Preprints.org

Technical Note

Not peer-reviewed version

Can the Six Traditional Tamil Flavors of Arusuvai Be Represented Computationally? A Six-Dimensional Framework for Food-Related Oral Chemosensation

Vimal Aditya Raj Varadharaj *

Posted Date: 27 November 2026

Keywords: Arusuvai; sensory data interoperability; oral chemosensation; missing data; computational representation; multisensory HCI; reproducible research

Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Preprints.org | NOT PEER-REVIEWED | Posted: 27 November 2026 https://doi.org/10.20944/preprints202611.2890.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Technical Note

Can the Six Traditional Tamil Flavors of Arusuvai Be Represented Computationally? A Six-Dimensional Framework for Food-Related Oral Chemosensation

Vimal Aditya Raj Varadharaj

Independent Researcher, USA; vimaladityarajv@gmail.com

Abstract

Sensory data are difficult to compare when attribute definitions, acquisition protocols or missing measurements differ. The Arusuvai Computational Taste Framework (ACTF) formalizes six Tamil oral sensory descriptors as named, normalized coordinates with explicit provenance and protocol metadata. A versioned JSON schema and Python implementation distinguish absent ratings from measured zero, reject incompatible protocol declarations, and compute exact geometric completion bounds for incomplete profiles. Eight constructed profiles round-tripped without information loss; 28 complete distances agreed with a separately written JavaScript implementation within 1.2×10^{-16}. Across 1,764 synthetic common-mask comparisons, every reference distance lay within its calculated bounds. Removing pungency and astringency produced 36 strict ordering reversals among 168 candidate-order comparisons under zero filling; both orderings were feasible for every comparison under the completion box. A secondary analysis of four published food-cluster means mapped four measured taste attributes into ACTF while retaining pungency and astringency as missing and savouriness outside the six-coordinate schema. All six resulting conditional distance intervals overlapped. This example highlights why an apparently precise partial profile need not support a full-profile ranking. ACTF contributes an inspectable, culturally indexed data contract and conservative comparison workflow, not a new biological taxonomy or validated model of human similarity. Prospective bilingual and psychophysical studies are specified to evaluate construct validity and practical utility.

Keywords: Arusuvai; sensory data interoperability; oral chemosensation; missing data; computational representation; multisensory HCI; reproducible research

1. Introduction

Flavor is a multisensory experience involving gustatory, olfactory and oral somatosensory inputs [1–3]. Computational descriptions must therefore specify which aspects of experience they preserve and which they omit. Descriptive sensory analysis already measures multiple attributes, and recent work on digital chemical senses and computational taste modeling illustrates both the opportunities and the limitations of numerical representations [4–7]. A six-number vector cannot, by itself, recreate a food experience.

ACTF addresses a narrower but consequential problem: exchanging a culturally indexed subset of oral sensory measurements without silently conflating missingness, coordinate identity or protocol differences. Its contributions are (i) a documented mapping of selected Tamil terms to provisional sensory constructs, (ii) a named, versioned record format with an executable comparison gate, (iii) exact completion bounds and numerical tests that expose the consequences of unavailable attributes, and (iv) a secondary demonstration using published sensory aggregates and a prospective validation program. These contributions complement established measurement methods rather than replace them.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 27 November 2026 https://doi.org/10.20944/preprints202611.2890.v1

The research questions are: RQ1, can compatible profiles be exchanged and compared consistently across code paths? RQ2, how do missingness, weighting and boundary assumptions alter computed relationships? RQ3, what can be represented from published sensory measurements that do not cover all six attributes? RQ4, can the proposed bilingual constructs be measured reliably and does cultural indexing confer incremental annotation or exchange utility? RQ1–RQ3 are examined computationally; RQ4 requires future participant studies.

2. Cultural Foundation and Relationship to Prior Work

2.1. Documentary Basis and Terminology

The University of Madras Tamil Lexicon entry அறுசுவை (aṟucuvai; Arusuvai), available through an online transcription [8], enumerates six flavors: கைப்பு, இனிப்பு, புளிப்பு, உவர்ப்பு, துவர்ப்பு and கார்ப்பு. The separate entry for கைப்பு identifies it with bitterness (கசப்பு) in the six-category vocabulary [9]. This is evidence of a lexically attested classification, not of an invariant order, exclusive origin, historical priority or receptor-level taxonomy. The original institutional lexicon is indexed by the Digital South Asia Library; the specific entries were checked through online transcriptions, not directly against the scanned volumes.

ACTF adopts the operational labels in Figure 1, in a fixed computational order. The kaippu/kasappu alternatives are retained because of their documented lexical relationship; equivalence of their contemporary assessment instructions is not established. Romanizations are practical identifiers, not phonetic transcriptions. Historical textual traditions and regional usage warrant specialist review before experimental deployment.

இனிப்பு inippu Sweetness (S) Gustatory	புளிப்பு pulippu Sourness (So) Gustatory	உவர்ப்பு uvarppu Saltiness (Sa) Gustatory
கைப்பு / கசப்பு kaippu / kasappu Bitterness (B) Gustatory	கார்ப்பு karppu Pungency (P) Chemesthetic	துவர்ப்பு tuvarppu Astringency (A) Oral sensory

Figure 1. ACTF vocabulary and fixed coordinate order. English glosses are proposed operational definitions. Sensory domains describe broad associations, not exclusive biological pathways or six basic tastes. Diagram prepared for this manuscript from the terminology in [8,9].

