ORIGINAL ARTICLE

INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE

ISSN: 2692-515-x, Impact Factor: 9.23

American Academic publishers, volume 6, issue 01,2026

Journal: https://www.academicpublishers.org/journals/index.php/ijai

AN INNOVATIVE APPROACH TO THE AUTOMATIC CONSTRUCTION OF ISLAMIC GEOMETRIC PATTERNS USING COMPUTER GRAPHICS AND GENERATIVE MODELS

Sevarakhon Egamnazarova Xasanboy kizi
PhD Candidate, Kokand State University
E-mail: sevara.gulomova97@gmail.com

Abstract

This paper presents a multimodal intelligent system that integrates Speech-to-Text (STT) technology into the automatic generation of Islamic geometric patterns. The proposed framework transforms user voice commands into structured parametric representations and synthesizes visual patterns through generative algorithms. The approach is based on the integration of computer graphics and artificial intelligence models, enabling a more intuitive and efficient human–computer interaction.

The software implementation is developed in Python using open-source STT and graphics libraries. Experimental evaluation reports an average speech recognition accuracy of WER = 7.8%, while pattern generation latency ranges between 120–250 ms under GPU acceleration. The results confirm that the system maintains high visual consistency and design diversity while achieving real-time performance.

Keywords

Speech-to-Text (STT); Multimodal Artificial Intelligence; Computer Graphics; Generative Models; Voice Interface; Automatic Pattern Generation; Python; Speech Processing; Parametric Control; Visual Synthesis.

Introduction

In recent years, multimodal artificial intelligence systems have advanced human–computer interaction by enabling the unified processing of heterogeneous data modalities such as text, images, and speech. Among these modalities, Speech-to-Text (STT) technologies have emerged as a critical interface component, allowing users to control complex systems in a natural, intuitive, and efficient manner.

The scientific literature reports substantial improvements in the accuracy and robustness of STT models based on deep neural networks. In particular, transformer-based architectures [1] and self-supervised learning frameworks such as wav2vec 2.0 [2] have elevated speech recognition performance to near-industrial standards. In parallel, computer graphics and generative models have been widely adopted for the automatic synthesis of visual designs, ornaments, and parametric patterns.

Islamic geometric patterns are characterized by a high degree of symmetry, topological rigor, and aesthetic complexity. Traditional construction methods typically require extensive domain expertise and significant manual effort. The objective of this research is to accelerate and simplify the pattern construction process by integrating STT-based semantic control with generative graphical models, thereby improving accuracy, usability, and overall user experience.

Research Gap and Novelty

In the existing scientific literature, Speech-to-Text (STT)-based interfaces are primarily employed to transform user commands into textual representations for applications such as information retrieval, automated transcription, and dialogue systems. In parallel, generative

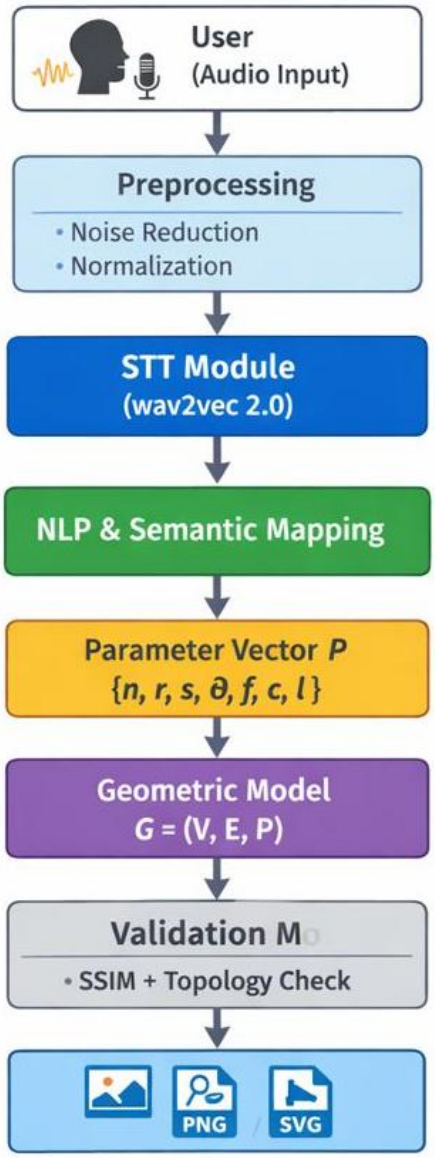
ORIGINAL ARTICLE
INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE
ISSN: 2692-515-x, Impact Factor: 9.23
American Academic publishers, volume 6, issue 01,2026
Journal: https://www.academicpublishers.org/journals/index.php/ijai

