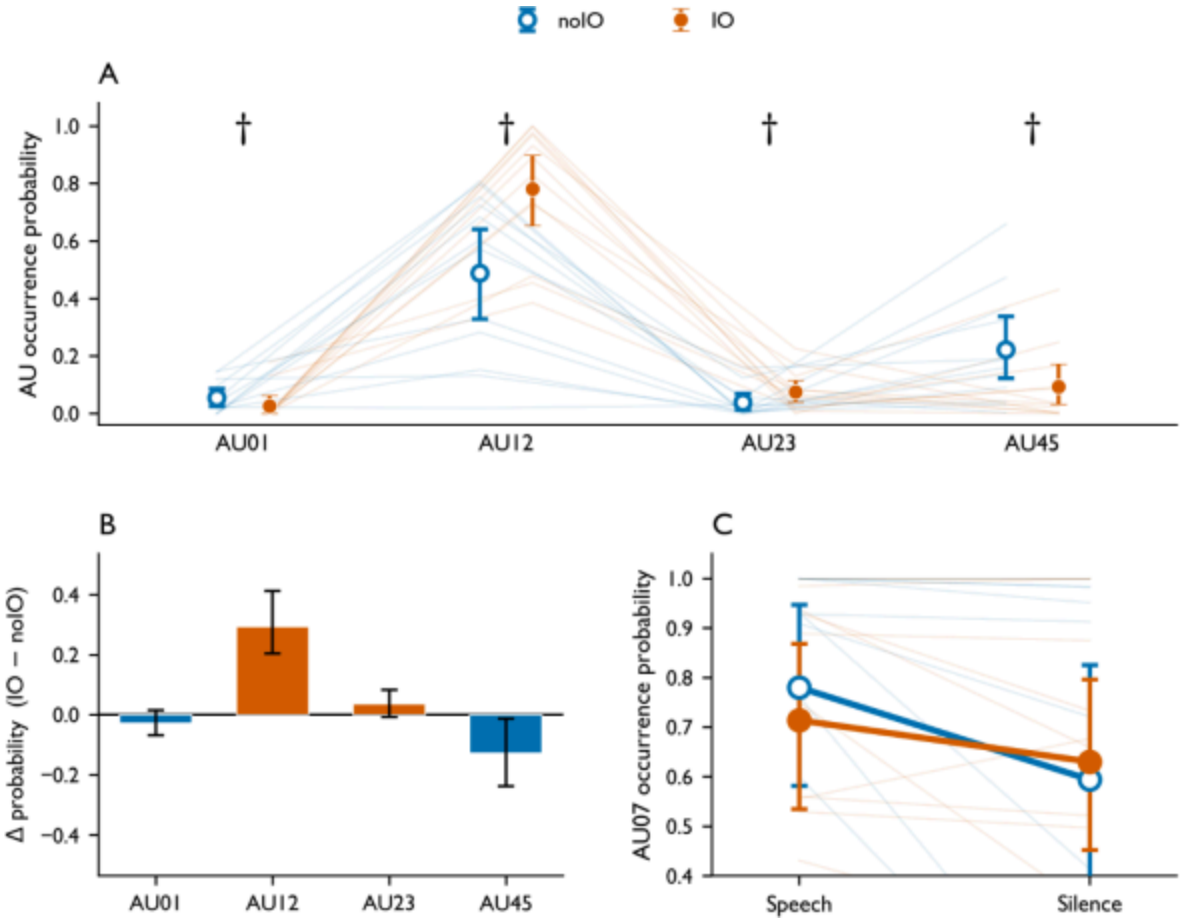


Figure 4

Two context-specific markers of Inside-Out performance. (A) Segment 1 (“opening reverie”) subtext signature. AU01 (inner-brow raiser), AU12 (lip-corner puller), AU23 (lip tightener), and AU45 (blink) under IO and noIO, plotted as in Figure 2. Daggers mark stringent BH significance from §3.2. (B) Paired difference (mean_IO − mean_noIO) for the same four AUs in Segment 1, over the 12 actors with 95 % bootstrap CIs, making the directional structure of the signature directly readable: AU12 rises strongly under IO (orange bar) while AU01, AU23, and AU45 fall (blue bars). (C) Segment 6 (“plain, plain, plain”) AU07 × speech-state interaction. Mean per-frame AU07 occurrence probability separately for Speech and Silence frames in each condition; thin lines are paired actor-level trajectories (24 = 12 × 2 in total). Under noIO the speech-vs-silence drop in AU07 is +0.19; under IO it is only +0.08, so the IO line is visibly flatter. The condition × speech-state interaction is BH-stringent ($q < .05$) in §3.5.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformationHSSC.docx