POLITECNICO DI TORINO

Master's Degree in ICT for Smart Societies

Masters's Degree Thesis

Mapping the Neighborhood of Microtonal Music Scales Using Self-Organizing Maps to Enhance Modulation Techniques and Discover Innovative Shifts

Candidate

KOUSHA NIKKAR

Supervisor

CRISTINA EMMA MARGHERITA ROTTONDI

February 2025

Abstract

This study presents a novel method for mapping music scales in two dimensions, incorporating microtones, using a self-organizing map. Modulation, key to shaping emotional experience, relies on interval similarities in Iranian music's Radif. This research systematically explores key and Gushe relationships, considering microtonal variations and the Shahed note as an emotional anchor.

The self organazing map algorithm was utilized to cluster eight distinct interval patterns spanning both Western and Persian musical theories based on variations in tonic and dominant notes, along with their microtones states. A total of 1,176 musical scale states were systematically generated and clustered, incorporating semi-tonic and microtonic variations. The training of self organizing map ensured both the representation of known scales and the emergence of new, previously unexplored patterns. Quantitative metrics such as quantization error, topographic error, and reconstruction percentage were employed to evaluate the clustering results, achieving high accuracy and retaining 100% of the original data in the final map.

A key application of the generated map is to provide smooth modulation pathways between musical scales, allowing for gradual transitions that remain imperceptible to the listener. Using Bresenham's Algorithm, modulation pathways were optimized for computational efficiency while ensuring an intuitive auditory experience. To further enhance user understanding, a melody generator was developed using Python's Mido library, which translates these modulation pathways into audible transitions, leveraging quarter-tone adjustments and note-centric weighting.

The evaluation process included both quantitative and qualitative methods. Quantitative analysis measured the smoothness of modulation using average step differences and hit/miss rates for predicted modulation pathways, yielding an 83% hit rate. Qualitative assessments involved a user survey with 72 participants, who rated 28 generated melodies based on their perceived mood continuity. The results confirmed the system's ability to propose modulation pathways that align with traditional practices while introducing innovative transitions.

This research demonstrates the potential of self organizing map for uncovering new pathways in musical modulation, bridging the gap between traditional and algorithmic approaches, and providing valuable tools for composers exploring nuanced emotional shifts in their work.

Acknowledgements

I would like to express my deepest gratitude to those who have supported me throughout my thesis journey. Their encouragement, patience, and belief in me have been invaluable.

Family Members

To my beloved parents and my two brothers, whose unwavering support and sacrifices have been the foundation of my journey. I am deeply grateful for your encouragement, patience, and belief in my pursuit of knowledge. Your love and motivation, even from a distance, have kept me focused and inspired every step of the way.

Supervisor

I extend my sincere appreciation to my supervisor, Professor **CRISTINA EMMA MARGHERITA ROTTONDI**, for her invaluable guidance and trust. Her insightful feedback, dedication, and encouragement have greatly shaped my academic growth. Her patience and commitment have been instrumental in my progress.

Friends and Colleagues

Lastly, to all my friends, colleagues, and anyone who contributed to my academic journey, thank you. Your presence, discussions, and shared experiences have made this path memorable. The conversations we had, the debates we engaged in, and the moments of shared struggle and triumph have all shaped my understanding and approach to research. Your friendship and camaraderie have made this journey not only intellectually stimulating but also personally fulfilling. Whether through words of encouragement, shared laughter, or simply being there in times of need, you have all contributed significantly to my success. This thesis is as much a reflection of my hard work as it is a testament to the incredible support system that has surrounded me.

Table of Contents

List of Figures

Acronyms

SOM
: Self-Organizing Maps

QE
: Quantization Error

TE
: Topographic Error

RP
: Reconstruction Percentage

TN
: Tonal Neighborhood

CT
: Computational Analysis

GC
: Genre Classification

HA
: Harmonic Analysis

MR
: Music Recommendation

GS
: Gushe

SM

Smooth Modulation

MP

Modulation Pathway

QI

Qualitative Input

AI

Artificial Intelligence

BMU

Best Matching Unit

MT

Microtone

SV

Sigma Value

LR

Learning Rate

NI

Number of Iterations

MIDI

Musical Instrument Digital Interface

TPR

True Positive Rate

FNR

False Negative Rate

Chapter 1

Introduction

Introduction

Music is a universal medium for expressing emotions, ideas, and stories, achieved through a combination of melody, harmony, rhythm, and tonal dynamics [1]. Across cultures and historical periods, music has played an essential role in human expression, serving as a means of storytelling, emotional release, and cultural identity [2]. One of its most fascinating aspects is its ability to evoke a wide range of emotions, creating a profound connection between the listener and the composition. Whether through a melancholic melody, an uplifting harmonic progression, or a dynamic rhythmic pattern, music has the power to influence mood and perception [3]. Among the many techniques composers use to shape a piece's emotional and structural narrative, modulation, the transition between musical keys, stands out as an essential tool. [4]. This technique allows composers to guide listeners through contrasting emotional landscapes while maintaining coherence and enhancing the overall complexity of the composition.

In any musical work, preserving the tonal integrity of the key is fundamental to creating a harmonious experience [5]. The choice of key defines the tonal center, influencing the emotional character of a piece and the relationships between notes and chords. However, composers often choose to modulate, or change the key, deliberately to introduce contrast, expand expressive possibilities, and sustain listener interest [6]. modulation is particularly effective when executed seamlessly, allowing the transition to unfold naturally without disrupting the flow of the composition. When done skillfully, key changes can create a sense of movement and evolution, leading the listener through different tonal environments while preserving musical unity [7].

In the context of Persian classical music, modulation is deeply rooted in the tradition of the Radif, a structured system of melodic patterns passed down through generations [8]. The Radif provides a framework for both composition and improvisation, guiding musicians in their exploration of different tonal spaces [9]. A defining feature of Persian music is its extensive use of Microtone (MT), which contribute to its rich tonal palette and unique expressive qualities [10]. These microtonal inflections allow for nuanced modulations that might not conform to the rigid structures found in Western tonal music. In Persian music, modulation often revolves around the axis notes, known as the Shahed, which serves as a melodic anchor and establishes a sense of resolution and emphasis [11]. The Shahed is a crucial element in defining the emotional direction of a piece, acting as a reference point for modulation between different scales or modes [12].

Traditionally, musicians have relied on intuition and empirical knowledge to navigate Modulation Pathway (MP), drawing on their understanding of the relationships between the tonic, dominant, and axis notes of different keys [13]. These relationships are influenced by the intervallic similarities between modes, facilitating smooth and coherent transitions [14]. Despite the significance of modulation in both composition and performance, a systematic and Computational Analysis (CT) of MPs has remained largely unexplored [15]. The absence of a formalized, data-driven approach has left much of the modulations process to the subjective interpretation and experience of musicians [16]. While this allows for artistic freedom, it also means that modulation techniques are not always structured or easily accessible for study.

This research seeks to bridge the gap between tradition and modernity by employing CTs methods to analyze MPs across Persian and Western classical music scales [17]. The primary objective is to construct a two-dimensional map of musical modes that visually represents their neighborhood relationships and provides a structured framework for modulation [18]. By utilizing Self-Organizing Maps (SOM), a type of Artificial Intelligence (AI) neural network, this study clusters modes based on their interval patterns, tonal characteristics, and shared modulation possibilities [19]. The resulting visualization not only offers theoretical insights into the relationships between different modes but also serves as a practical tool for musicians and composers [20]. This approach enables a systematic exploration of SMs pathways, expanding the expressive possibilities of music while preserving the richness of both Persian and Western classical traditions [21]. By integrating CTs with traditional music theory, this research aims to provide a novel perspective on modulation, offering both theoretical advancements and practical applications in composition and performance [22].

Thesis Structure

This thesis is structured in a way that facilitates a comprehensive exploration of the research objectives, methods, and findings. The document is divided into the following chapters:

1. **Introduction:** The first chapter introduces the research problem, objectives, and scope of the study. It provides an overview of the importance of modulation in music, particularly within the context of Persian and Western music scales, and presents the research questions that guide the thesis.

2. **Literature Review:** Chapter 2 presents an extensive review of the relevant literature. This includes a discussion on modulation in music theory, the use of Microtone (MT) in Persian classical music, and the application of Self-Organizing Maps (SOM) in music theory and pattern recognition. The literature review also covers the principles behind the SOM algorithm and its application in other domains.

3. **Methodology:** In Chapter 3, the methodology used for this research is described in detail. This chapter covers the data collection process, including the musical scales and interval patterns considered in the study. It also explains the use of the SOM algorithm for clustering musical scales and generating modulation paths, as well as the technical aspects of the algorithm, such as the initialization of nodes and the calculation of modulation paths.

4. **Results and Discussion:** Chapter 4 presents the results of the SOM algorithm applied to Persian and Western music scales. The generated map of musical scales is analyzed, and the modulation paths between different scales are discussed. The findings are also compared with existing theories on musical modulation. The chapter further includes a qualitative evaluation of the AI-generated modulation paths through field surveys and user feedback. This evaluation covers the criteria for assessing the smoothness, emotional impact, and effectiveness of the modulation transitions.

5. **Conclusion and Future Work:** The final chapter summarizes the main findings of the research and discusses their implications for music theory and composition. It also suggests potential areas for future research, including the refinement of the modulation system and its application to other musical traditions.

Chapter 2

Research Background

2.1 Theoretical and Musical Context

The process of modulation has been an integral aspect of music theory and composition for centuries, playing a crucial role in the structural and emotional development of musical pieces. Modulation allows composers to shift between different keys or tonal centers, providing emotional contrast, dynamic movement, and variety within a piece. It serves as a powerful tool to sustain listener engagement by introducing new harmonic colors, expanding the expressive range, and creating a sense of development and resolution [23].

In Western classical music, modulation is typically analyzed in terms of key relationships, harmonic functions, and formal structures. The study of tonal harmony has placed significant emphasis on the rules governing diatonic and chromatic harmony, cadence progressions, and MPs that define the transitions between keys [24]. These modulations are often guided by harmonic sequences, chord progressions, and pivot chords that ensure smooth key changes while maintaining coherence [25].