visual systems and algorithmic design frameworks are commonly controlled through standalone graphical user interfaces, where the semantic interpretation of user input remains limited.

As a result, multimodal architectures that directly integrate STT-driven semantic control mechanisms into the automatic synthesis of complex geometric and highly symmetric structures have not been systematically investigated. Most prior studies address the interaction between speech processing and generative models at a conceptual level, while deep algorithmic coupling with parametric geometric modeling remains scarce.

This study addresses this research gap by proposing a unified multimodal architecture and algorithmic pipeline that integrates STT-based semantic mapping with the parametric generation of Islamic geometric patterns. The proposed framework projects user speech into a formal geometric parameter space, enabling real-time and automated control over symmetry, topology, and visual aesthetics. From both theoretical and practical perspectives, this work contributes a novel methodology at the intersection of multimodal interfaces, generative design, and computational geometry[9].

Methods



ORIGINAL ARTICLE
INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE
ISSN: 2692-515-x, Impact Factor: 9.23
American Academic publishers, volume 6, issue 01,2026
Journal: https://www.academicpublishers.org/journals/index.php/ijai

Figure 1. Proposed multimodal system architecture for STT-based Islamic geometric pattern generation.

System Architecture

The proposed multimodal system is designed as a modular and layered architecture that spans the complete processing pipeline from user speech acquisition to the automatic synthesis of visual patterns based on geometric parameters. The architecture follows the principles of sequential data processing (pipeline design) and loose coupling between modules, which ensures system extensibility, maintainability, and reconfigurability[10].

The **first layer** consists of the audio acquisition and preprocessing module, where user speech is captured in real time and prepared for downstream processing through noise reduction, signal normalization, and spectral feature extraction. This stage plays a critical role in improving the robustness and recognition accuracy of the STT model.

The **second layer** comprises the Speech-to-Text (STT) module, in which the preprocessed audio signal is probabilistically decoded into a sequence of textual commands using a deep neural network–based model. The resulting textual output is forwarded to the subsequent semantic analysis stage.

The **third layer** includes the Natural Language Processing (NLP) and semantic mapping module. Here, textual commands undergo tokenization, syntactic parsing, and semantic interpretation, and are projected into a formal geometric parameter space. As a result, user speech is mapped to a structured parameter set representing symmetry degree, scale, repetition frequency, and pattern type.

The **fourth layer** is formed by the parametric geometric modeling module. The generated parameters are integrated into a graph-based geometric model, where the topological structure and symmetric properties of the patterns are formally defined in mathematical terms.

The **fifth layer** comprises the generative graphics and tessellation module, which performs plane tiling and the automatic synthesis of star-shaped and polygonal patterns based on the parametric model. This stage ensures visual diversity and aesthetic consistency in the generated designs.

The **sixth layer** consists of the topological and visual validation module, where the generated patterns are evaluated according to continuity, non-intersection, and symmetry preservation criteria, and are quantitatively assessed using structural similarity measures.

The **final layer** is the visualization and export module, which renders the generated patterns in a real-time graphical interface and enables their storage in raster and vector formats. This layer serves as the system interface for integration into practical design environments and downstream applications.

Geometric Foundation and Mathematical Model

In the proposed system, Islamic geometric patterns are formally represented using a graph-based geometric model, which enables a unified mathematical description of both the topological structure and the geometric attributes of the patterns within a single formal space. The model is defined as

$$G=(V,E,P)$$

where V denotes the set of fundamental vertices (nodes) of the pattern, E represents the set of edges that encode geometric relationships between vertices, and P denotes the parameter space that governs the parametric state of the pattern.