2.2. What the Framework Adds

Standard sensory profiles and lexicons already provide attribute definitions, reference stimuli and trained-panel measurements [4,5]. Multilingual lexicons have also been developed and assessed for specific products [10]. ACTF differs in its selection of a fixed Arusuvai-indexed subset and its explicit exchange format, not in its use of vectors. Electronic tongues classify instrument responses, including in work on traditional Ayurvedic categories [11]; molecular resources such as FlavorDB relate compounds to ingredients and flavor descriptors [12]. Neither is automatically a human-rated ACTF profile. Recent computational taste research addresses prediction from molecules [7], whereas ACTF stores externally acquired or explicitly estimated ratings. In HCI, the distinction between representing a sensation and physically delivering it is equally important [6].

A conventional six-attribute profile with identical definitions, values, metadata and geometry is mathematically equivalent to ACTF. Any incremental value must therefore arise from cultural indexing, consistent documentation or exchange, not from a purportedly superior numerical formula.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 27 November 2026 https://doi.org/10.20944/preprints202611.2890.v1

2.3. Position Within Sensory Data Infrastructure

The FAIR principles emphasize discoverability, access, interoperability and reuse [13]. Visalli et al. [14] specifically propose a typology and minimum reporting standards for product-focused sensory measurements, accompanied by a data-sharing template [15]. Parma et al. [16] identify fragmented chemosensory data as a barrier to reusable computational research. ACTF is therefore not the first structured sensory format. It is a deliberately restricted, culturally indexed profile that could be embedded in richer measurement records. Its contribution is not equivalent to FAIR compliance: repository identifiers, licenses, controlled terminology and cross-system mappings remain necessary.

Generic sensory tables can preserve more than six attributes, temporal traces, replicate-level ratings and richer annotations; ACTF projects a subset into a fixed interface. A name-based mapping preserves field identity but does not validate semantic equivalence. The present evaluation tests internal exchange and conditional geometry, not conformity to the external template or interoperability across arbitrary production systems.

3. Sensory Constructs and Operational Scope

Four coordinates are conventional gustatory descriptors. Umami is an established gustatory quality [1], but it is not independently encoded in this selected traditional taxonomy. Pungency primarily concerns oral chemical irritation; capsaicin-evoked burning and menthol-evoked cooling are distinct experiences with different receptor mechanisms [17–20]. Astringency includes drying, roughness and puckering, and involves more than a single tactile mechanism [21,22]. The six columns consequently do not imply six biologically independent channels or statistically independent variables.

Table 1. Prospective attribute instructions. Reference compounds illustrate sensory qualities, not calibrated concentrations or a human-testing protocol.

Coordinate	Proposed assessment instruction	Measurement caution
Sweetness	Rate sucrose-like sweet sensation.	Distinguish sweetness from aroma associations.
Sourness	Rate acid-like sour sensation.	Do not equate acidity with drying.
Saltiness	Rate sodium-chloride-like salty sensation.	Specify anchor and food matrix.
Bitterness	Rate quinine- or caffeine-like bitter sensation.	Adjudicate Tamil lexical alternatives.
Pungency	Rate chemically induced oral burning or sharpness.	Cooling and tingling are not automatically included.
Astringency	Rate the declared drying, roughness, tightening or puckering construct.	A single score can mask divergent subqualities.

ACTF excludes aroma, independent umami intensity, texture, temperature as a sensation, visual appearance, sound and hedonic preference. Temperature and other presentation factors are nevertheless recorded as contextual metadata. A profile describes a specified sample and protocol, rather than a generic food name. Temporal instructions are particularly important for irritation and astringency, which can persist or accumulate across presentations. Taste-mixture interactions can alter individual attribute ratings, so ingredient intensities cannot simply be added to predict a final profile [23].

4. Numerical and Data Specification

4.1. Ordered Profile and Scale

For sample f, assessment context c and declared time window t, the complete ACTF profile is the vector

$$x(f; c, t) = [S, So, Sa, B, P, A] \in [0, 1]^6. \quad (1)$$

For a zero-anchored intensity scale with a fixed maximum $M > 0$ and rating r_j, the normalized coordinate is

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 27 November 2026 https://doi.org/10.20944/preprints202611.2890.v1

$$x_j = r_j / M,\ 0 \le r_j \le M. \quad (2)$$

M is the declared scale endpoint, not the maximum observed in a dataset. A nonzero lower anchor L requires an explicitly declared affine transformation $(r_j - L)/(U - L)$, $U > L$. Normalization does not establish equal perceptual increments, common salience or interval-scale properties; psychophysical anchor selection and cross-group scale comparability require independent attention [24,25].

4.2. Record Structure and Workflow

The ACTF JSON Schema (draft 2020-12) defines six required named fields S, So, Sa, B, P and A. Each is a number in [0,1] or explicit null. Unlike a bare array, exchanging an object with these field names is insensitive to key ordering. Unknown or duplicate field names, numeric strings, booleans, nonfinite values and blank required text are rejected by the schema and strict Python ingestion layer. A null value means unavailable, never measured absence.

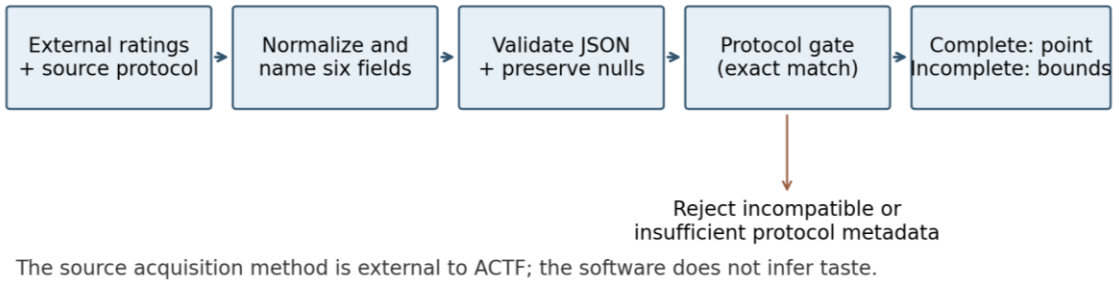


Figure 2. ACTF acquisition-to-exchange workflow. A protocol gate precedes comparison; point distance requires complete profiles, while incomplete profiles can yield conditional bounds. The workflow does not infer ratings or reproduce sensations. Diagram generated by the ACTF workflow script [32].