2.1.1 Modulation in Persian Classical Music

In Persian classical music, modulation plays a more nuanced role, shaped by the specific characteristics of its scales, the use of MTs, and the unique modal system it employs [26]. Unlike Western classical music, which relies heavily on harmonic relationships, modulation in Persian music is primarily melodic and microtonal, emphasizing the continuous flow of melodic phrases rather than harmonic progression [27].

The concept of modulation in this tradition is deeply rooted in the Radif, a structured collection of melodic motifs and modal structures passed down through generations [28]. The Radif provides both a theoretical framework and a practical guide for composition and improvisation, outlining the relationships between different melodic units and establishing the conventions for modulation within the Persian musical system [29].

The Radif consists of a series of "Gushe (GS)" or smaller melodic motifs, that are grouped into larger modal systems called "Dastgah." Each "Dastgah" represents a unique tonal and emotional space, and the transitions between "GSs" within a "Dastgah" or between different "Dastgah"s form the basis for modulation [30].

2.1.2 Computational Approaches to Modulation Analysis

Despite the rich theoretical foundation of modulation in Persian music, there has been limited computational exploration of how MPs can be systematically mapped and analyzed [31]. Traditionally, musicians and composers have relied on their deep understanding of the Radif and the inherent emotional qualities of different scales to make intuitive decisions about modulation [32].

With advancements in computational tools and AIs, new opportunities have emerged for analyzing musical structures in an objective and systematic manner [33]. The development of machine learning models, particularly neural networks, has enabled researchers to identify patterns in musical data, classify different tonal relationships, and explore new approaches to music analysis [34].

2.1.3 Self-Organizing Maps (SOM) in Music Analysis

Among these computational techniques, the SOMs algorithm has shown particular promise in clustering and visualizing complex relationships between different musical elements [35]. Developed by Teuvo Kohonen, SOMs are a type of unsupervised AIs neural network designed for dimensionality reduction and pattern recognition [36].

SOMs have been widely applied in fields such as image processing, speech recognition, and financial forecasting, but their use in music theory and analysis remains relatively novel [37]. The ability of SOMs to cluster data based on similarity makes them a powerful tool for organizing musical modes into structured maps, revealing hidden relationships between different tonal systems [38].

2.1.4 Research Objectives and Contributions

This research seeks to address the gap by leveraging SOMs to generate a structured, visual representation of modes based on their intervallic relationships [39]. The goal is to provide a computational tool that can identify smooth MPs, taking into account both Western and Persian musical traditions [40].

Furthermore, this study introduces an innovative method for evaluating MPs by analyzing the distances between adjacent steps in the modulation process [41]. The research also incorporates qualitative evaluations through field surveys, helping validate the effectiveness of the proposed MPs [42].

Additionally, the study considers the potential applications of these findings in both compositional practice and music education [43]. By providing composers with a structured framework for exploring modulations, this research offers practical tools that can enhance the creative process [44].

2.2 Research Objectives and Scope

The primary objective of this research is to explore and develop a computational framework for understanding, analyzing, and generating MPs between musical scales, with a particular focus on Persian and Western music traditions. Modulation, as a crucial element in musical composition, enables composers to transition between different tonal centers, providing dynamic contrast and emotional depth. This study employs the SOMs algorithm to systematically map musical scales and their relationships by considering key elements such as tonic, dominant, and axis notes, while also taking into account the MTs nature of Persian music. Through this approach, the study aims to bridge traditional theoretical knowledge with modern computational methods.

To achieve this overarching goal, the research is structured around the following key objectives:

1. Examine the Neighborhood of Keys and GSs in Persian Music.
2. Analyze Interval Changes and Clustering of Key Classes.
3. Apply the SOMs Algorithm to Musical Scales.
4. Evaluate the Effectiveness of AI-Generated MPs.
5. Investigate the Role of the Evident Note (Shahed) in Modulations.
6. Compare Traditional and Computational Approaches to Modulations.
7. Develop a Practical Application for Composers and Music Theorists.

2.2.1 Scope of the Research

The scope of this research encompasses both Persian and Western musical traditions, focusing specifically on how modulation operates within these systems. In Persian music, the emphasis lies in understanding microtonal transitions and the unique modal structures defined by the Radif and Dastgah frameworks. For Western music, the study examines harmonic progressions and key relationships based on diatonic and chromatic principles.

The computational aspect of the research involves the application of SOMs for dimensionality reduction and visualization of tonal relationships. By combining quantitative evaluations, such as Quantization Error (QE), Topographic Error (TE), and Reconstruction Percentage (RP), with qualitative assessments from musicians and listeners, this research ensures a holistic understanding of modulation. The ultimate goal is to develop practical tools that support both compositional creativity and theoretical exploration, bridging traditional insights with modern computational techniques.

2.3 Modulation in Music Theory

Modulation refers to the process of changing the key or tonal center in a piece of music. It is a key concept in both Western and non-Western musical traditions. It allows composers to explore different emotional landscapes, expand musical possibilities, and create contrast in a musical composition. Modulations provide a sense of progression and transformation, and increase the emotional depth of a piece. It is a technique that enables the composer to develop musical ideas by changing the tonal center, creating a sense of movement and variety. Modulations can occur in a variety of ways, smoothly or abruptly, with the latter typically creating a more dramatic effect that can evoke stronger emotional responses. Modulations are often intended to evoke a sense of change or development. They guide the listener's emotional journey and add a dynamic quality to the music. The transition between keys can highlight contrasts between different emotional states, giving rise to feelings of tension, release, or resolution [45].

Figure 2.1 illustrates the number of the smallest musical units, called commas, between each musical note within an octave, totaling 53 commas. Altering the spacing of notes within the octave generates a new algorithm, evoking various emotional and psychological responses in the listener. Additionally, a fundamental condition for forming a musical scale is that, in any configuration of note intervals within an octave of that scale, there must always be 53 commas [46].

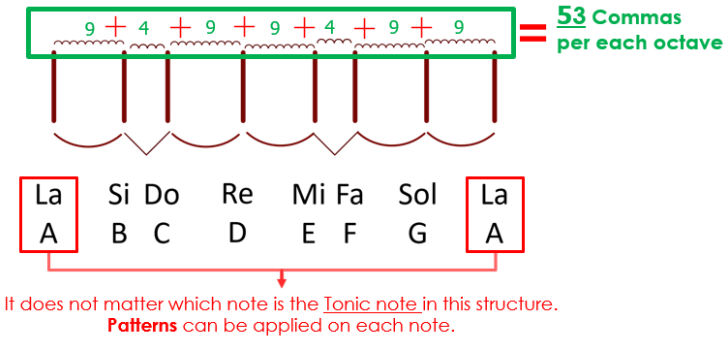


Figure 2.1: The intervals between musical notes create a distinct pattern.

In Western classical music, modulations are traditionally achieved through a variety of techniques. These include the pivot chord method, direct modulation, and common-tone modulation. In the pivot chord method, a chord common to both the current key and the new key is used to bridge two tonal centers. Direct modulation involves an immediate and clear change from one key to another, often without any transitional chords, resulting in an abrupt but often effective change. Common-tone modulations use a single note or tone that remains constant while the surrounding harmonies change, making the change less noticeable. These techniques help ensure that modulations are smooth and consistent, allowing for a seamless transition from one key to another. The smoothness of modulations is considered crucial to maintaining the flow of the music and ensuring that the listener does not perceive the change as jarring or disruptive. This ability to modulate smoothly enhances the emotional continuity of the music, providing a sense of unity even as the tonal center changes [46].

Figure 2.2 illustrates that the position of each note can be represented as an index in an array, where each index corresponds to a comma of spacing. Any musical scale depicted in this system assigns the tonic note a fixed index of 0. For example, if the second degree of the scale has an index of 9 and the third degree has an index of 13, there is a distance of 4 commas (equivalent to a musical semitone) between them. In this figure, the pattern of the A minor scale is visible, with each degree having an index corresponding to its position [47].

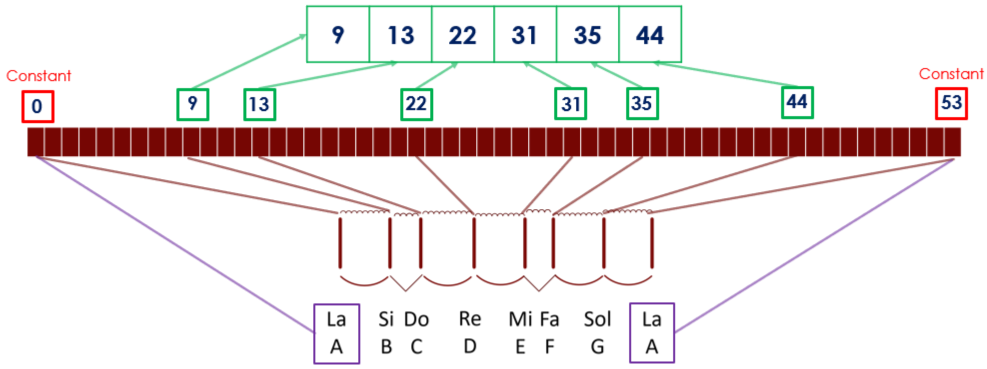


Figure 2.2: Example of the A minor scale, showing the index of each degree on the list of available commas.

In Persian classical music, known as Radif, modulations is similarly an essential element for emotional expression. The Radif consists of a series of musical phrases and melodic patterns that serve as a structure for improvisation and composition. These phrases are organized in a way that guides the performer in creating music that is both structured and free-flowing. The concept of modulations in Persian music, however, differs from Western practices, as it involves shifts between various modes, called "Dastgah." These Dastgah are distinct melodic structures that define the emotional and tonal landscape of a piece. The use of MTs, which are intervals smaller than those found in the Western tuning system, plays a key role in the modulations process. Persian music often involves modulating between Dastgah that are similar in terms of their intervallic relationships, creating subtle changes in tone without abrupt shifts. The smoothness of modulations in Persian music is largely dependent on the degree of similarity between different Dastgah, which influences how pleasant or abrupt the modulations feels to the listener. When the transition between two Dastgah is smooth and gradual, it enhances the emotional impact of the music, providing a sense of continuity and development. However, if the modulations is less smooth, it can create a sense of tension or surprise, adding dramatic effect to the piece.

