Title: *Digital Encoding of Perfumes Using Scent Vector Codes (SVC)*

Author: Alraj Kadivar

Affiliation: Independent Researcher

Abstract

Perfumes represent one of the most complex and nuanced domains of human sensory experience. They are composed of dozens to hundreds of volatile organic compounds that interact both chemically and perceptually to generate a unique olfactory profile. Unlike sight and sound, which have long benefited from standardized digital encodings such as RGB for color and MIDI for music, smell lacks a universal digital language. At present, perfumes are studied either through **chemical analysis techniques** such as gas chromatography–mass spectrometry (GC-MS), which provide detailed molecular compositions but are resource-intensive and not scalable to large collections, or through **human evaluation panels**, which capture subjective perception but suffer from variability, cultural bias, and limited reproducibility. These constraints highlight a significant gap: while our ability to **digitally model and compute on visual and auditory data** has advanced enormously, olfactory data remains underrepresented and poorly structured.

This research proposes a new computational framework called **Scent Vector Codes (SVCs)**, designed to act as **digital fingerprints for fragrances**. The methodology involves extracting dominant **olfactory accords**—such as woody, floral, amber, or citrus—and quantifying their intensity values. These values are normalized and organized into **high-dimensional numerical vectors**, ensuring that each perfume can be encoded in a consistent mathematical structure. The vectors are then compressed into two complementary forms: a **lossless SVC code**, which preserves the full accord information in a reversible format, and a shorter **Display Code (DC)**, which serves as a compact, human-readable identifier. Together, these dual representations make scent data both machine-tractable and interpretable by humans, providing a bridge between raw chemistry and perceptual experience.

The advantages of this system are multifold. First, SVCs enable **objective comparison and clustering** of perfumes, something that has traditionally been limited to subjective similarity judgments. By mapping scents into a shared numerical space, it becomes possible to perform large-scale **fragrance similarity searches, clustering analyses, and visualization on 2D or 3D maps**. Second, SVCs facilitate **AI-driven applications**, such as recommendation systems, generative perfume design models, and adaptive fragrance personalization for consumers. Third, the encoding provides a foundation for **digital perfume cataloging**, where vast fragrance libraries can be indexed and retrieved with computational precision. Finally, because SVCs are inherently reproducible and scalable, they present a pathway toward **machine olfaction research**, potentially informing the development of digital smell sensors, wearable olfactory devices, and immersive AR/VR systems that integrate scent.

Beyond its technical contributions, this work introduces the idea of a **computational language of scent**—a framework that could transform not only perfumery but also adjacent fields such as neuroscience, chemistry, consumer technology, and even art. Just as RGB encoding enabled a revolution in how we capture, manipulate, and transmit color, SVC encoding aspires to play a similar role for fragrances. By formalizing a digital structure for olfactory data, we take an early but crucial step toward **a future where scents can be stored, searched, transmitted, and recombined with the same ease as images and sounds**.

In summary, this study presents SVCs as a novel and practical system for the **digital representation of perfumes**. It bridges the gap between chemical composition and human perception, offering a scalable and reproducible encoding that supports computational analysis, AI integration, and creative exploration. The introduction of SVCs is not simply a technical advancement but a conceptual shift: it reimagines scent as data, opening the door to new scientific insights and technological innovations in the digital era of perfumery.

Introduction

Traditionally, the analysis and understanding of perfumes has relied heavily on **human sensory evaluation**. Trained perfumers, consumer focus groups, and fragrance testers form the backbone of how scents are classified, described, and marketed. While this human-centric approach provides valuable insight into emotional and cultural responses, it is also **highly subjective**. Perception of smell varies significantly between individuals depending on biological sensitivity, prior experiences, cultural background, and even temporary conditions such as fatigue, mood, or health. This inherent variability creates challenges for research and industry alike, where **consistency, reproducibility, and comparability** are critical.

In contrast, fields such as color and sound have long benefited from the development of **standardized digital representations**. For instance, RGB encoding in digital media allows colors to be expressed precisely as numerical values, enabling them to be stored, compared, transmitted, and reproduced without ambiguity. Similarly, music can be encoded in systems like MIDI, which capture notes, timing, and dynamics in a machine-readable form. These digital standards not only facilitated scientific study but also unlocked revolutions in technology, from digital photography and printing to music streaming and editing. By comparison, the world of **olfactory data** has no universally accepted digital language. Despite the fact that perfumes are among the most complex sensory products we encounter, their representation remains either **chemically grounded** (lists of volatile compounds obtained via laboratory analysis) or **linguistically grounded** (human-generated labels such as “floral” or “woody”). Neither approach offers the **structured, quantitative digital encoding** needed for computational operations at scale.