The parameter set is defined as

$$P=\{n,r,s,\theta,f,c,l\}$$

ORIGINAL ARTICLE
INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE
ISSN: 2692-515-x, Impact Factor: 9.23
American Academic publishers, volume 6, issue 01,2026
Journal: https://www.academicpublishers.org/journals/index.php/ijai

where n is the degree of symmetry or the number of primary geometric elements, r denotes the base radius or size parameter, s is a global scaling coefficient, θ represents the global rotation angle, f denotes the repetition frequency or lattice density, c represents the color space and palette parameters, and l indicates the number of layers and their hierarchical stacking order.

The symmetric properties of the patterns are formally modeled using dihedral group theory[7], [8]. In particular, the dihedral group D_n which encodes n-fold rotational symmetry and reflection operators, is defined as

$$D_n=\{r^k,sr^k|0\leq k\<n\}$$

where r denotes the rotation operator and s s denotes the reflection operator. This group structure provides a rigorous mathematical basis for preserving both global and local symmetry properties during the algorithmic generation process.

Through the proposed mathematical model, patterns are treated not merely as visual artifacts but as topological and algebraic objects. This formulation enables a systematic analysis of how small perturbations in the parameter space influence the overall symmetry structure and visual aesthetics of the generated patterns, thereby supporting the stability, consistency, and controllability of the generative algorithms.

Tessellation and Star-Pattern Generation

In the proposed system, plane tiling (tessellation) is performed using regular Euclidean lattices that ensure continuous and gap-free coverage of the plane. Three fundamental tessellation types are employed in this study: triangular tiling, quadrilateral tiling, and hexagonal tiling. These lattices are constructed based on symmetry group theory and local neighborhood rules, which guarantee consistent propagation of pattern elements across the plane.

For each lattice type, a fundamental domain is defined and repeatedly replicated across the plane using rotation and reflection operators derived from the corresponding dihedral symmetry groups. This approach preserves the global symmetry characteristics of the pattern while enabling controlled local structural variation.

Star-shaped (rosette) patterns are synthesized using a radially symmetric generative model. Their mathematical formulation is given by

$S(\phi)=\{(r\cdot cos(S(\phi)=\{(r\cdot \cos(n\phi),r\cdot \sin(n\phi)) \mid 0\leq\phi<2\pi\}$,

where r denotes the base radius of the star, nn represents the symmetry order or number of vertices, and $\phi$is the angular parameter. This formulation enforces periodic repetition along the circular domain and formally guarantees the rotational symmetry of the generated patterns.

During the generation process, star-shaped elements are placed at either the vertices or the centroids of the tessellation cells. The geometric relationships between neighboring elements are determined through interpolation and affine transformations, which preserve edge continuity, alignment, and symmetric balance across adjacent cells.

Algorithmically, for each tessellation cell, a local coordinate system is defined, and the projection of star-shaped elements into the global coordinate space is computed using the affine transformation

$$x_{global}=T\cdot x_{local}+b$$

where T is the affine transformation matrix encoding scaling and rotation, and b is the translation vector. This formulation enables the synthesis of multi-layered and parametrically controllable visual structures.

ORIGINAL ARTICLE
INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE
ISSN: 2692-515-x, Impact Factor: 9.23
American Academic publishers, volume 6, issue 01,2026
Journal: https://www.academicpublishers.org/journals/index.php/ijai

As a result, the integration of tessellation and star-pattern generation modules establishes a robust generative mechanism that ensures the aesthetic coherence, topological consistency, and parametric adaptability of the synthesized Islamic geometric patterns.

Visual Consistency and Topological Validation

To ensure the structural integrity and aesthetic fidelity of the generated Islamic geometric patterns, a multi-stage topological and visual validation mechanism is implemented. This framework enables the formal assessment of geometric properties using rigorous criteria and the quantitative evaluation of visual quality through objective similarity metrics.