Therefore, while the basic concept of modulations is common in both Western and Persian music, the methods and effects of modulations differ significantly due to distinct musical structures and cultural contexts. The use of MTs and the emphasis on smooth transitions in Persian classical music offer unique challenges and opportunities for composers and performers, distinguishing it from the approaches used in Western classical traditions. These differences underscore the richness of modulations as a tool for emotional expression, as it allows musicians to explore the full range of tonal possibilities within their respective musical systems [48].

2.4 Bresenham's Algorithm

Bresenham's Algorithm, developed by Jack Bresenham in 1962, is a highly efficient algorithm used to draw straight lines on a raster grid (pixel-based grid). It is an integer-based algorithm that minimizes the need for floating-point operations, making it particularly useful in computer graphics, especially on devices with limited processing power. This section provides a step-by-step explanation of the algorithm, including the relevant mathematical formulas and key steps of the method. The details of Bresenham's Algorithm can be found in [49].

2.4.1 Step 1: Initialization

The first step in Bresenham's line-drawing algorithm is to initialize the decision parameters. Given two points (x_0, y_0) and (x_1, y_1) that define the line, the algorithm computes the differences in the x and y directions:

$$\Delta x = x_1 - x_0, \quad \Delta y = y_1 - y_0, \tag{2.1}$$

where Δx is the horizontal change and Δy is the vertical change between the two endpoints.

Next, the initial decision parameter p_0 is computed. This parameter helps to determine the next pixel to be selected. The initial decision parameter is given by:

$$p_0 = 2\Delta y - \Delta x. \tag{2.2}$$

This decision parameter is used to determine whether the next pixel will be directly to the right of the current pixel (horizontal step) or one pixel up and one pixel to the right (diagonal step).

2.4.2 Step 2: Iterative Process

Starting from the initial pixel (x_0, y_0), the algorithm determines the next pixel to plot by checking the value of the decision parameter p_k, which is updated at each step. If p_k is less than zero, the next pixel is chosen horizontally (i.e., one step to the right). If p_k is greater than or equal to zero, the next pixel is chosen diagonally (i.e., one step up and to the right).

The decision parameter is updated as follows:

$$p_{k+1} = p_k + 2\Delta y, \quad \text{if} \quad p_k < 0, \tag{2.3}$$

$$p_{k+1} = p_k + 2(\Delta y - \Delta x), \quad \text{if} \quad p_k \geq 0. \tag{2.4}$$

This process continues until the endpoint (x_1, y_1) is reached.

2.4.3 Step 3: Symmetry Considerations

Bresenham's algorithm is symmetric, meaning that the same principles can be applied to lines with different slopes (positive, negative, shallow, or steep). Depending on the values of Δx and Δy, the algorithm can be adapted to handle lines that go in different directions:

For lines with a slope less than 1 (shallow lines), the algorithm prioritizes horizontal movement, and the decision parameter is updated based on Δx. For lines with a slope greater than 1 (steep lines), the algorithm prioritizes vertical movement, and the decision parameter is updated based on Δy.

To handle these cases, the algorithm swaps Δx and Δy accordingly, ensuring that the process works for all types of lines.

2.4.4 Step 4: Algorithm Termination

The algorithm continues the iterative process until it reaches the final point (x_1, y_1). At each step, it selects the appropriate pixel based on the decision parameter and updates the decision parameter according to the formula provided. Once the final point is reached, the line has been successfully drawn between the two given points.

The final output is a set of pixels that approximate a straight line between the points (x_0, y_0) and (x_1, y_1). The simplicity and efficiency of the algorithm make it an excellent choice for real-time line drawing in raster-based graphics systems.

2.4.5 Example

To illustrate the algorithm, let's consider an example where the starting point is $(x_0, y_0) = (2, 3)$ and the ending point is $(x_1, y_1) = (10, 7)$. The steps are as follows:

- Initialize: $\Delta x = 10 - 2 = 8$, $\Delta y = 7 - 3 = 4$, and $p_0 = 2 \times 4 - 8 = 0$.
- At each step, update the decision parameter and choose the appropriate pixel (either horizontal or diagonal).
- Continue this process until the line reaches the endpoint $(10, 7)$.

This approach ensures that the line is drawn efficiently, pixel by pixel, while minimizing floating-point operations.

2.4.6 Advantages of Bresenham's Algorithm

Bresenham's Algorithm offers several advantages over other line-drawing algorithms:

- Efficiency: The algorithm uses only integer arithmetic, making it very fast and suitable for hardware implementation.

- Accuracy: It produces a visually accurate approximation of a straight line.
- Simplicity: The algorithm is simple to implement and has a low computational cost.

These advantages have made Bresenham's Algorithm widely used in computer graphics, especially in raster graphics editors like in Figure 2.3, rendering engines, and real-time graphics applications [49].

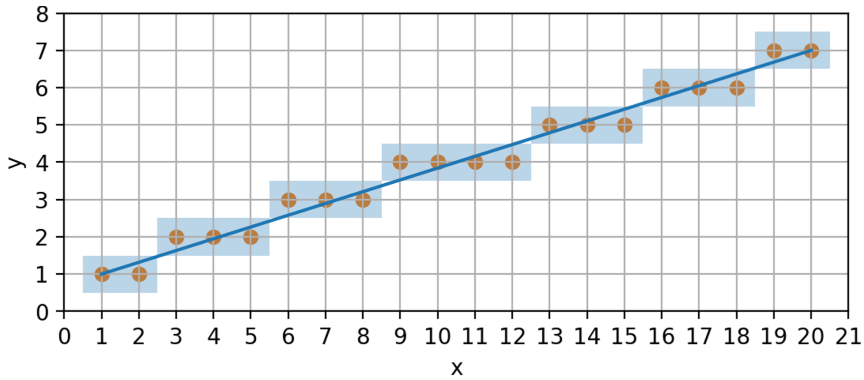


Figure 2.3: A graphical representation of the line drawn using Bresenham's Algorithm [49].

2.5 MTs in Persian Classical Music

MTs are intervals smaller than a semitone and play a significant role in Persian classical music. Persian music scales often include intervals that fall between the notes of the Western chromatic scale, creating a distinct sound that is not typically present in Western music. These MTs are an essential part of the Persian musical system, and their use enables composers and performers to create more subtle tonal variations, adding layers of emotional depth and nuance to the music. MTs provide the musician with the ability to explore fine gradations of pitch, which allows for a richer and more varied emotional expression. The flexibility of MTs opens up a broader spectrum of tonal possibilities, giving the music a unique character that can be both intricate and profoundly expressive.

In Persian classical music, the term "Dastgah" refers to a system of modes, each characterized by a particular arrangement of intervals that define the mood and atmosphere of the music. These Dastgah are the building blocks of Persian music and serve as a guide for improvisation and composition. Each Dastgah contains a series of MTs, such as quarter tones and other intervals that lie between the standard Western pitches. The use of these MTs allows Persian music to achieve subtle tonal variations that evoke a range of emotions, from serenity to intensity, making the music deeply emotional and engaging. These microtonal shifts create a sense of fluidity in the tonal landscape, where the pitch can gently bend and flow, enhancing the expressive qualities of the performance.

The "Shahed" (evident note) plays a pivotal role in many Dastgah, serving as a reference note around which the melody revolves. This note acts as an anchor point in the tonal structure of the piece, providing both stability and direction. The Shahed is crucial for defining the emotional and tonal center of the piece, guiding the listener's emotional journey and providing a sense of resolution when the melody returns to it. The role of the Shahed is central to the overall expressiveness of the music, as it helps to shape the mood and character of the performance. It is around this note that the performer improvises, creating variations and embellishments that highlight the unique characteristics of the Dastgah [50].

In music, the central note of a melody is represented by its scale degree. For example, in Figure 2.4, the second degree of the A minor scale is shown as the central note. The second degree of the A minor scale is the note B. Therefore, the scale degree chosen as the central note is determined based on the tonic of the scale.

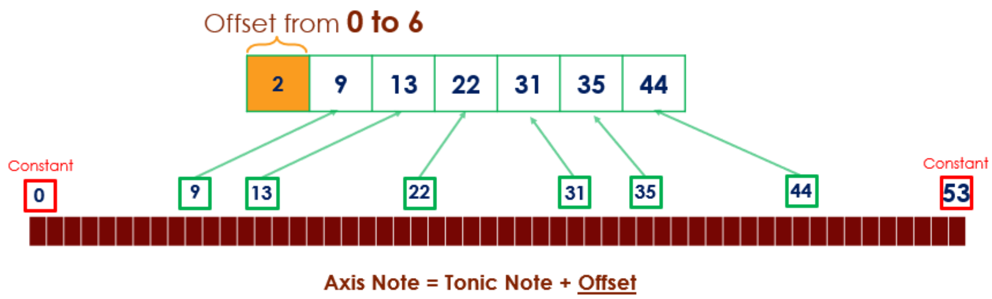


Figure 2.4: The "Shahed" (axis) note in Persian music.

The presence of MTs in Persian music challenges the standard Western theory of modulations, which is based on the assumption of equal-tempered tuning and fixed intervals between notes. Traditional methods of modulation used in Western music, such as axial chords and common-note modulations, rely on the concept of fixed intervals between notes. However, the MTs in Persian music do not fit neatly into the Western system of tuning, making it difficult to apply these standard techniques directly. As such, understanding the nature of MTs and how they influence modulations in Persian classical music is crucial for developing a model that can effectively handle modulations between Persian and Western music scales. This requires a more flexible approach that accounts for the unique characteristics of MTs and their role in shaping the emotional and tonal qualities of Persian music. A deeper understanding of how MTs function in modulations will provide valuable insights into how Persian and Western music can interact and how modulations can occur smoothly between the two systems.

By recognizing the importance of MTs in Persian classical music, this research seeks to bridge the gap between Western and Persian music theory, developing a framework that respects the nuances of both traditions. The study aims to provide a comprehensive understanding of how microtonal shifts impact the modulations process and how they can be incorporated into a unified model for analyzing and generating MP across different musical systems.