This project seeks to **bridge that gap** by introducing a novel framework for the **numerical encoding of perfume scent profiles**. Unlike approaches aimed at physically reproducing or duplicating smells, the purpose here is not to simulate the olfactory experience itself, but rather to create a **computationally tractable digital representation**. In other words, our goal is to answer: *How can we represent a perfume in numbers, in a way that is consistent, reproducible, and meaningful across contexts?*

The proposed method is built on the following key objectives:

1. **Dissection of fragrances into discernible accords:**
   Every perfume is broken down into its fundamental olfactory components, known as **accords**. These accords—such as *fresh, woody, floral, amber, green,* or *musky*—represent dominant perceptual categories that shape how the perfume is experienced. Identifying and weighting these accords provides a bridge between the chemistry of the formula and the language of perception.

2. **Quantification of each accord using intensity values:**
   To make accords computationally useful, they must be **quantified**. Each accord is assigned a numerical intensity value, indicating its relative prominence in the fragrance profile. By normalizing these values, perfumes of varying complexity can be represented in a consistent numerical structure. This allows precise comparisons between fragrances, regardless of whether they contain 5 accords or 50.

3. **Transcoding into dual codes for usability and precision:**
   To ensure that the numerical representation is both **human-readable and machine-tractable**, we introduce a dual coding system. The **Display Code (DC)** is a short, compact identifier designed for quick use in user interfaces, comparisons, or cataloging. It is non-reversible, meaning it does not expose the underlying accord details, but acts as a concise label for reference. The **Scent Vector Code (SVC)**, on the other hand, is a **lossless encoding** of the

perfume's accord intensities. It preserves the full detail of the scent profile and allows accurate reconstruction of the original accord values. Together, DC and SVC balance the needs of **usability and data integrity**.

4. **Enabling computational operations on perfumes:**
 With perfumes translated into numerical form, a range of computational techniques become possible. Perfumes can now be **indexed like data points**, searched efficiently, and visualized in 2D or 3D maps using dimensionality reduction techniques such as PCA or t-SNE. They can also be **clustered automatically** based on similarity, enabling the discovery of new fragrance families or hidden relationships between perfumes. Beyond analysis, this representation opens the door to **AI-driven applications**, such as personalized recommendation systems, similarity search engines, and even computational tools for novel perfume design.

This study is significant because it proposes, for the first time, a **scalable, reproducible, and human-readable digital representation of perfumes**. By formalizing scent into numerical codes, we establish the foundation for a new era of **digital perfumery**, where fragrances can be studied, indexed, and manipulated with the same rigor and precision as other forms of sensory data. Such a framework is not only valuable for fragrance researchers and the perfume industry but also holds potential for adjacent fields such as **artificial intelligence, neuroscience, human-computer interaction, and virtual reality**, where scent could play an increasingly important role.

In summary, the introduction of **Scent Vector Codes (SVCs)** represents a first step toward a standardized digital language of scent. Much like RGB for color, this system provides a universal format for representing olfactory information—enabling computation, comparison, and creativity in ways that were previously inaccessible.

- This approach is significant because it introduces a novel, reproducible, and human-readable method for digital perfume representation. By providing a consistent and structured way to encode fragrances, the system lays the groundwork for robust perfume databases, AI-assisted fragrance analysis, and new computational research into olfactory patterns and similarities. Unlike traditional evaluation methods, this digital representation removes subjectivity while retaining meaningful information about the perfume’s character, making it both scientifically rigorous and practically useful.

Perfume Accord Extraction

Perfumes are decomposed into fundamental accords, which are perceived scent categories. These accords map to specific chemicals.

Sample accords and chemicals used in this study:

Accord	Chemicals
citrus	Linalool, Citronellol
fresh	Linalool, Hedione
floral	Geraniol, Hedione
white floral	Hedione, Geraniol
woody	Iso E Super, Ambroxan
amber	Ambroxan, Vanillin
musky	Musk Ketone
spicy	Eugenol

Example perfume profile:

Index	Accord	Intensity
0	white floral	100
1	fresh	97.6
2	woody	80.5

3	green	78.4
4	floral	78.2
5	mossy	71.6
6	aldehydic	68.8
7	yellow floral	64.7
8	amber	60.9
9	earthy	59.1

These intensity values are **normalized to a scale of 0–100** and represent the **strength of each accord**.