The record separates sample descriptors (identity, preparation, batch, concentration description, temperature in °C, portion in g) from protocol declarations (ID, definitions, language, assessor population, acquisition method, rating scale, anchors, time window in seconds and aggregation). Source and data kind distinguish synthetic, observed and estimated profiles. Numeric temperature, portion and time are structured; other fields remain textual. Missing numeric context is null. The conservative gate requires identical data kind and protocol fields and rejects unknown time windows in non-synthetic records. It does not require identical sample temperature or preparation, because such conditions may be deliberate comparison factors; analysts must check that their contrasts are scientifically justified. Exact declarations do not prove equivalent measurements.

Illustrative coordinate object (excerpt, not a complete record):

{“S”: 0.10, “So”: 0.15, “Sa”: 0.05, “B”: 0.10, “P”: null, “A”: null}

The complete schema-valid examples, Python API and separate JavaScript complete-profile comparator are included in the supplement. The API exposes a guarded point distance, guarded bounds and an explicitly geometry-only completion-bounds operation for conditional sensitivity calculations.

4.3. Aggregation and Uncertainty

For n comparable independent observations of a coordinate, the arithmetic mean and sample variance are

$$\mu_j = (1/n)\ \Sigma_i\ x_{ij};\ s_j^2 = \Sigma_i(x_{ij} - \mu_j)^2/(n - 1),\ n \ge 2. \quad (3)$$

The sample standard deviation describes dispersion; under independent identically distributed observations, $s_j/\sqrt{n}$ estimates the standard error of the mean. Repeated sessions and crossed samples and assessors require suitable hierarchical models. Distance between mean vectors is not generally

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 27 November 2026 https://doi.org/10.20944/preprints202611.2890.v1

equal to mean assessor-level distance. No population uncertainty is estimated in the synthetic experiment.

4.4. Geometry, Weighting and Collisions

For two complete, compatible profiles x and y, define Euclidean distance D_2, normalized distance D_n and geometric proximity Q:

$$D_2 = \sqrt{(\Sigma_j(x_j - y_j)^2)};\ D_n = D_2/\sqrt{6};\ Q = 1 - D_n. \quad (4)$$

All sums over j run from 1 to 6. Because $D_2 \le \sqrt{6}$, both D_n and Q are in [0,1]. Q is a monotone transformation of a geometric distance, not a probability or calibrated perceptual similarity. A weighted alternative is

$$D_w = \sqrt{(\Sigma_j\, w_j(x_j - y_j)^2)},\ w_j \ge 0,\ \Sigma_j\, w_j = 1. \quad (5)$$

Uniform weights recover D_n. Strictly positive fixed weights define a metric via coordinate rescaling; zero weights yield a pseudometric. The mapping from physical foods to vectors is many-to-one, so two distinct foods may share an ACTF profile. A zero distance means equality of represented attributes, not identity of foods or experiences.

4.5. Exact Completion Bounds for Incomplete Profiles

Let x and y be partially observed profiles, with each missing coordinate allowed to vary independently over [0,1]. For every coordinate j, define squared-difference limits ℓ_j and u_j. When both values are known, both equal $(x_j - y_j)^2$. When one value v is known, $\ell_j = 0$ and $u_j = \max(v, 1 - v)^2$. When both are missing, $\ell_j = 0$ and $u_j = 1$. Then the exact minimum and maximum weighted distances over the unrestricted completion box are:

$$L = \sqrt{(\Sigma_j\, w_j\, \ell_j)},\ U = \sqrt{(\Sigma_j\, w_j\, u_j)}. \quad (6)$$

Proof. Every squared difference is bounded by its coordinatewise limits. Nonnegative weights preserve those inequalities and the square root is monotone. Each coordinate extremum can be attained independently by selecting the missing values within [0,1], so both summed extrema are achievable. With complete profiles $L = U = D_w$. The result is conditional on the scale, weights and independent box constraints; correlations or known physiological restrictions may narrow the feasible set. It is not a statistical confidence interval. For a query profile, $U(\text{query}, a) < L(\text{query}, b)$ is a sufficient certificate that a is nearer than b under every admissible completion. Overlap alone is not proof that both rankings are attainable, particularly when the query has shared missing values.

Zero filling both missing coordinates sets their squared differences to zero, creating the lower endpoint rather than a justified point estimate. Renormalizing over available coordinates instead changes the geometry being compared. ACTF stores incomplete records, returns bounds only when the protocol gate passes, and otherwise refuses a guarded comparison. The separate geometry-only API exposes conditional bounds with explicit responsibility for contextual compatibility.

5. Computational Evaluation

5.1. Input Data and Reproducibility

Eight hand-assigned profiles from the original study are retained (Table 2). Food names are mnemonic labels for constructed numbers. The evaluation uses Python 3.13.5, NumPy 2.3.5, Matplotlib 3.10.8, jsonschema 4.26.0 and Node.js 22.16.0. The supplement includes the input vectors, ACTF schema, package source, full-precision outputs, dependencies and execution instructions. The research supplement (v2.0.2) is archived on Zenodo [32] (see Section 9).