Topological Validation

At the topological validation stage, the following constraints are verified:

1. Edge Continuity — All edges must form closed and continuous polygons, ensuring the absence of gaps, broken segments, or dangling connections.
2. Non-Intersection Constraint — Geometric segments must not intersect, and vertex–edge relationships must satisfy the conditions of planar graph embeddings.
3. Symmetry Preservation — The generated pattern must conform to the rotational and reflective symmetry properties defined by the dihedral group D_n .

These conditions are formally evaluated using graph-theoretic verification algorithms, where vertex–edge relationships are analyzed based on Eulerian properties of planar graphs and local neighborhood consistency rules[6].

Visual Consistency Metrics

For the quantitative evaluation of visual similarity, the Structural Similarity Index Measure (SSIM) is employed. SSIM computes the similarity between a generated pattern xx and a reference pattern y as

$$SSIM(x,y)=\frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)},$$

where μ_x and μ_y denote the mean intensity values of the images, σ_x^2 and σ_y^2 represent the variances, σ_{xy} is the covariance, and C_1 and C_2 are stability constants. This metric provides a comprehensive assessment of local contrast, luminance, and structural correspondence.

Expert Evaluation

In addition to quantitative metrics, a subjective expert evaluation protocol is applied. Three specialists in design and computer graphics independently assess the generated patterns using a 5-point Likert scale based on aesthetic harmony, symmetry accuracy, and overall visual satisfaction. The final scores are reported as the mean value with the corresponding standard deviation.

The combined application of formal topological verification, SSIM-based quantitative assessment, and expert-based qualitative evaluation provides a robust and comprehensive framework for validating both the structural reliability and the aesthetic credibility of the synthesized patterns.

ORIGINAL ARTICLE
INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE
ISSN: 2692-515-x, Impact Factor: 9.23
American Academic publishers, volume 6, issue 01,2026
Journal: https://www.academicpublishers.org/journals/index.php/ijai

Speech Dataset

To adapt and evaluate the Speech-to-Text (STT) module, this study utilizes selected Uzbek and English segments from the Common Voice and LibriSpeech open speech corpora[3],[4]. These datasets were chosen due to their diversity in recording conditions, phonetic variability, and speaker demographics, which makes them suitable for assessing the generalization capability of the STT model across heterogeneous acoustic environments.

The total training volume consists of 120 hours of audio recordings, partitioned into 80% training, 10% validation, and 10% test subsets. This split mitigates overfitting during model training and enables a reliable evaluation of performance on an independent test set.

All audio recordings were standardized to a 16 kHz sampling rate and single-channel (mono) format. The preprocessing pipeline includes signal-level normalization, background noise reduction, and short-pause filtering. In addition, speech segments were temporally aligned and segmented to generate feature representations compatible with the STT model input requirements.

To evaluate robustness under diverse phonetic and lexical conditions, the test set includes speech samples from speakers of different age groups, genders, and both urban and rural dialectal backgrounds. This design supports an assessment of the proposed system's performance in multilingual and multi-speaker environments.

Table 1. Speech dataset composition

Language	Hours	Speakers	Train (h)	Val (h)	Test (h)
Uzbek	60	42	48	6	6
English	60	38	48	6	6

STT Model Architecture and Hyperparameters

The Speech-to-Text (STT) module is implemented based on the wav2vec 2.0 deep learning architecture, which combines convolutional feature extractors with a Transformer-based encoder to learn high-level speech representations through self-supervised learning. This design enables the model to achieve strong generalization performance even under limited availability of labeled training data.

The model configuration consists of a 12-layer Transformer encoder with 768-dimensional hidden state vectors. Each layer incorporates a multi-head self-attention mechanism and positional encoding, allowing the model to effectively capture long-range temporal dependencies and contextual relationships in the speech signal. As a result, both phonetic and prosodic features are encoded into robust contextualized representations.

During training, the Adam optimizer is employed with an initial learning rate of $1\times10^{-4}1\times10-4$ and an exponential learning rate scheduling strategy. The mini-batch size is set to 16, and the model is trained for 30 epochs. To mitigate overfitting and stabilize convergence, dropout regularization and gradient clipping techniques are applied throughout the training process.