2.6 SOM Algorithm

The SOM algorithm, developed by Teuvo Kohonen in the 1980s, is an unsupervised learning method used for clustering and dimensionality reduction. SOM maps high-dimensional data into a lower-dimensional grid while preserving topological relationships. This section will describe the SOM algorithm step by step, with mathematical formulas and detailed explanations of each phase of the algorithm. The details of the algorithm, as well as its application in various fields, can be found in the work of Kohonen and other studies [51].

2.6.1 Step 1: Initialization

The first step in the SOM algorithm is to initialize the weights of the grid nodes. Each node in the grid has an associated weight vector, which has the same dimension as the input data vectors. The weight vectors are typically initialized with small random values. Let the weight vector of node i at time $t = 0$ be denoted as:

$$\mathbf{w_i}(0) = \text{random initialization}, \tag{2.5}$$

where $\mathbf{w_i}(0)$ is the weight vector for node i, and the initialization values are chosen randomly from a small range.

2.6.2 Step 2: Competition Phase

Once the weights are initialized, the next step is to determine the Best Matching Unit (BMU). The BMU is the node whose weight vector is closest to the input vector. The distance between the input vector $\mathbf{x}$ and each node's weight vector $\mathbf{w_i}$ is computed, typically using the Euclidean distance:

$$d(\mathbf{x}, \mathbf{w_i}) = \|\mathbf{x} - \mathbf{w_i}\|_2, \tag{2.6}$$

where $\mathbf{x}$ is the input vector, $\mathbf{w_i}$ is the weight vector of node i, and $d(\mathbf{x}, \mathbf{w_i})$ is the Euclidean distance between them. The BMU $\mathbf{w_{bm}}$ is the node that minimizes this distance:

$$\mathbf{w_{bm}} = \arg\min_i d(\mathbf{x}, \mathbf{w_i}). \tag{2.7}$$

2.6.3 Step 3: Adaptation Phase

After identifying the BMU, the next step is to update the weight vectors of the BMU and its neighboring nodes to make them more similar to the input vector. The update rule for the weights is given by:

$$\mathbf{w_i}(t+1) = \mathbf{w_i}(t) + \eta(t)h_{i,bm}(t)(\mathbf{x} - \mathbf{w_i}(t)), \tag{2.8}$$

where:

- $\mathbf{w_i}(t)$ is the weight vector of node i at time t.
- $\mathbf{x}$ is the input vector.
- $\eta(t)$ is the Learning Rate (LR) at time t.
- $h_{i,bm}(t)$ is the neighborhood function that determines the influence of the BMU on neighboring nodes.
- t is the current iteration of the algorithm.

The neighborhood function $h_{i,bm}(t)$ typically decreases over time, meaning that the influence of the BMU on neighboring nodes becomes smaller as the algorithm progresses. A common form of $h_{i,bm}(t)$ is:

$$h_{i,bm}(t) = \exp\left(-\frac{d(i, bm)^2}{2\sigma(t)^2}\right), \tag{2.9}$$

where $d(i, bm)$ is the distance between node i and the BMU, and $\sigma(t)$ is the radius of the neighborhood, which also decreases over time.

Figure 2.5 illustrates a graphical representation of sigma (neighborhood radius) and the LR, showing how updates affect neighboring nodes based on their distance to the center.

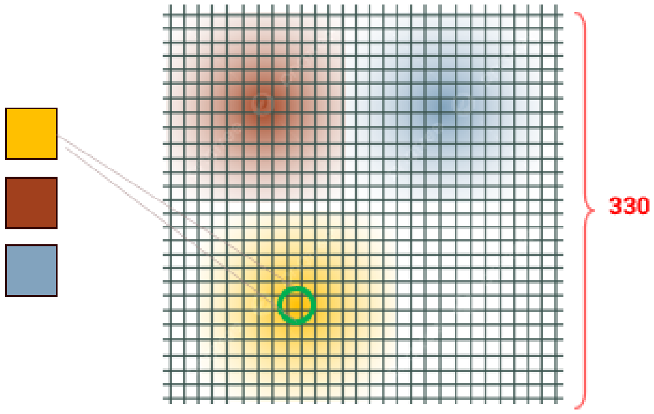


Figure 2.5: A graphical representation of sigma (neighborhood radius) and LR (the effect of updates on neighboring nodes based on distance to the center).

2.6.4 Step 4: Iteration and Convergence

The SOM algorithm proceeds iteratively. In each iteration, a new input vector $\mathbf{x}$ is presented to the network, and the BMU and weight updates are computed. Over time, the LR $\eta(t)$ and the neighborhood radius $\sigma(t)$ decrease, and the map becomes more refined. The LR typically follows a schedule such as:

$$\eta(t) = \eta_0 \exp\left(-\frac{t}{T}\right), \tag{2.10}$$

where η_0 is the initial LR, t is the current time step, and T is the total number of iterations.

The algorithm continues until the weight vectors converge, meaning that the positions of the nodes no longer change significantly with additional iterations.

2.6.5 Step 5: Final Map

Once the algorithm converges, the nodes in the network represent similar clusters. The input data points of the final map represent the underlying structure of the data. Similar data points are grouped together in the same or adjacent nodes. The SOM map can be visualized as a two-dimensional grid, where each node represents a cluster of input vectors. The organization of the network reflects the relationships between data points, with similar data points mapped to adjacent nodes.

SOM algorithms are capable of discovering complex structures in the data, and the final map provides a clear visual representation of these structures, as illustrated in Figure 2.6. This approach has been successfully applied to various fields, including music theory, where SOM have been used to model the relationships between musical scales, harmony, and modulation pathways, as discussed in [51].

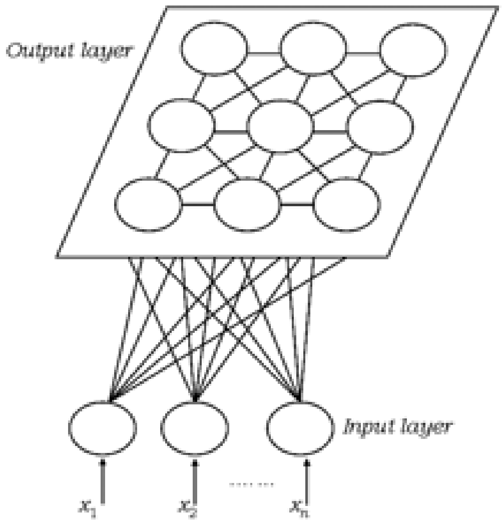


Figure 2.6: A schematic representation of the SOM network [51].

In the context of music theory, SOM have been used to model various aspects of musical structure, such as harmony, tonality, and modulation. By clustering musical scales based on their intervallic structures, SOM can reveal relationships between different musical modes and help identify smooth modulation pathways. This approach allows for a deeper understanding of how different scales are related and how modulations between them can occur seamlessly [51].

Chapter 3

Literature Review

The application of SOM in music theory has been extensively explored in various studies, primarily focusing on Western music scales and key modulations. However, limited research has addressed the complexities of microtonal variations and the modulation processes in Persian classical music. This chapter compares existing frameworks with our proposed approach, highlighting how our method addresses key limitations in previous work. Additionally, this chapter expands the discussion with more references and detailed comparisons to meet the required depth.

3.1 Existing Frameworks for SOM in Music Theory

Several researchers have employed SOM for analyzing key modulations and scale relationships in Western music. For example, the solution proposed in [52] successfully clustered Western musical scales based on interval patterns and tonal characteristics. However, their framework lacks the capability to incorporate microtonal intervals, which are essential in Persian classical music. Our proposed framework overcomes this limitation by integrating microtonal variations into the clustering process, providing a more comprehensive analysis that respects the tonal nuances of Persian music. Moreover, unlike previous works that focus solely on tonal relationships, our approach captures the subtleties of pitch differences, which are crucial in shaping the emotional content of Persian music.

The study in [53] applied SOM to analyze modulation pathways between major and minor scales. While the research demonstrated smooth transitions between Western scales, it did not address the complexity of quarter-tone adjustments found in Persian music. The solution proposed in [53] lacks the flexibility required to handle these variations. Our approach extends this work by including quarter-tone variations, allowing for the generation of modulation pathways that align with the unique tonal characteristics of Persian classical music. This makes our framework suitable for applications in Persian music, where microtonal nuances play a significant role in defining the emotional tone of a performance.

Additionally, previous works such as [54] have focused on hierarchical clustering methods for scale analysis. However, these methods often fail to reveal emergent patterns that are not predefined by the hierarchical structure. Our SOM-based framework overcomes this limitation by allowing previously unexplored patterns to emerge during the clustering process, thereby offering novel insights into modulation pathways that may not have been discovered using traditional clustering techniques.

3.2 Addressing Microtonal Variations

Traditional modulation techniques often overlook microtonal intervals, making them unsuitable for Persian music. The approach in [55] attempted to consider microtonal variations but lacked an efficient computational framework for practical application. The solution proposed in [55] lacks real-time analysis capabilities and cannot dynamically adjust to different microtonal contexts. In contrast, our framework employs Bresenham's Algorithm to optimize modulation pathways, ensuring computational efficiency while maintaining an intuitive auditory experience. This advancement allows for the practical implementation of microtonal modulations, bridging a significant gap in the literature.

Furthermore, while previous research often treats microtonal variations as peripheral, our framework places them at the core of the modulation analysis. By systematically incorporating quarter-tone and semi-tonic adjustments into the SOM clustering process, we ensure that these essential components of Persian music are accurately represented and analyzed. Our method also supports adaptive learning, where the SOM continues to refine its clusters based on new input data, thereby maintaining relevance as new musical scales and compositions are introduced.

The use of microtonal notes is not limited to Persian and Eastern music. Western instruments such as the violin also possess the capability to perform these notes. This is because the violin and similar instruments lack keys or frets, enabling continuous frequency changes between two notes. This unique characteristic allows for seamless transitions and microtonal variations that are essential for expressing subtle musical nuances [46].