Research software and computational procedures were developed with assistance from ChatGPT (OpenAI), as disclosed in Section 9. The archived scripts and outputs [32] document the numerical procedures used in this study.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 27 November 2026 https://doi.org/10.20944/preprints202611.2890.v1

All 28 unordered pairs and five fixed weight schemes (140 weighted results) were evaluated. All 36 tests passed locally, including the original 29 tests and seven release regression tests covering protocol key order, extreme normalization and 100 randomized asymmetric completion cases. Reruns reproduced nine numerical and record outputs and three generated figures byte for byte. This is local reproducibility, not an external audit.

Table 2. Exact synthetic input profiles. Values were manually assigned and have no empirical sampling interpretation.

label	S	So	Sa	B	P	A
Jalapeno pepper	0.10	0.15	0.05	0.10	0.90	0.05
Chai tea latte	0.70	0.05	0.08	0.12	0.22	0.18
Vanilla ice cream	0.90	0.02	0.03	0.02	0.00	0.00
Lemon juice	0.05	0.95	0.01	0.08	0.02	0.04
Black coffee	0.02	0.12	0.01	0.75	0.02	0.40
Unsweetened black tea	0.02	0.08	0.01	0.52	0.01	0.70
Salted potato chips	0.05	0.02	0.82	0.06	0.02	0.03
Dark chocolate	0.15	0.04	0.01	0.70	0.01	0.48

5.2. Pairwise and Weighting Results

For the illustrative jalapeño and vanilla vectors, the squared Euclidean distance is 1.4762, yielding D_2 = 1.2149897, D_n = 0.4960175 and Q = 0.5039825. Across the 28 constructed pairs, the smallest normalized distance is 0.0734 (black coffee–dark chocolate) and the largest is 0.5153 (vanilla ice cream–lemon juice). The closest constructed profile to jalapeño is chai tea latte (0.3765). These relationships arise entirely from the assigned values.

Table 3. Weight sensitivity for the constructed jalapeño–vanilla pair. No weights are fitted to human judgments.

Weight scheme	Weights in fixed coordinate order	D_w
uniform	(1/6, 1/6, 1/6, 1/6, 1/6, 1/6)	0.4960
pungency high	(.08, .08, .08, .08, .60, .08)	0.7344
pungency low	(.19, .19, .19, .19, .05, .19)	0.4088
astringency high	(.08, .08, .08, .08, .08, .60)	0.3455
gustatory high	(.225, .225, .225, .225, .05, .05)	0.4358

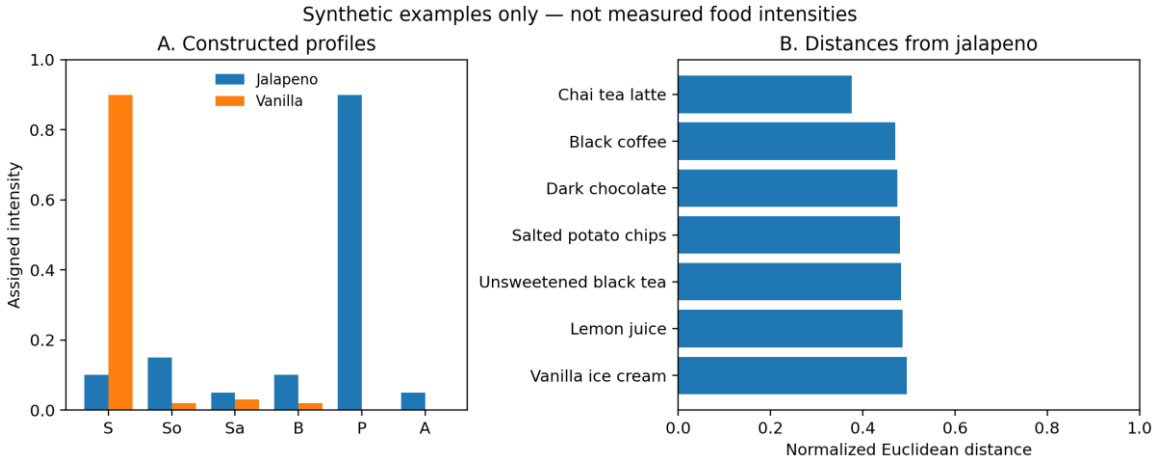


Figure 3. A: comparison of two constructed six-coordinate vectors. B: normalized distance from the synthetic jalapeño profile to the remaining seven profiles. The figure is generated directly from the supplemental CSV; the rankings are not empirical sensory findings. Original plot generated by the ACTF analysis script [32].

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 27 November 2026 https://doi.org/10.20944/preprints202611.2890.v1

5.3. Scale, Correlation and Omitted Information

For a single-coordinate difference $\delta = 0.1$, weighted distance is $0.1\sqrt{w_j}$. Under uniform weights, this is approximately 0.0408. Duplicating an identical latent contrast in two coordinates increases its contribution by $\sqrt{2}$ under the same uniform geometry, relative to placing the contrast in one coordinate. Monotone transformations (such as squaring) retain within-coordinate order but alter gaps: the contrast from 0.1 to 0.2 becomes 0.03 after squaring, whereas 0.8 to 0.9 becomes 0.17. Thus normalized ranges alone do not guarantee interpretable geometry. Covariance-adjusted or learned distances require suitable real data; none is fitted here.

5.4. Boundary Sensitivity

For each assumed Gaussian perturbation SD $\sigma \in \{0, 0.025, 0.05, 0.10, 0.20\}$, 5,000 paired draws perturb the two reference vectors independently by coordinate. The random generator is NumPy PCG64 with SeedSequence([20261103, level_index]); draw shape is (5000, 2, 6). Clipping $C(z) = \min(1, \max(0,z))$ and reflection $R(z) = 1 - |1 - (z \bmod 2)|$ operate on identical draws. Both are illustrative boundary rules, not empirically calibrated noise models.