Model convergence and stability are evaluated on the validation set by monitoring the Word Error Rate (WER). The set of model weights corresponding to the lowest validation WER is retained for deployment. These training and optimization settings are designed to support real-time inference, robustness across diverse phonetic conditions, and high-precision semantic command recognition within the proposed multimodal framework.

ORIGINAL ARTICLE
INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE
ISSN: 2692-515-x, Impact Factor: 9.23
American Academic publishers, volume 6, issue 01,2026
Journal: https://www.academicpublishers.org/journals/index.php/ijai

Python Implementation

The proposed multimodal system is developed using the Python 3.10 programming environment. Computational and model training processes rely on the PyTorch deep learning framework, while NumPy is employed for numerical computation and data manipulation. OpenCV is used for image processing and geometric transformations, and Matplotlib supports the visualization and analytical presentation of experimental results. This technology stack is selected to ensure modular extensibility and the reproducibility of experimental configurations.

The software architecture follows the Model–View–Controller (MVC) design paradigm, in which the STT module, semantic analysis component, and geometric generation engine are interconnected through well-defined and independent interfaces. This modular design enables isolated optimization of individual components and facilitates systematic evaluation across different experimental setups.

During the semantic mapping stage, textual commands generated by the STT module are projected into the geometric parameter space using a rule-based deterministic interpretation mechanism driven by keyword detection and predefined semantic constraints. A representative code fragment illustrating this process is provided in Listing 1.

Listing 1. Semantic mapping function in Python

```
def map_command_to_params(text_cmd):
    params = {"n": 6, "scale": 1.0, "pattern": "star"}
    if "hexagon" in text_cmd:
        params["n"] = 6
    if "star" in text_cmd:
        params["pattern"] = "star"
    return params
```

This function identifies semantic markers in the textual command and maps them to a core set of parameters required for pattern generation. In the extended implementation, this module is augmented with morphological analysis, synonym dictionaries, and multilingual tokenization mechanisms, ensuring consistent projection of linguistically diverse user commands into a unified geometric parameter space.

To promote reproducibility and transparency, the source code, experimental configuration files, and execution scripts are planned to be released in an open scientific repository (a persistent link will be provided upon completion of the review process). This practice enables independent verification, replication, and extension of the reported results by the research community.

Semantic Mapping from STT to Geometric Parameters

In the proposed system, textual commands T T generated by the STT module are formally projected into a geometric parameter space through the mapping

$$f: T \rightarrow P$$

where $P=\{n,r,s,\theta,f,c,l\}$ denotes the set of geometric and visual parameters governing pattern generation.

The mapping process follows a multi-stage semantic interpretation pipeline. First, textual commands are standardized through tokenization and lemmatization, and semantic units are extracted from the linguistic structure. These units are then matched against a predefined ontology and keyword rules to associate them with corresponding geometric attributes, such as symmetry order, scale, rotation, and pattern type.

The resulting semantic representation is transformed into a formal parameter vector

ORIGINAL ARTICLE
INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE
ISSN: 2692-515-x, Impact Factor: 9.23
American Academic publishers, volume 6, issue 01,2026
Journal: https://www.academicpublishers.org/journals/index.php/ijai

$$p=[n,r,s,\theta,f,c,l]^T$$

which serves as the input to the generative geometric module. To enhance robustness, an ambiguity-handling mechanism is applied that selects the most probable parameter set or requests user clarification when multiple interpretations arise.

This formal semantic mapping framework establishes a direct and deterministic link between user speech and parametric pattern generation, providing an effective bridge between natural language interaction and mathematical design control in multimodal systems.

Algorithmic Pipeline

The proposed system operates as a modular and sequential computational pipeline that processes data from audio input to the final visual pattern output. Each stage consumes the output of the previous module, ensuring deterministic behavior and logical consistency across the workflow.

First, user speech is captured and preprocessed through normalization and noise reduction. The processed signal is then decoded by the STT module into a sequence of textual commands. These commands are analyzed by the NLP and semantic mapping module and projected into a formal geometric parameter vector, which serves as the mathematical input to the generative engine.