As illustrated in Figure 3.1, a comparison between the piano and the violin highlights how microtonal notes can be performed. Unlike the piano, which is restricted to discrete semitones, the violin allows for smooth frequency shifts, making microtonal performance feasible. It would be advantageous for contemporary music to incorporate these frequencies more extensively, offering greater diversity and richness in musical expression.

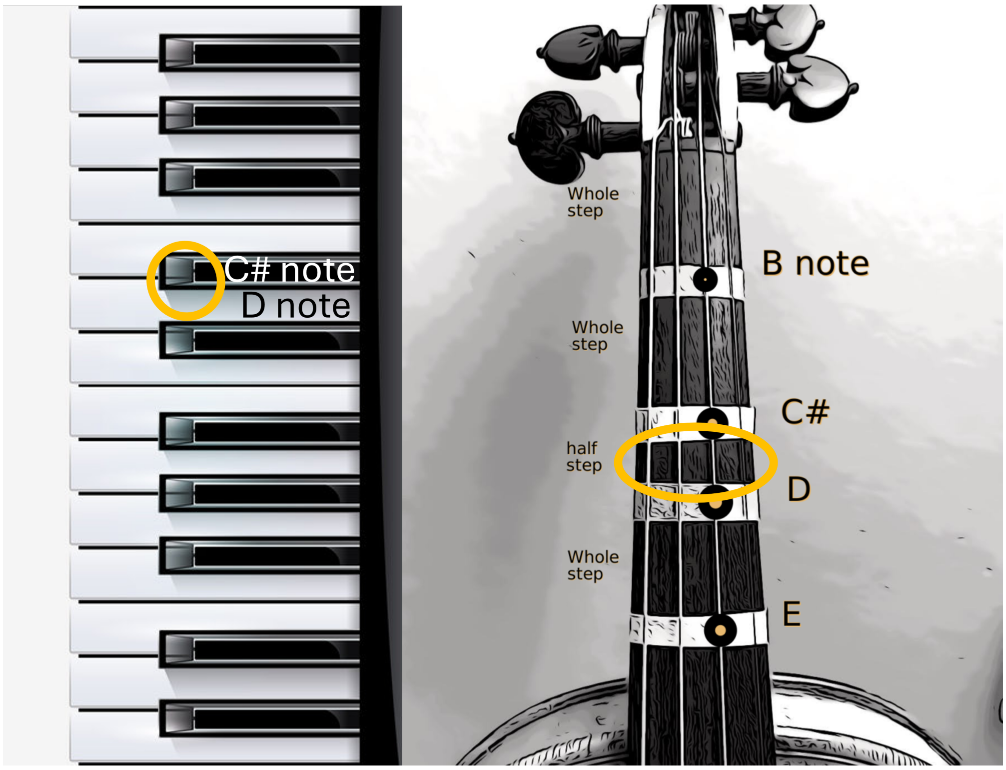


Figure 3.1: The capability of performing microtonal notes in some Western instruments, such as the violin, compared to the piano [56].

3.3 Emotional Anchors and Shahed Notes

Previous models, such as those discussed in [57], have explored the relationships between different Dastgahs in Persian music. However, these studies did not systematically consider the role of Shahed notes as emotional anchors in the modulation process. The solution proposed in [57] lacks a structured mechanism for evaluating how Shahed notes influence modulation pathways. Our proposed framework addresses this gap by incorporating Shahed note analysis, ensuring that the modulation pathways reflect both theoretical transitions and their corresponding emotional impacts.

Our approach differs significantly from previous works by treating Shahed notes not only as structural elements but also as central emotional anchors that influence the listener's perception of continuity during modulation. By assigning weight factors to these notes during the clustering process, we ensure that the generated modulation pathways align more closely with traditional performance practices, offering a more nuanced understanding of modulation.

3.4 Real-Time Auditory Validation

While prior research has focused on theoretical analyses of modulation pathways, few studies have provided real-time auditory validation. For instance, the framework in [58] mapped relationships between scales but did not offer practical tools for auditory evaluation. The solution proposed in [58] lacks dynamic interaction capabilities that allow composers to hear and adjust modulation pathways in real time. Our research advances this area by developing a melody generator using Python's Mido library [59], translating modulation pathways into audible transitions. This feature allows composers to evaluate modulation pathways in real time, validating theoretical predictions through auditory feedback.

Additionally, our framework incorporates user-centric features such as tempo and rhythm adjustments during real-time playback. This provides a more comprehensive auditory experience and allows for further customization based on individual compositional needs. Our approach ensures that modulation pathways are not only theoretically sound but also practically useful in live performance and composition contexts.

3.5 Comprehensive Comparison with Existing Frameworks

To summarize, the key limitations identified in previous research include:

- **Lack of integration of microtonal variations** ([52], [55]): Previous works did not incorporate microtonal intervals critical for Persian music.
- **Absence of efficient computational frameworks** ([55]): Earlier frameworks lacked optimization techniques like Bresenham's Algorithm.
- **Neglect of Shahed notes as emotional anchors** ([57]): Existing studies failed to highlight the emotional role of Shahed notes in modulation.
- **Lack of real-time auditory validation tools** ([58]): Prior research did not provide interactive, real-time feedback mechanisms.
- **Limited adaptability to new musical scales** ([54]): Hierarchical clustering approaches were inflexible, preventing the discovery of new modulation patterns.

Our proposed framework addresses these limitations by:

- Integrating microtonal variations into the clustering process.

- Optimizing modulation pathways using Bresenham's Algorithm for computational efficiency.
- Incorporating Shahed note analysis for emotional continuity.
- Developing a melody generator for real-time auditory validation.
- Supporting adaptive learning to accommodate new musical scales and compositions.

These contributions demonstrate the novelty and effectiveness of our approach, bridging the gap between traditional music theory and algorithmic analysis in the context of Persian classical music. By addressing the gaps left by existing research, our framework offers a more comprehensive, adaptable, and practical solution for exploring modulation pathways, especially in microtonal musical systems.

Chapter 4

Methodology

This chapter outlines the research methodology used to generate smooth MPs between various musical scales, particularly in the context of Western and Persian music, which includes microtonal intervals. The aim is to develop an effective computational model that enables smooth transitions between scales while retaining the emotional and tonal integrity of the music. The methodology combines music theory, computational techniques, and machine learning algorithms, particularly the SOM, to achieve this goal. By using SOM, the research seeks to uncover hidden patterns in the relationships between different scales and to model the subtle nuances that define key modulation. The approach is designed to provide a comprehensive understanding of the underlying structure of musical scales, offering insights into how different tonalities relate to one another and how they can transition smoothly within a musical context.

The chapter discusses the steps taken for data collection, design, and implementation of the SOM algorithm, as well as the generation and evaluation of MPs. The data collection process involves gathering a variety of musical scales from both Western and Persian traditions, paying particular attention to the microtonal intervals that distinguish Persian music. These scales are carefully analyzed to identify the key intervals and tonal relationships that will inform the creation of MPs. The SOM algorithm is then applied to map these scales onto a two-dimensional grid, where the proximity of scales on the map reflects their relative tonal similarity. The goal is to produce a map that accurately represents the relationships between scales, enabling the identification of smooth MPs that transition from one key to another without disrupting the musical flow. Once the MPs are generated, they are evaluated through both computational and qualitative methods to assess their emotional impact and their effectiveness in maintaining musical coherence. The chapter also discusses the criteria used for evaluating the MPs, including the smoothness of the transitions, the emotional tone of the music, and the perceptual ease with which listeners can follow the modulation.

Through this methodology, the research aims to contribute to a deeper understanding of how musical scales, both traditional and microtonal, can be integrated into a computational model for key modulation. By employing SOM, this study explores the potential of machine learning algorithms to enhance music theory and composition, offering new tools for composers, musicians, and researchers to explore key transitions and MPs in a wide range of musical traditions. The methodology presented in this chapter provides the foundation for the subsequent chapters, which will detail the implementation and analysis of the model, as well as the results of the MP evaluations.

4.1 Data Collection

Data collection is the basis of the methodology, as it serves as input for the computational model. In this study, the data represent interval patterns from both Iranian and Western classical music scales. The initial focus is on capturing the relationships between musical notes in these scales, including fine-grained differences in pitch for microtonal scales that are meaningful in Persian music. This nuanced approach is crucial to understand the subtle tonal shifts that occur in Persian music and to incorporate these into the MPs between scales.

The dataset used for training the SOM includes a variety of scales, each with its own specific interval pattern. These scales are carefully selected to represent a broad spectrum of musical traditions, particularly from both Persian and Western classical music [29]. The scales selected for the study include:

- **Major Scale (Mahur)**: This scale follows the common Western pattern of major scales, which serve as the cornerstone of Western tonal music theory.

- **Melodic Minor Scale**: This scale has distinct ascending and descending forms and is used in both Western and Iranian music, highlighting its flexibility and adaptability across cultures.

- **Harmonic Minor Scale**: Known for its strange sound, especially in Persian music, this scale has an amplified seventh interval that gives it a dramatic, otherworldly quality.

- **Natural Minor Scale**: This scale follows the diatonic pattern common in Western music and provides a basic basis for comparison with other scales.

- **Homayoun Mode**: One of the important modes in Iranian classical music, known for its distinct emotional quality, often evoking feelings of peace and nostalgia.

- **Shur Mode**: This Persian mode is characterized by a mixture of minor and major intervals and is often used to express melancholy, creating a complex emotional tone.

- **Chahargah Mode**: This mode from Persian classical music evokes a heroic and dramatic tone, frequently used to convey strength, courage, and passion.

- **Segah Mode**: A mode used in Persian classical music that has a unique combination of intervals, creating a mystical and solemn atmosphere, often associated with spiritual or introspective themes.

This information has been derived from the study of [29]. Figure 4.1 illustrates that the collected data consists of a list of nodes, where each node is itself a list. The first index of each node represents the root note, while the remaining items correspond to the step degrees. In this figure, the major scale pattern is depicted, referring to seven root notes. The collected data list can follow this same structure for all musical scales. Figure 4.1 illustrates that the collected data consists of a list of nodes, where each node is itself a list. The first index of each node represents the root note, while the remaining items correspond to the step degrees. In this figure, the major scale pattern is depicted, referring to seven root notes. The collected data list can follow this same structure for all musical scales.