Table 4. Paired Monte Carlo results. Sim. SD describes the distribution of simulated distances; MC SE = SD/√5000 quantifies numerical uncertainty in the simulation mean. Outside % is the fraction of coordinate draws outside [0,1] before applying either rule.

Rule	σ	Mean D_n	Sim. SD	MC SE	Outside %
clipping	0.000	0.4960	0.0000	0.0000	0.00
reflection	0.000	0.4960	0.0000	0.0000	0.00
clipping	0.025	0.4935	0.0132	0.0002	13.10
reflection	0.025	0.4902	0.0132	0.0002	13.10
clipping	0.050	0.4913	0.0251	0.0004	19.69
reflection	0.050	0.4841	0.0248	0.0004	19.69
clipping	0.100	0.4840	0.0452	0.0006	29.55
reflection	0.100	0.4630	0.0421	0.0006	29.55
clipping	0.200	0.4694	0.0744	0.0011	38.29
reflection	0.200	0.4120	0.0687	0.0010	38.29

At $\sigma = 0.20$, 38.29% of draws require boundary handling, and mean distance differs by 0.0574 between clipping and reflection. The difference is a consequence of model assumptions and the constructed inputs; it is not a human sensory confidence interval.

5.5. Record Exchange and Cross-Implementation Calculation

The evaluation imports the eight CSV rows by named attribute, constructs ACTF records, validates each, serializes to JSON, reads the records back and compares the entire objects for equality. The Python guarded distance operates directly on those records. A separate ECMAScript routine reads the exchanged JSON and independently computes normalized distances by scalar summation. All eight records round-tripped exactly; all 28 distances agreed within 1.11×10^{-16} (absolute difference). This tests two code paths within one study, not an external implementation team or a comprehensive interoperability certification. The JavaScript routine is a complete-profile comparator, not a general schema validator.

The adversarial test suite checks blank provenance, wrong keys, out-of-range values, booleans, numeric strings, NaN, duplicate JSON keys, reversed scale endpoints and time windows, incompatible protocol or data-kind declarations, nulls and incorrect weights. It also verifies invariance to JSON coordinate-field order and checks the one-unknown and two-unknown analytic extrema. These tests provide evidence of rejection behavior on specified cases; they cannot detect mislabeled measurements that are syntactically correct.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 27 November 2026 https://doi.org/10.20944/preprints202611.2890.v1

5.6. Missingness Stress Test

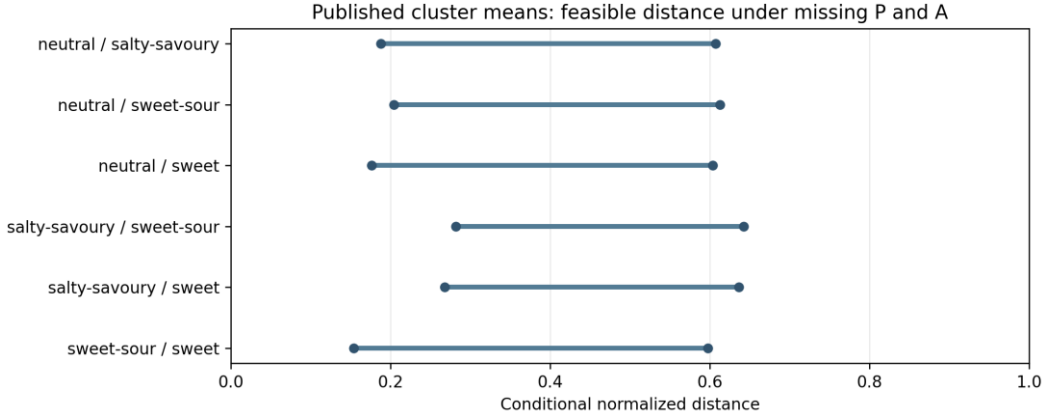
For each of the 63 nonempty subsets of the six attributes, the selected fields were masked in both members of all 28 synthetic pairs, yielding 1,764 comparisons. The original complete vector served as a hidden numerical reference; no missingness mechanism or food distribution was estimated. All 1,764 reference distances were contained in the corresponding exact completion interval. For the mask removing P and A only, zero filling underestimated complete distance by a mean of 0.0752 and the mean interval width was 0.3409. These are finite-dataset descriptive quantities, not population estimates.

To evaluate ranking effects, each of the eight profiles served as an anchor; the other seven candidates formed 21 unordered candidate pairs per anchor, totaling 168 order comparisons. With both P and A masked across all records, zero filling produced 36 strict order reversals relative to the complete synthetic profiles (ties were not counted as reversals). No comparison was certified by strictly disjoint distance intervals. Moreover, both orderings are feasible for all 168 comparisons: the difference between the known squared-distance contributions is strictly smaller in magnitude than 2/6. Assigning the missing P and A of the anchor and one candidate to zero and those of the other candidate to one adds 2/6 to one squared distance; reversing the assignments reverses that contribution. This demonstrates the risk of presenting missing values as zero in this particular example; it does not estimate the frequency of ranking errors for real sensory data.

5.7. Secondary Case Study Using Published Sensory Measurements

Viskaal van Dongen et al. [26] reported intensity ratings of five basic tastes for 50 foods evaluated by 19 subjects, using anchored 0–15 scales. Their Table 2 (p. 145) provides four cluster-level means: neutral (25 foods), salty/savoury (6), sweet/sour (9) and sweet (10). This secondary case uses those reported means, not participant-level or individual-food observations. Sweetness, sourness, saltiness and bitterness were provisionally mapped to S, So, Sa and B and divided by 15; P and A were set to null because they were not assessed. Reported savouriness remains in the source-side metadata, outside ACTF. The source cluster assignments were constructed using five tastes and are not a test of ACTF classification.