Chemical Mapping and Formula Generation

Each accord maps to **specific chemicals**.

Step-by-step:

1. Normalize accord keys (lowercase, strip spaces).
2. Initialize a **10-chemical zero vector**.
3. Distribute accord intensity across mapped chemicals.
4. Normalize the chemical vector to sum 100.
5. Encode as a **Formula Code**:

Example formula code for the above perfume:

F-23-15-10-12-08-07-06-09-05-05

This represents **percentage contributions of the 10 chemicals**.

SVC Encoding

The **Scent Vector Code (SVC)** is a **lossless, unique digital fingerprint** for each perfume.

Steps

1. Normalize accord dictionary.
2. Sort keys and convert to **JSON string**.
3. Compress with `zlib`.
4. Encode in **base64**.
5. Append **CRC32 checksum** for verification.

```
svc_code = encode_svc_from_accords(accords)
```

Example SVC for sample perfume:

SVC-eNpVkl1S6zAMhbfC5LmTSe9A_3ajOgrxYFu9stxQGPaOCkiBN32W5HMs-50qsFv5vIGuCKf_CM6QKeUXEusCrURS5K3xBgADkMUPGZDQj882BAiUQNI6lgJp2FG7mLGhpaeI0TehtVOLk2ggssyo8HfvtppsSMaj4_tD_U2Ksc3c67vvdDzzUSwzmZ-IWxeCZEctX4-OmU9dnsDfMVNCqIvvsEgRRJ8HximV0WwlBZqcMHIv5zyiQ1nVkzbCLZapVxfbbu-Xc6osp6zDcKzXbM739snAB0RG3FI1pUUPWxFTNAbf8E1VIfmuFK3l1zeTSleCyxpP8GWKlZq-sC6JtTOgMIdinECZt8UdK09-52rnqj0wJjCIlv2YBzn_kljkKPtiat8P3IdGoBYehf9p0N0yJFi_ZPfb7j0-jJe40-EDE37F33

This **SVC string is reversible** to retrieve original accords.

Dual-code System

To facilitate **quick UI display**:

- **SVC Vector:** Deterministic numeric vector aligned to axes.
- **Display Vector:** 3D compressed representation for plotting.
- **Display Code (DC):** Short SHA-based non-reversible code for easy reference.

```
codes = create_dual_codes_from_accords(accords)
```

Outputs:

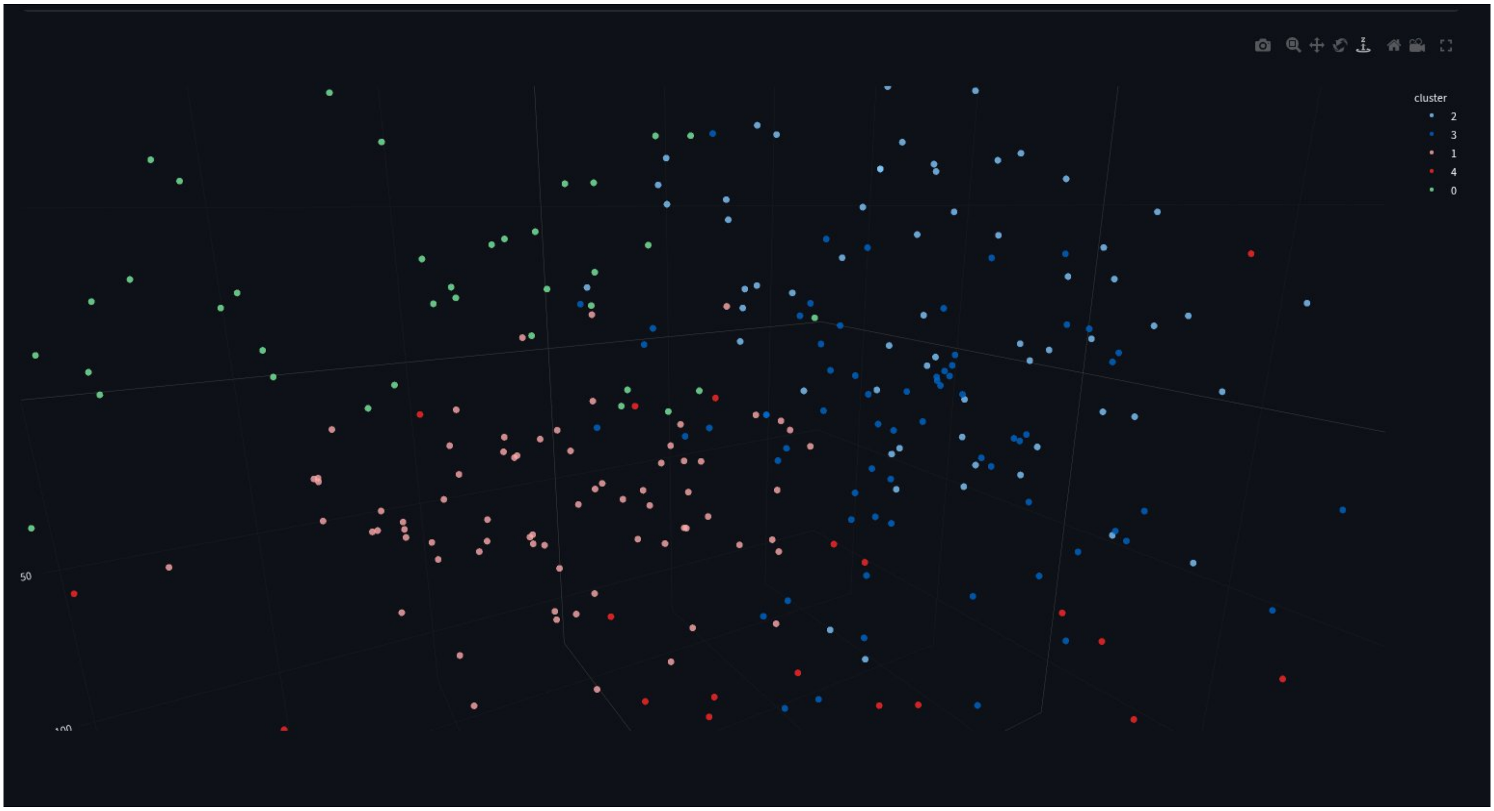
- `codes["svc_code"]` → SVC string
- `codes["display_code"]` → DC, e.g., `DC-1A2B3C4D`

- codes["svc_vector"] → normalized vector for clustering
- codes["display_vector"] → 3D display vector for plotting

Visualization and Clustering

- **2D/3D PCA or t-SNE maps** are used for clustering perfumes.
- Color can represent **clusters** or **SVC-derived scent vectors**.

- Example clusters show **similar perfumes grouped**



Step-by-step:

1. Compute PCA/t-SNE embedding from svc_vector.
2. Apply KMeans clustering.
3. Plot in 2D/3D using matplotlib or Plotly.
4. Annotate with perfume names and top accords.

Step-by-Step Example

Given the perfume table:

1. Generate chemical formula:

F-23-15-10-12-08-07-06-09-05-05

2. Generate SVC:

   SVC-eNpVkl1S6zAMhbfC5L...

3. Build svc_vector:

   [0.12, 0.11, 0.09, 0.08, 0.08, 0.07, 0.06, 0.06, 0.05, 0.04, 0.14, 0.1]

4. Compute display_vector for plotting:

   [0.33, 0.35, 0.32]

5. Generate DC:

   DC-1A2B3C4D

Applications

1. AI-driven perfume recommendation systems
By encoding each scent as a unique SVC code, recommendation systems can move beyond simple user reviews or sales data and instead match users to fragrances based on **quantifiable scent similarity**. For example, a user who prefers a perfume with floral-top and woody-base accords could be matched with other fragrances that have similar SVC codes,

enabling highly personalized and **data-driven fragrance suggestions**. This approach could also support dynamic recommendations that evolve with user preferences, seasonal trends, or even regional scent popularity.

2. Digital perfume cataloging and search
Traditional perfume catalogs are often textual and descriptive, relying on subjective notes like "woody," "citrus," or "musky." ScentByte introduces **numerical representations of scent profiles**, allowing perfumes to be cataloged and searched in a structured, computational way. Users could search for perfumes by specific scent characteristics, or even by a target SVC code, effectively creating a **digital library of scents** that is searchable, sortable, and analyzable at scale.

3. Fragrance similarity analysis
SVC codes allow for precise **comparative analysis between perfumes**, something previously limited to subjective evaluations or expert panels. By mapping perfumes into a multi-dimensional scent space, it becomes possible to calculate **quantitative similarity scores**, identify clusters of closely related fragrances, and uncover subtle relationships between scents that would otherwise go unnoticed. This capability can benefit retailers, fragrance researchers, and even consumers exploring alternatives to their favorite perfumes.