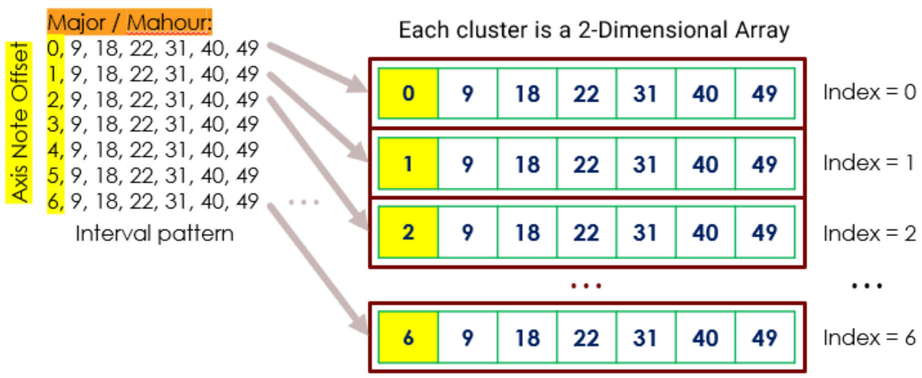


Figure 4.1: A set of seven Major/Mahour scales with different axis notes.

Each scale consists of seven notes and the corresponding intervals that define the relationship between the notes. These intervals can be derived through analysis of both Western and Persian musical structures. In Persian music, these intervals can include microtones (e.g., quarter tones or other subtle pitch variations) which are essential for creating the distinctive sound of Persian music. These microtonal variations often form the basis of the emotional expressiveness in Persian compositions. The inclusion of microtones in this study is critical for representing the fine distinctions in pitch that shape the musical experience in both cultures [57].

In this study, each note is represented numerically, where its pitch is denoted relative to a starting tonic note, known as the axis note (shahed), which serves as the tonal center for the scale. This numerical representation allows for precise analysis and manipulation of the data. For example, a scale like Mahour may have a tonic note at C, and the subsequent notes would be derived based on their intervals from the tonic. The microtonal variations are included by representing these intervals with higher precision (e.g., quarter tones, half steps, and other finer gradations of pitch), thus providing a more accurate mapping of musical intervals for computational analysis [60].

This data was collected from various musical theory sources, including both Western and Persian classical music literature, as well as contemporary analyses of microtonal music systems. The data collection process is designed to be exhaustive and comprehensive, capturing the full range of scales and modes used across both musical traditions. The collected data is then encoded into a numerical format suitable for processing by the SOM algorithm. This encoding process ensures that the microtonal intervals are preserved and can be effectively used in the generation of MPs [55].

The contents of each item in the node list are shown in Figure 4.2. The first item represents the scale degree, where the axis note defines the melody.

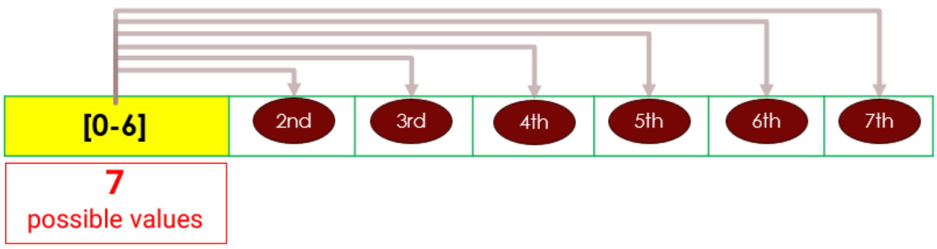


Figure 4.2: Structure of the node list, illustrating how the axis note determines the melody based on scale degrees.

Since each musical note, considering microtones, can exist in five different forms (sharp, flat, natural, koron, and sori), Figure 4.3 illustrates that each item in the node representing scale degrees can have five possible variations.

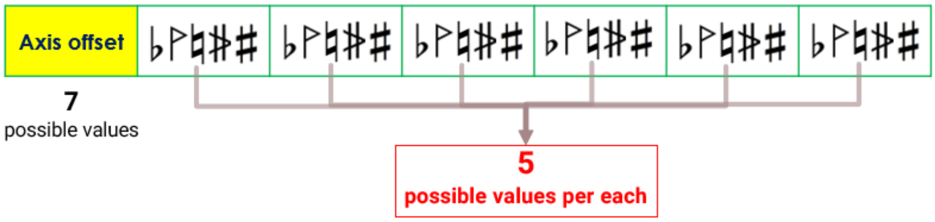


Figure 4.3: The range of values for each item in the collected data, considering the microtones used in Persian music.

Image 4.4 illustrates how the different states of each note can be represented using numbers. These numbers correspond to the position of each note within the list that defines the range of comma counts in a scale.

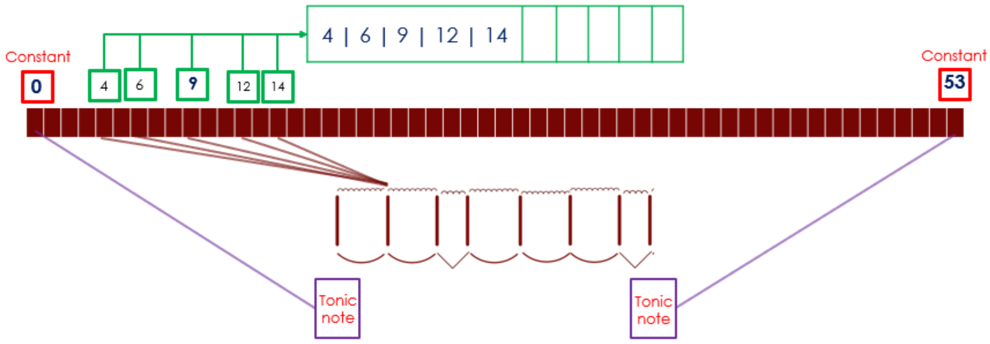


Figure 4.4: Representation of first degree's possible positions in a scale using index numbers based on the total count of commas in the scale.

This expanded data set is essential for the model, as it allows for the generation of modulation paths that respect both the microtonal subtleties of Persian music and the broader structural relationships in Western music. With this rich set of data, the SOM will be able to generate nuanced and accurate paths for modulation between scales, accounting for both the emotional and technical aspects of musical transitions.

4.2 Designing the SOM

The design process for the SOM involved several key components, which were critical to ensuring the model effectively captured the complexities of musical scales, including their MT features. The steps are as follows: [52]

- **Grid Dimensions:** The SOM was initialized with a two-dimensional grid of nodes. Each node in the grid represents a cluster of similar scales, and its position reflects the relationship of scales with one another based on their interval patterns. The grid's size was carefully chosen after testing several configurations, balancing computational efficiency with the model's ability to capture complex relationships between scales. A larger grid allowed for more detailed clustering, but required more computational power, while a smaller grid could lead to oversimplification. Therefore, finding the optimal size for the grid was a key design decision.

  Figure 4.5 demonstrates the initial estimation of the map dimensions by counting the possible nodes based on the permutations of different states of each node's items.

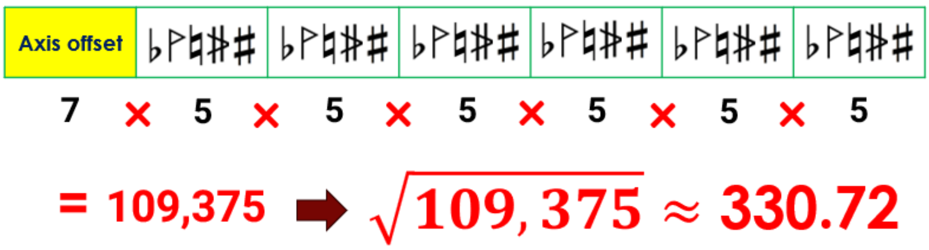


Figure 4.5: An initial estimate of the map dimensions required for the SOM algorithm.

- **Data Representation:** Each musical scale was represented as a vector The numerical values of these values correspond to the pitch relationships between notes, with particular attention to the inclusion of microtonal intervals. The minor tones of Iranian music, which are smaller than Western semitones are particularly important to include, as they form a central aspect of the Persian musical system. The notes within each scale were encoded as a set of features, where each feature indicates the relationship of a note to the tonic (axis note) and other notes in the scale. This numerical representation enabled the SOM to analyze and process the data efficiently, while still preserving the intricate details of each scale's structure.

  Since the product of possible states for each node is 109,375 and the nodes are supposed to be adjacent to each other in a two-dimensional grid, the potential dimensions of our map at its largest will be the square root of this number. As shown in Figure 4.6, each node is 7-dimensional, and the entire map system is two-dimensional, with the adjacency of the nodes determining the layout.

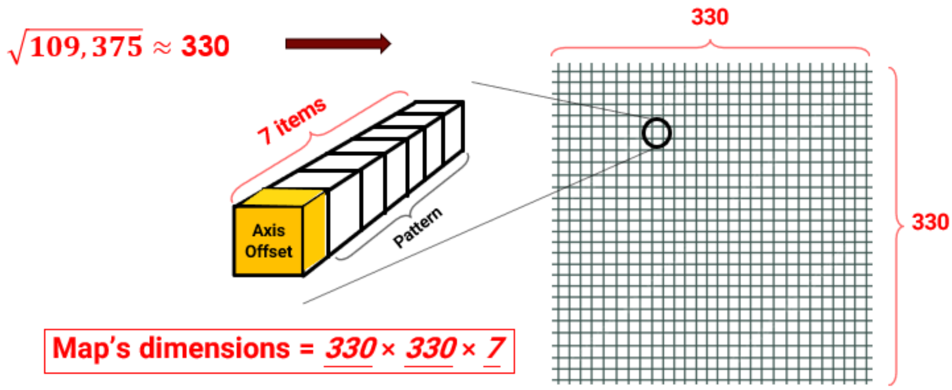


Figure 4.6: Representation of the 7-dimensional nodes within a 2-dimensional map, where the layout is determined by the adjacency of the nodes.