For the four aggregates, 16 of 24 coordinate positions are populated (four attributes per cluster). The source protocol provides reference scales and acquisition procedures but no exact comparable post-contact time window for these aggregate records; therefore the strict gate rejects all six pairwise comparisons. To quantify missingness alone, the geometry-only operation was used under the explicitly conditional assumption of common scale and construct interpretation. For example, the sweet/sour versus sweet clusters have lower and upper normalized distances of 0.1542 and 0.5976. The intervals for all six comparisons overlap (Figure 4), and for each pair of distinct comparisons, the known squared-distance difference is less than 2/6 in magnitude. The same endpoint construction (assigning the shared sample to zero if present) permits either ordering under the completion box. This is not a claim about actual unmeasured pungency or astringency.



© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 27 November 2026 https://doi.org/10.20944/preprints202611.2890.v1

Figure 4. Conditional completion bounds calculated from four published aggregate profiles [26], with P and A unobserved. Endpoints are feasible under the independent [0,1] completion box; they are not confidence limits. The strict protocol gate rejects these records for unknown timing. Original plot of derived results; no source figure or table is reproduced [26,32].

The case illustrates both a feasible mapping and a limit of reuse: the published data support partial records but not fully observed ACTF distances. A broader sensory dataset or new collection with the missing attributes and standardized timing would be necessary. The aggregate means are rounded by the source; within-cluster variation, subject-level dependence and measurement uncertainty are not recoverable from this calculation. The original study is cited and its factual means are transcribed in the supplemental script; no original participant data are redistributed.

6. Prospective Empirical Validation

The next research phase must distinguish whether the six proposed constructs can be measured from whether the culturally indexed format adds utility. Neither study has been conducted. The designs below are proposed protocols and will require pilot calibration and preregistration before confirmatory recruitment.

6.1. Objective A: Content and Construct Validity

A bilingual panel of Tamil linguists, culturally knowledgeable sensory specialists and trained assessors should first review documentary sources, operational definitions, regional vocabulary and alternative labels. Forward and back translation, adjudicated cognitive interviews, and independent clarity and relevance ratings can inform revisions. A feasibility pilot should assess the safety and suitability of reference stimuli and anchors, burden, timing, carryover, and variance across samples, assessors and sessions.

A subsequent sensory study would use standardized samples, randomized presentation, repeat sessions and prespecified recovery periods. Estimate repeatability and absolute-agreement reliability for each coordinate, selecting an intraclass correlation coefficient appropriate to the rater and session design and reporting uncertainty [27–29]. Reliability is necessary for many applications but is not equivalent to construct validity [30]. Evaluate possible information loss in P and A using separate quality-specific ratings and matched-intensity comparisons, rather than treating cooling as part of the default burning/sharpness definition.

For convergent evidence, independent blinded judges would rate dissimilarity explicitly over the six declared attributes. Hold out entire foods (not merely pairs) before any fitting of weights, scale transformations or covariance corrections. Use separate training and locked evaluation sets and food- and assessor-aware uncertainty for held-out associations. Generic representational dissimilarity methods provide precedent for comparing geometries with external observations [31]. A numerical association with judgments over the same six attributes can be partly tautological; it does not demonstrate superiority to ordinary sensory profiling or validity for full-flavor judgments.

6.2. Objective B: Comparative Utility of Cultural Indexing

A preregistered 2 × 2 × 2 factorial study would independently manipulate (i) bilingual Tamil-indexed versus English-only labels, (ii) fixed machine-readable schema versus flexible but equally documented entry, and (iii) required structured metadata versus optional free-text context. All eight conditions use the same sensory definitions, anchors, task content, available information, software functionality and training time. The label contrast tests the combined effect of bilingual presentation and cultural indexing; a follow-up matched bilingual control is required to isolate cultural familiarity from the mere availability of translation.

Recruit participants with documented Tamil and English proficiency; stratify by proficiency and relevant experience, and randomize assigned conditions. Use counterbalanced task sets, with

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 27 November 2026 https://doi.org/10.20944/preprints202611.2890.v1

independent source and destination implementations. Define the primary endpoint as the proportion of prespecified semantic or field- mapping errors in data transfer, adjudicated by blinded raters against a fixed reference codebook. Secondary outcomes include omission rate, completion time, annotation repeatability and comprehension. Analyze the primary error outcome using a binomial mixed-effects model with participant and item effects, factorial main effects and prespecified interactions, including label-by-language proficiency. Report absolute error-rate differences and uncertainty, not only significance tests.

6.3. Falsifiable Hypotheses and Decision Criteria

The feasibility study should determine practical effect sizes, error distributions and acceptable precision before confirmatory thresholds are selected. Sample-size simulations must reflect crossed participants and items, repeated sessions, anticipated missingness and the factorial allocation. Preregister primary outcomes, equivalence or superiority margins where relevant, exclusions, multiplicity control, missingness handling, stopping rules, and all model-selection procedures. Keep feasibility data separate from the confirmatory evaluation. Identical six- coordinate geometry is a required null-equivalence control; it cannot establish a prediction advantage for cultural labels.

Table 5. Prospective hypotheses. All require newly collected evidence; none is supported by the synthetic numerical demonstration alone.