4. Computational perfume design research
While ScentByte is **not intended for cloning fragrances**, representing scents as SVC codes opens new avenues for computational research. Researchers can study the relationships between chemical compositions and perceived scent profiles, identify trends in popular fragrance families, or simulate how adjusting specific accords shifts a perfume’s position in the scent

space. This numerical framework can also support the development of **next-generation AI tools** for fragrance ideation, predicting consumer responses, or optimizing blends based on target scent characteristics—without ever needing to physically mix chemicals.

5. Market and trend analysis
SVC-coded perfumes enable large-scale analysis of the fragrance market. Brands and researchers can track the **evolution of scent trends**, identify emerging popular scent families, and perform data-driven forecasting of consumer preferences. Unlike subjective surveys, this approach provides a **quantitative, reproducible method** for understanding and predicting market behavior.

6. Enhanced educational and research tools
ScentByte's numerical representation system can serve as a **learning and research platform** for perfumers, chemists, and fragrance enthusiasts. By visualizing perfumes in a multi-dimensional scent space, users can explore **relationships between chemical compositions, accords, and perceptual experiences**, facilitating more effective training, experimentation, and hypothesis testing in fragrance studies

Discussion

- **Strengths**
  The primary strength of the ScentByte framework lies in its **reproducibility and computational accessibility**. By encoding perfumes as SVC codes, every scent can be represented numerically in a structured, standardized manner, which allows for **consistent analysis across datasets and experiments**. Unlike subjective human evaluations, which can vary depending on mood, expertise, or environment, SVC codes provide a

reversible mapping between chemical compositions and perceived scent profiles. This reversibility ensures that any computational analysis—whether clustering, similarity searches, or trend mapping—can be traced back to the original chemical components and accords. Additionally, the system is **highly scalable**: new perfumes can be incorporated into the database without disrupting existing analyses, and the computational nature of the approach allows for integration with a wide variety of AI and data analytics tools.

- **Limitations**
  Despite its advantages, ScentByte has inherent limitations. Most notably, it **does not reproduce scents physically**, meaning the SVC codes are purely analytical and cannot replace actual fragrance experiences. The system also **depends on the accuracy of chemical accord extraction**; incomplete, inconsistent, or ambiguous data can affect the fidelity of SVC representation and subsequent analyses. Another limitation is that **subjective perception nuances**—such as the influence of concentration, volatility, or human olfactory sensitivity—cannot be fully captured by numerical codes alone. Finally, as a computational model, ScentByte requires continuous **data validation and curation**, particularly as new perfumes, chemicals, or accords are introduced into the system.
- **Future Work**
  ScentByte lays the groundwork for several promising research directions. A key avenue is the **integration of machine learning models** to predict perfume relationships and potential new blends based on SVC similarity scores. Such models could identify hidden patterns in scent spaces, uncover underexplored fragrance combinations, and assist perfumers in computational ideation. Additionally, **cross-modal analysis** could be explored, linking SVC-coded scents with other sensory or consumer data,

such as color palettes, product packaging, or purchase behavior. Expanding the framework to include **dynamic temporal profiles**—capturing how a perfume evolves over time after application—could further enhance predictive capabilities. Lastly, ScentByte could serve as a **foundation for collaborative databases**, enabling researchers and brands to share scent information in a reproducible, standardized format, fostering innovation and more objective fragrance research.

Conclusion

The SVC and dual-code system provide a robust, standardized framework for the digital representation of perfumes, analogous to how RGB codes represent colors. By encoding each fragrance numerically, this approach enables quantitative analysis, clustering, and similarity assessments across a wide range of perfumes. The system allows for computational visualization of scent spaces, facilitating a deeper understanding of relationships between fragrances and their chemical accords. Importantly, while it does not reproduce scents physically, it provides a reproducible and reversible method for mapping perfumes into a structured, analyzable format.

This framework opens new avenues for research, AI-driven recommendation systems, and computational exploration of fragrances, offering tools for perfumers, researchers, and educators alike. Furthermore, the integration of SVC codes with machine learning and predictive models promises to enhance data-driven fragrance ideation, trend analysis, and consumer insights, ultimately bridging the gap between traditional olfactory studies and modern computational methods. As digital scent representation matures, systems like ScentByte can serve as a foundation for collaborative fragrance research, enabling standardized, scalable, and objective analysis in perfumery.