- **Initial Weights:** The initial weights for each node in the SOM grid were set randomly, but within a carefully defined range that reflected the potential values of musical intervals. These intervals were chosen to encompass the full range of possible pitch relationships found in both Western and Persian musical systems. Western music typically operates within a twelve-tone temperament framework, in which basic intervals consisting of whole steps and semitones form the basis of most musical structures. In contrast, Persian music, with its deep history of microtonality, utilizes intervals that fall between Western semitones, resulting in a unique system of microtonal steps. The inclusion of these microtonal intervals in the initial range of weights was essential for ensuring that the SOM could accommodate the complexities of Persian music, which features a rich variety of intervals that are not present in traditional Western music.

  To ensure that the SOM was capable of effectively learning and clustering scales from both traditions, the range of initial weights was chosen to reflect this duality. The weight of each node in the network therefore corresponded to the potential pitch distances present in the scales analyzed, ensuring that nodes could represent different relationships between internal substructures of both Western and Persian scales. By initializing the weights in this way, we provided the SOM with a starting point that was both flexible and capable of adapting to the diverse pitch intervals that would be encountered during the training process. This flexibility was crucial in allowing the SOM to learn meaningful representations of the different scales, facilitating the clustering of related scales and the generation of smooth transitions between them.

  As the SOM was trained on the musical data, the random initialization of weights played an important role in allowing the algorithm to explore the full range of possible pitch intervals. The initial weights provided a foundation for the learning process, ensuring that the system did not favor any specific set of intervals over another. Instead, the weights were adjusted gradually through the training process, with the system refining its understanding of musical structures and the relationships between scales. This adaptive process allowed the SOM to generate clusters of scales that reflected the musical logic and structure of both Western and Persian traditions, while also respecting the MTs nuances that characterize Persian music.

Since each scale degree has only 5 different frequency states and we have converted these states into numbers, we can generate random nodes. For example, as shown in Figure 4.7, the possible numerical values for the fifth degree of each scale are 26, 28, 31, 34, and 36. We repeat this process to create all possible nodes. To assist with the learning process, nodes that share the same scale but have different root notes are created consecutively so that, in the initial map, these nodes are deliberately placed as neighbors. This helps accelerate learning and facilitates greater interaction during the process.

The randomness of the initial weights also helped prevent the SOM from converging prematurely to a suboptimal solution. By starting with random values within the defined range, the SOM had the opportunity to explore different configurations of pitch intervals and scale structures. This exploration was essential for ensuring that the system could generate meaningful and musically coherent results, rather than becoming trapped in a local optimum that might not reflect the true diversity of musical scales. Over time, as the training process progressed, the weights were adjusted in such a way that they corresponded more closely to the actual pitch intervals and scale structures in the training data, leading to the formation of meaningful clusters that represented different musical scales.

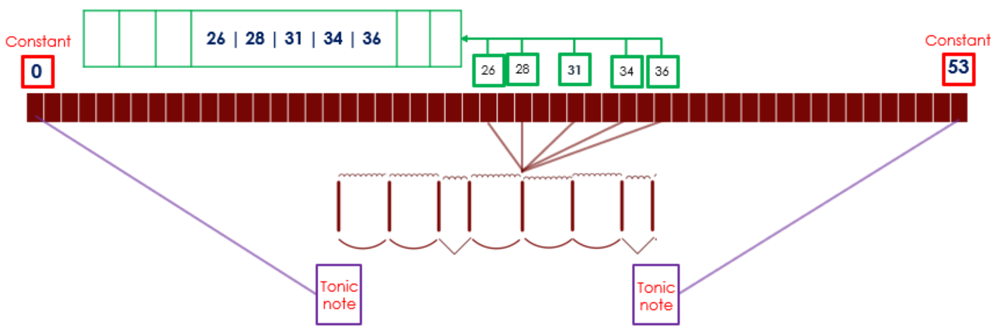


Figure 4.7: An example of numerical values for the fifth scale degree, showing how random nodes are generated from frequency states.

- **Normalization:** Normalizing the data is crucial in this study because the SOM algorithm uses a distance calculation between two vectors, and unnormalized data can bias the results toward larger values and the seventh degree notes of the scale. Since each degree in the vector of each node has specific values and only 5 possible states, we normalize these values using a formula that changes the minimum value to 0 and the maximum value to 1:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{4.1}$$

  where x is the original value, $x_{\min}$ is the minimum value in the vector, and $x_{\max}$ is the maximum value in the vector.

  Thus, in the end, all items of each node have values between 0 and 1, and each value represents a specific state of the degree's position of the note within the list of commas in the musical scale.

- **Identifying the BMU:** The process of identifying the BMU and updating the weights of the BMU and its neighboring nodes continued until the SOM had sufficiently learned to represent the relationships between the musical scales. This involved multiple cycles of training, with the SOM being exposed to various input scales in each cycle. As the training progressed, the weights of the nodes converged, and the SOM produced a map that meaningfully clustered the musical scales based on their interval patterns [61]. The resulting map provided an effective representation of the musical scales, which could then be used to generate modulation paths between them. These modulation paths, based on the clustering and relationships learned by the SOM, allowed for the generation of smooth transitions between scales, respecting both the MT intervals of Persian music and the traditional intervals of Western music.

  Figure 4.8 shows that the system finds the BMU by calculating the Euclidean distance between two 7-dimensional nodes.

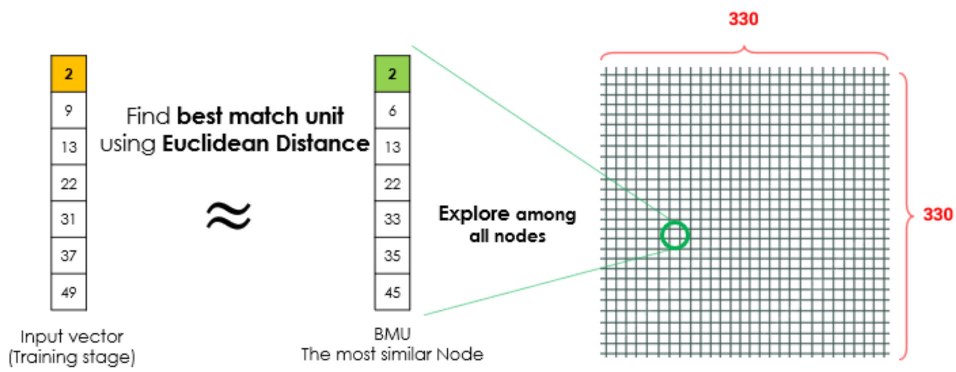


Figure 4.8: Identifying the Best Match Unit according to the input.

Also, Figure 4.9 shows that to update the BMU, all 7 dimensions of the nodes correspondingly update based on the 7 items of the input node.

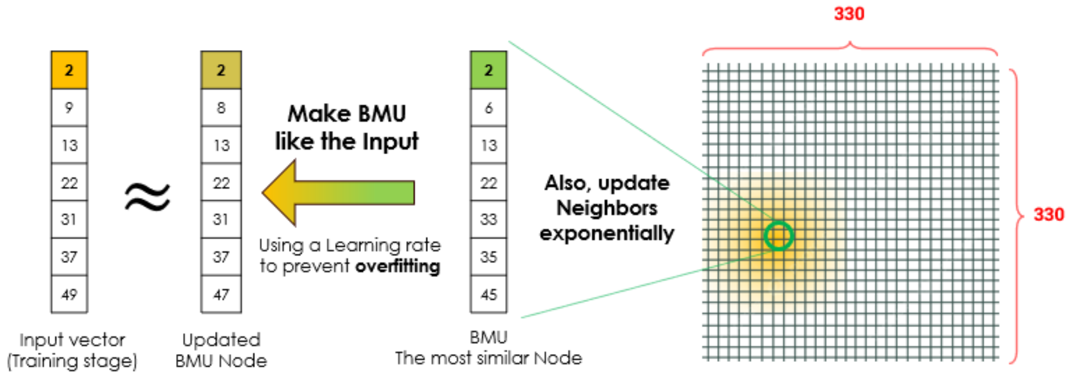


Figure 4.9: The BMU is selected based on the smallest distance, and its weights, along with its neighbors, are adjusted.

- **Identifying Neighborhood:** The neighborhood defines how the weights of nodes are updated during each iteration. Initially, the neighborhood function is wide, meaning that many neighboring nodes are adjusted in response to each training step. As training progresses, the neighborhood function gradually narrows, so only the BMU and a small number of its nearest neighbors are updated. This ensures that the SOM begins with a broad consideration of scale relationships and then refines its clustering as the training continues [62]. The gradual narrowing of the neighborhood radius allows the map to fine-tune its representation of the data over time, improving the clustering accuracy and the quality of modulation paths [54]. By adjusting the neighborhood function in this way, the SOM algorithm can adapt to the inherent structure of musical scales, facilitating a more accurate representation of the relationships between scales and their potential transitions. SOMs typically consider node neighborhoods either as rectangular grids or hexagonal (octagonal) grids. Our choice is to use eight neighbors for each node, providing more connections between nodes. This is because each node has 7 dimensions, and musical scales may have several neighboring steps.

Figure 4.10 illustrates how, in our proposed model, the neighbors of each node are related based on their indices in the neuron list.

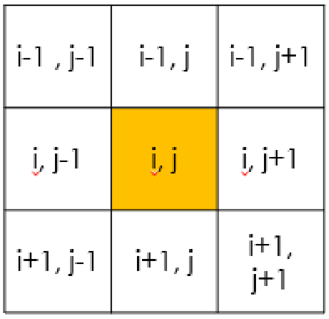


Figure 4.10: In this method, 8 neighboring nodes are considered for each node.