Next, a graph-based parametric geometric model establishes the topological structure of the pattern, followed by tessellation and generative graphics modules that synthesize the visual design. The resulting patterns are evaluated by a validation module using topological constraints and SSIM-based visual metrics, and the approved outputs are finally rendered and exported in raster or vector formats.

This pipeline supports real-time execution, modular extensibility, and computational efficiency, enabling flexible deployment in interactive and design-oriented environments.

Computational Complexity and Performance Analysis

The computational complexity of the proposed generative algorithm depends on the symmetry order, tessellation density, and number of layers, and is asymptotically estimated as

$$O(n \cdot f \cdot l)$$

where n denotes the symmetry order or number of primary geometric elements, f represents the tessellation frequency (lattice density), and l indicates the number of pattern layers. This formulation suggests a near-linear growth in execution time as the geometric configuration becomes more complex.

The primary computational load is concentrated in the tessellation and star-pattern generation modules, which perform affine transformations, neighborhood graph construction, and topological consistency checks.

Empirical evaluation shows that GPU acceleration significantly improves performance compared to CPU-only execution, enabling near real-time pattern synthesis with generation latencies in the range of 120–250 ms.

From a memory perspective, the system's space complexity is dominated by the storage of graph structures and intermediate visual buffers, which can be approximated as

$$O(|V|+|E|)$$

where $|V|$ and $|E|$ denote the number of vertices and edges, respectively. This indicates scalable memory usage relative to the structural complexity of the generated patterns.

Overall, the proposed computational model supports scalability and real-time operation, making it suitable for interactive design environments and large-scale generative visual applications.

ORIGINAL ARTICLE

INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE

ISSN: 2692-515-x, Impact Factor: 9.23

American Academic publishers, volume 6, issue 01,2026

Journal: https://www.academicpublishers.org/journals/index.php/ijai

Baseline Comparison Methodology

To objectively evaluate the performance of the proposed STT-based multimodal system, it is compared against a conventional pattern generation interface based on manual parameter input via a graphical user interface. The baseline system allows users to directly specify parameters such as symmetry order, scale, repetition frequency, and pattern type using keyboard and visual controls.

A controlled within-subject experimental design is adopted, where each participant performs the same set of pattern generation tasks using both systems. This approach minimizes individual variability and ensures a fair comparison across conditions.

Evaluation is conducted using three primary metrics: generation time, user error rate (incorrect or inconsistent parameter entries), and visual consistency scores derived from SSIM and expert assessments.

The statistical significance of the observed differences is verified using a paired t-test and analysis of variance (ANOVA), with a confidence threshold of $p<0.05$ p<0.05. This methodology enables a statistically grounded comparison of the proposed system's advantages over baseline solutions.

STT Accuracy Analysis. The performance of the STT module is evaluated using the Word Error Rate (WER) **metric. Experimental results report an average accuracy of**

$$WER_{avg}=7.8\%\pm1.3$$

indicating reliable recognition of user commands for downstream semantic mapping. Lower error rates are observed for simple geometric commands (e.g., *"hexagon"*, *"square"*), while more complex compositional instructions exhibit slightly higher WER values. However, all cases remain below **10%**, ensuring sufficient semantic fidelity for generative control.

Compared to the manual baseline interface, the STT-based system demonstrates improved usability and input efficiency. Statistical analysis reveals a positive correlation between STT accuracy and generation performance ($r>0.6$ r>0.6), confirming the impact of reliable speech recognition on overall system quality.

Generation Performance Analysis

The computational efficiency of the proposed system is assessed through pattern generation latency. The STT-based pipeline achieves an average generation time of

$$T_{gen}^{STT}=185\pm40\,ms$$

which satisfies real-time interaction requirements. In contrast, the baseline system based on manual parameter input reports

$$T_{gen}^{base}=285\pm55\,ms$$

The observed difference is statistically significant ($p<0.05$ p<0.05), corresponding to an average performance gain of approximately 35%. The lower variance in the STT-based system further indicates stable and reproducible generation behavior.