Hypothesis	Independent evidence	Evidence against the claim
H1. Ratings can be repeated reliably	Repeated standardized presentations; agreement intervals	One or more coordinates fail prespecified precision / agreement requirements
H2. Geometry tracks six-quality judgments	Locked-food independent dissimilarity ratings	Weak or unreliable held-out association
H3. Aggregate P/A can lose quality information	Matched-intensity quality-specific judgments	No supported loss of relevant distinctions
H4. Learned weighting improves a defined prediction task	Locked-food paired evaluation against uniform baseline	No meaningful held-out improvement
H5. Tamil-indexed bilingual labels improve exchange	Factorial annotation and transfer study	Null or adverse label effect on primary error rate

7. Discussion

The principal technical result is that missing attributes can be represented without inventing values, while explicitly communicating what a distance calculation can and cannot identify. ACTF provides a named exchange format, a conservative protocol gate and an exact completion-bound calculation. The exchange tests show agreement on the supplied records, and the secondary aggregate example identifies a practical reuse barrier. Neither establishes superiority to other formats: any conventional schema with equivalent missingness, metadata and geometry can implement the same operations.

For multisensory HCI, ACTF can be viewed as an intermediate representation between external acquisition and downstream interaction. An acquisition layer records ratings and provenance; a representation layer standardizes named fields and validates records; a processing layer calculates distances or abstains; an application layer may visualize or exchange the results. A future actuation layer would need separate calibration of stimulus delivery, user-specific psychophysics and safety constraints. This division prevents an intensity vector from being mistaken for a device control signal. Bilingual education and sensory documentation are nearer-term applications; their usefulness must be evaluated through controlled user tasks.

Important limitations include provisional cross-language mapping, omission of independent umami and nonoral modalities, possible compression of astringency and pungency subqualities, scale dependence and unknown correspondence to human judgments. The exact bounds use only box constraints; the protocol gate is conservative and textual metadata can conceal semantic mismatch. The eight synthetic examples do not represent a sampled food population. The four published cluster means provide a partial reuse example, not a validation dataset. No predictive

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 27 November 2026 https://doi.org/10.20944/preprints202611.2890.v1

model, formal external-template crosswalk, independent external audit or device actuation experiment was conducted.

8. Conclusions

ACTF defines a culturally indexed, six-coordinate representation with an executable schema and comparison rules. Local exchange and numerical experiments show deterministic calculation and illustrate why incomplete profiles should not be converted silently to precise distances. Published sensory aggregates demonstrate the feasibility and limits of partial mapping. The next scientific questions concern bilingual content validity, reliable sensory acquisition, perceptual relevance and comparative usefulness. These require independent empirical evidence rather than more synthetic examples.

Supplementary Materials: The supplementary information can be downloaded at the website of this paper posted on Preprints.org

Funding: No external funding was received for this research.

Institutional Review Board Statement: No participants were recruited or contacted, and no identifiable or individual-level participant data were analyzed. The secondary example uses published aggregate results. The prospective studies would require appropriate ethics review and informed consent before implementation.

Data Availability Statement: ACTF research supplement v2.0.2 is archived on Zenodo [32]: https://doi.org/10.5281/zenodo.24515244. GitHub release: https://github.com/vimaladityaraj/ACTF/releases/tag/v2.0.2 (commit dfa7c474576b9c2219f029eb9dfaaafc9c2ecee6). This release aligns documentation and Figure 1 with this manuscript; computational inputs, methods and numerical results are unchanged from v2.0.1. The archive contains source code, schema, data [26], scripts, tests, outputs, figures and checksums. No original human participant data were collected.

Acknowledgments: ChatGPT (OpenAI) assisted with manuscript preparation, literature retrieval and checking, mathematical development, software implementation and testing, figure preparation, and formatting.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Chandrashekar, J.; Hoon, M.A.; Ryba, N.J.P.; Zuker, C.S. The receptors and cells for mammalian taste. Nature 2006, 444, 288–294. https://doi.org/10.1038/nature05401
2. Small, D.M.; Prescott, J. Odor/taste integration and the perception of flavor. Exp. Brain Res. 2005, 166, 345–357. https://doi.org/10.1007/s00221-005-2376-9
3. Spence, C. Multisensory flavor perception. Cell 2015, 161, 24–35. https://doi.org/10.1016/j.cell.2015.03.007
4. International Organization for Standardization. ISO 13299:2016, Sensory Analysis—Methodology—General Guidance for Establishing a Sensory Profile. ISO: Geneva, 2016. https://www.iso.org/standard/58042.html
5. Suwonsichon, S. The importance of sensory lexicons for research and development of food products. Foods 2019, 8, 27. https://doi.org/10.3390/foods8010027
6. Spence, C.; Obrist, M.; Velasco, C.; Ranasinghe, N. Digitizing the chemical senses: Possibilities & pitfalls. Int. J. Hum.-Comput. Stud. 2017, 107, 62–74. https://doi.org/10.1016/j.ijhcs.2017.06.003
7. Ashikhmina, M.S.; Dutta, K.; Normatov, S.; Filozop, V.S.; Volodarskiy, M.O.; Volkova, O.; Ulasevich, S.A.; Skorb, E.V. Molecular and computational basis of taste perception: A review toward the “digital language of taste”. ACS Omega 2026, 11, 26093–26102. https://doi.org/10.1021/acsomega.5c05861
8. University of Madras. Tamil Lexicon (1924–1936), entry அறுசுவை. Digital South Asia Library institutional edition: https://dsal.uchicago.edu/dictionaries/tamil-lex/; entry transcription: https://agarathi.com/word/அறுசுவை (accessed 24 November 2026).