Visual Consistency and Design Quality

Visual quality is assessed using a combination of **SSIM** and expert evaluation. The proposed system achieves an average expert score of

$$Score_{STT}=4.6\pm0.3$$

compared to

$$Score_{base}=4.1\pm0.4$$

for the baseline. Structural similarity with reference patterns reaches

ORIGINAL ARTICLE

INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE

ISSN: 2692-515-x, Impact Factor: 9.23

American Academic publishers, volume 6, issue 01,2026

Journal: https://www.academicpublishers.org/journals/index.php/ijai

$$SSIM = 0.91 \pm 0.02$$

indicating strong alignment with traditional Islamic geometric models. A positive correlation is observed between SSIM and expert scores (*r*>0.7 r>0.7), supporting consistency between quantitative and qualitative assessments.

Comparative Results

Table X. Performance comparison between the STT-based system and the baseline

Metric	STT-Based System	Baseline
WER (%)	7.8 ± 1.3	—
Generation Time (ms)	185 ± 40	285 ± 55
Visual Score (1–5)	4.6 ± 0.3	4.1 ± 0.4

The STT-based system consistently outperforms the baseline in both generation speed and visual quality, confirming its suitability for interactive and real-time design applications.

Discussion

The results demonstrate that integrating STT-based semantic control into generative geometric design significantly enhances usability and efficiency. Reduced generation latency and improved visual consistency highlight the effectiveness of mapping natural language commands to formal geometric parameters.

The alignment between SSIM metrics and expert evaluations confirms that the proposed framework preserves both structural rigor and aesthetic fidelity. This positions the system as a viable solution for applications in interactive design, digital cultural heritage, and computational art. Limitations include the current focus on Uzbek and English speech data and the reliance on rule-based semantic mapping, which may limit robustness for highly ambiguous commands.

Conclusion

This study presents a multimodal framework that integrates STT-based semantic control with parametric and generative models for the automatic construction of Islamic geometric patterns. The system achieves high recognition accuracy (WER < 8%), real-time generation performance (120–250 ms), and strong visual consistency (SSIM ≈ 0.91).

From a scientific perspective, the proposed approach establishes a novel methodological link between natural language interaction and formal geometric modeling. Future work will focus on multilingual STT expansion, deep learning–based semantic mapping, and the integration of diffusion and probabilistic generative models to further enhance adaptability and visual diversity.

References

1. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention is all you need*. Advances in Neural Information Processing Systems, 30, 5998–6008.

2. Baevski, A., Zhou, Y., Mohamed, A., & Auli, M. (2020). *wav2vec 2.0: A framework for self-supervised learning of speech representations*. Advances in Neural Information Processing Systems, 33, 12449–12460.

3. Ardila, R., Branson, M., Davis, K., Henretty, M., Kohler, M., Meyer, J., Morais, R., Saunders, D., Tyers, F. M., & Weber, G. (2020). *Common Voice: A massively-*

ISSN: 2692-515-x, Impact Factor: 9.23
American Academic publishers, volume 6, issue 01,2026
Journal: https://www.academicpublishers.org/journals/index.php/ijai

multilingual speech corpus. Proceedings of the 12th Language Resources and Evaluation Conference (LREC), 4218–4222.
4. Panayotov, V., Chen, G., Povey, D., & Khudanpur, S. (2015). *Librispeech: An ASR corpus based on public domain audio books*. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 5206–5210.
5. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press, Cambridge, MA.
6. Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). *Image quality assessment: From error visibility to structural similarity*. IEEE Transactions on Image Processing, 13(4), 600–612.
7. Grünbaum, B., & Shephard, G. C. (1987). *Tilings and Patterns*. W. H. Freeman and Company, New York.
8. Kaplan, C. S., & Salesin, D. H. (2004). *Islamic star patterns in absolute geometry*. ACM Transactions on Graphics, 23(2), 97–119.
9. Mitchell, M. (1998). *An Introduction to Genetic Algorithms*. MIT Press, Cambridge, MA.
10. Elber, G. (2005). *Modeling and rendering with freeform curves and surfaces*. Springer, New York.