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 27 November 2026 https://doi.org/10.20944/preprints202611.2890.v1

9. University of Madras. Tamil Lexicon (1924–1936), entry கைப்பு. Online transcription: https://agarathi.com/word/கைப்பு (accessed 24 November 2026).
10. Lippi, N.; Senger, E.; Karhu, S.; et al. Development and validation of a multilingual lexicon as a key tool for the sensory analyses and consumer tests of blueberry and raspberry fruit. Agriculture 2023, 13, 314. https://doi.org/10.3390/agriculture13020314
11. Jayasundar, R.; Ghatak, S.; Kumar, D.; Singh, A.; Bhosle, P. No ambiguity: Chemosensory-based ayurvedic classification of medicinal plants can be fingerprinted using E-tongue coupled with multivariate statistical analysis. Front. Pharmacol. 2022, 13, 1025591. https://doi.org/10.3389/fphar.2022.1025591
12. Garg, N.; Sethupathy, A.; Tuwani, R.; et al. FlavorDB: A database of flavor molecules. Nucleic Acids Res. 2018, 46, D1210–D1216. https://doi.org/10.1093/nar/gkx957
13. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 2016, 3, 160018. https://doi.org/10.1038/sdata.2016.18
14. Visalli, M.; Schlich, P.; Mahieu, B.; Thomas, A.; Weber, M.; Guichard, E. First steps towards FAIRization of product-focused sensory data. Food Qual. Prefer. 2023, 104, 104765. https://doi.org/10.1016/j.foodqual.2022.104765
15. Visalli, M.; Galmarini, M.V. Template to share sensory evaluation data. Recherche Data Gouv, 2023, V1. https://doi.org/10.57745/B35XCS
16. Parma, V.; Mainland, J.D.; Hwang, L.D.; Kedziora, R.J.; Pineau, N.; Gerkin, R.C.; Niv, M.Y. FAIR chemosensory data: Unlocking AI for flavour, food and health. Trends Food Sci. Technol. 2026, 169, 105545. https://doi.org/10.1016/j.tifs.2026.105545
17. Green, B.G. Chemesthesis: Pungency as a component of flavor. Trends Food Sci. Technol. 1996, 7, 415–420. https://doi.org/10.1016/S0924-2244(96)10043-1
18. Caterina, M.J.; Schumacher, M.A.; Tominaga, M.; Rosen, T.A.; Levine, J.D.; Julius, D. The capsaicin receptor: A heat-activated ion channel in the pain pathway. Nature 1997, 389, 816–824. https://doi.org/10.1038/39807
19. McKemy, D.D.; Neuhausser, W.M.; Julius, D. Identification of a cold receptor reveals a general role for TRP channels in thermosensation. Nature 2002, 416, 52–58. https://doi.org/10.1038/nature719
20. Roukka, S.; Puputti, S.; Aisala, H.; Hoppu, U.; Seppä, L.; Sandell, M.A. The individual differences in the perception of oral chemesthesis are linked to taste sensitivity. Foods 2021, 10, 2730. https://doi.org/10.3390/foods10112730
21. Bajec, M.R.; Pickering, G.J. Astringency: Mechanisms and perception. Crit. Rev. Food Sci. Nutr. 2008, 48, 858–875. https://doi.org/10.1080/10408390701724223
22. Wang, S.; Smyth, H.E.; Olarte Mantilla, S.M.; Stokes, J.R.; Smith, P.A. Astringency and its sub-qualities: A review of astringency mechanisms and methods for measuring saliva lubrication. Chem. Senses 2024, 49, bjae016. https://doi.org/10.1093/chemse/bjae016
23. Keast, R.S.J.; Breslin, P.A.S. An overview of binary taste–taste interactions. Food Qual. Prefer. 2003, 14, 111–124. https://doi.org/10.1016/S0950-3293(02)00110-6
24. Green, B.G.; Dalton, P.; Cowart, B.; Shaffer, G.; Rankin, K.; Higgins, J. Evaluating the "Labeled Magnitude Scale" for measuring sensations of taste and smell. Chem. Senses 1996, 21, 323–334. https://doi.org/10.1093/chemse/21.3.323
25. Bartoshuk, L.M.; Duffy, V.B.; Green, B.G.; et al. Valid across-group comparisons with labeled scales: The gLMS versus magnitude matching. Physiol. Behav. 2004, 82, 109–114. https://doi.org/10.1016/j.physbeh.2004.02.033
26. Viskaal van Dongen, M.; van den Berg, M.C.; Vink, N.; Kok, F.J.; de Graaf, C. Taste–nutrient relationships in commonly consumed foods. Br. J. Nutr. 2012, 108, 140–147. https://doi.org/10.1017/S0007114511005277
27. Shrout, P.E.; Fleiss, J.L. Intraclass correlations: Uses in assessing rater reliability. Psychol. Bull. 1979, 86, 420–428. https://doi.org/10.1037/0033-2909.86.2.420
28. Koo, T.K.; Li, M.Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J. Chiropr. Med. 2016, 15, 155–163. https://doi.org/10.1016/j.jcm.2016.02.012
29. Erratum to "A guideline of selecting and reporting intraclass correlation coefficients for reliability research". J. Chiropr. Med. 2017, 16, 346. https://doi.org/10.1016/j.jcm.2017.10.001

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 27 November 2026 https://doi.org/10.20944/preprints202611.2890.v1

30. Cronbach, L.J.; Meehl, P.E. Construct validity in psychological tests. Psychol. Bull. 1955, 52, 281–302. https://doi.org/10.1037/h0040957
31. Kriegeskorte, N.; Mur, M.; Bandettini, P.A. Representational similarity analysis—Connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2008, 2, 4. https://doi.org/10.3389/neuro.06.004.2008
32. Varadharaj, V.A.R. Arusuvai Computational Taste Framework (ACTF): Research Supplement. Version v2.0.2 [software]. Zenodo, 24 November 2026. https://doi.org/10.5281/zenodo.24515244

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.