- **Convergence Criteria:** Training continues until the algorithm converges, meaning that the weights of the nodes no longer change significantly. Several factors influence convergence in the SOM algorithm:
  - **Decay of the Learning Rate:** During training, the LR, denoted as $\eta(t)$, gradually decreases over time. This ensures that as the algorithm progresses, the updates to the node weights become smaller, allowing the map to stabilize. A common approach is to update the LR as:

    $$\eta(t) = \eta_0 \cdot \exp\left(-\frac{t}{\tau}\right) \tag{4.2}$$

    where η_0 is the initial LR, t is the number of iterations, and τ is a constant that controls the rate of decay. As $\eta(t)$ approaches zero, the updates to the weights become negligible, signaling convergence.
  - **Decay of the Neighborhood Radius:** Similar to the LR, the neighborhood radius $\sigma(t)$ also decays over time. This means that in the beginning, the weight updates affect a large group of neighboring nodes, but as training progresses, only nodes that are very close to the BMU are updated. A typical update rule is:

    $$\sigma(t) = \sigma_0 \cdot \exp\left(-\frac{t}{\tau_\sigma}\right) \tag{4.3}$$

    where σ_0 is the initial neighborhood radius and τ_σ is the time constant that controls the decay rate.
  - **No Significant Change in Weights:** Convergence is also identified when the changes in the weights of the nodes fall below a certain threshold. In other words, after a certain number of iterations, the weights of the nodes are considered stable, and further updates are minimal. The

convergence criterion can be expressed as:

$$\sum_{i=1}^{N} |\Delta \mathbf{w}_i(t)| < \epsilon \tag{4.4}$$

where $\Delta \mathbf{w}_i(t)$ is the weight change of the i-th node at time t, and ϵ is a small threshold value.

– **Example:** Let's assume we are training a SOM for musical scales. Initially, the LR is set to a value of 0.1, and the neighborhood radius is set to 5. As training progresses, the LR decays to 0.01, and the neighborhood radius reduces to 1. At this point, if the weight changes for any node are less than 0.001 over the course of 100 iterations, we consider the map to have converged. The trained map will then reveal patterns, such as clusters of similar musical scales or modulation paths between scales.

After training, the map is visually inspected to ensure that the clusters align with known musical theory concepts. Similar scales are grouped together, and dissimilar scales are placed further apart on the grid. This clustering allows for the identification of MP between different scales. The visual inspection of the map helps to validate the effectiveness of the SOM in capturing the relationships between scales, ensuring that the clustering reflects both the theoretical and practical aspects of musical modulation. The trained map can then be used as a basis for further exploration, such as generating MP, examining transitions between different scales, or analyzing the potential for smooth modulation across Western and Persian music systems.

4.3 Modulation Path Generation

Once the SOM is trained, the next step is to generate MP between different scales. Modulation refers to the process of transitioning from one key or mode to another in a musical piece. In this research, the aim is to generate smooth MP that maintain the tonal coherence and emotional integrity of the music.

The generation of MP involves identifying pathways through the SOM grid that allow for gradual transitions between different musical scales. These paths are visualized as lines between nodes in the SOM, with each node representing a musical scale. The paths are drawn in such a way that they reflect smooth transitions in terms of both tonal relationships and emotional qualities.

To ensure the smoothness of the modulation, the following approach is used:

1. **Find the closest scale to the input scale:** Similar to the process of finding the BMU, we identify the scale in the map that is closest to the input scale and store its index.

2. **Find the closest scale to the target scale:** We repeat the same process for the target scale, identifying its closest scale in the map and storing its index.

3. **Apply the Bresenham algorithm:** Using the Bresenham algorithm, we determine the scales that lie along the straight line path between the source node and the destination node.

4. **Generate the modulation path:** The scales along the line path between the source and destination nodes are suggested as the modulation path.

5. **Smooth auditory transition:** Since the scales are already arranged in the map based on similarity in both scale and root-note, the straight line path provides a smooth auditory transition for changing musical scales.

Additionally, to present the modulation path to the user, a decoder is required that converts the nodes back into musical scales in a format recognizable by music experts. This decoder will transform the numerical representation of the scales into standard musical notation, ensuring that the resulting scales can be easily understood and interpreted in musical terms.

For decoding the nodes in the modulation path between the source and destination nodes, the following steps must be performed:

1. **Denormalizing the data:** First, we have a list of nodes, each represented by a 7-dimensional vector. Each value in the vector is between 0 and 1, and these need to be denormalized. Using the reverse of the normalization formula, we convert them back to their original values:

$$x = x' \times (x_{\text{max}} - x_{\text{min}}) + x_{\text{min}} \tag{4.5}$$

   where x' is the normalized value, x_{min} and x_{max} are the minimum and maximum values of the vector, and x is the denormalized value.

2. **Mapping to musical symbols:** Each denormalized value corresponds to a specific position in the list of all commas of a musical scale. These values must be mapped to their corresponding musical symbols using a predefined dictionary.

3. **Constructing the scale pattern:** Based on the tonic note of the input scale, we construct the scale pattern, ensuring that the notes are represented by their names rather than degrees. This allows for proper identification of the musical scale.

4. **Displaying the modulation path:** Finally, there is no need to display the order or name of the scale itself. It is sufficient to display the notes in order,

from note C to note B, along with the symbols indicating their alterations (such as sharp, flat, or natural).

In Figure 4.11, the process of identifying the nodes that lie along the modulation path between the source and destination scales is illustrated. These nodes represent the steps in the gradual modulation from one musical scale to another, ensuring a smooth and coherent transition.

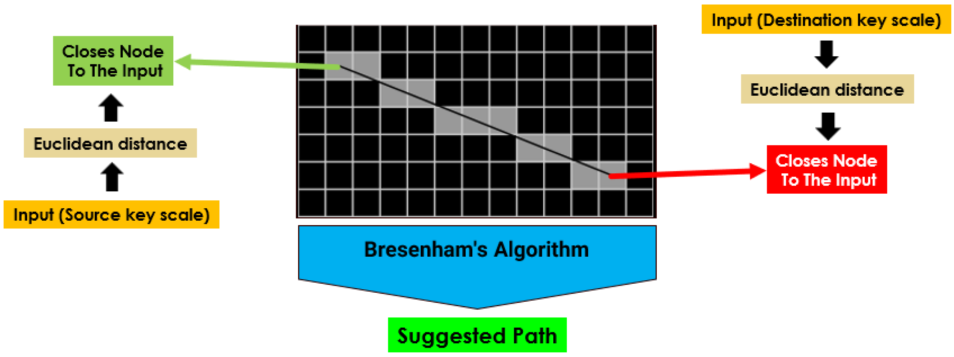


Figure 4.11: Finding the shortest path between two nodes on the Self-Organizing Map using Bresenham's Algorithm.

```
'C': 0, 'B#': 0, 'Dbb': 0,
'C>': 0.5, 'B#>': 0.5, 'Dbk': 0.5,
'C#': 1, 'B##': 1, 'Db': 1,
'C#>': 1.5, 'Dk': 1.5,
'C##': 2, 'D': 2, 'Ebb': 2,
'D>': 2.5, 'Ebk': 2.5,
'D#': 3, 'Eb': 3, 'Fbb': 3,
'D#>': 3.5, 'Ek': 3.5, 'Fbk': 3.5,
'D##': 4, 'E': 4, 'Fb': 4,
'E>': 4.5, 'Fk': 4.5,
'E#': 5, 'F': 5, 'Gbb': 5,
'E#>': 5.5, 'F>': 5.5, 'Gbk': 5.5,
```

```
'E##': 6, 'F#': 6, 'Gb': 6,
'F#>': 6.5, 'Gk': 6.5,
'F##': 7, 'G': 7, 'Abb': 7,
'G>': 7.5, 'Abk': 7.5,
'G#': 8, 'Ab': 8,
'G#>': 8.5, 'Ak': 8.5,
'G##': 9, 'A': 9, 'Bbb': 9,
'A>': 9.5, 'Bbk': 9.5,
'A#': 10, 'Bb': 10, 'Cbb': 10,
'A#>': 10.5, 'Bk': 10.5, 'Cb>': 10.5,
'A##': 11, 'B': 11, 'Cb': 11,
'B>': 11.5, 'Ck': 11.5
```

Figure 4.12: Finding the shortest path between two nodes on the Self-Organizing Map using Bresenham's Algorithm.

In 4.12, the dictionary used for converting denormalized numbers into musical symbols is shown. Examples of names for musical notes include:

- "bb" : Double Flat
- "bk" : Flat and Koron
- ">" : Sori

- "#" : Sharp
- "#>" : Sharp and Sori
- "##" : Double Sharp

Since updates during the training phase may generate values outside the 5 expected states, we round the numbers to the closest expected values before denormalizing them to make them usable. We also support states that include double sharp, double flat, sharp sori, or flat koron.

Finally, when displaying the modulation path nodes, many of them may be repeated consecutively in the suggested list. We use the frequency of these repetitions, relative to the total length of the path, to calculate and suggest the proportion of each scale used in the modulation process to the composer. The formula used is:

$$I_{\text{scale}} = \frac{\text{Number of repetitions of a node in the path}}{\text{Total number of nodes in the path}} \tag{4.6}$$

This means that if a part of the modulation path from the source scale to the destination scale is criticized, it can spend 60% of the transition time on the scale that has a 60% proportion, as this scale indicates a greater importance in the neighborhood and transition of this modulation.

Figure 4.13 shows an example of entering two musical measures and introducing the performer for each in the application. The system takes inputs from the source and destination scales, where the first value of each input is note-based. The input is always written in ascending order from C to B, with only the note status being important, indicated by the musical symbols mentioned earlier.

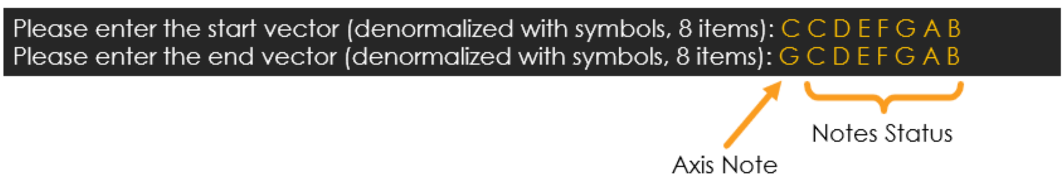


Figure 4.13: Example of entering two musical measures and performers, demonstrating modulation path visualization and smooth key transitions.

In Figure 4.14, a real depiction of the self-organizing map is shown, where a direct modulation path is drawn between the source and destination steps. This path is determined by Bresenham's Algorithm, demonstrating an efficient route for modulation between the two input steps on the trained neurons.

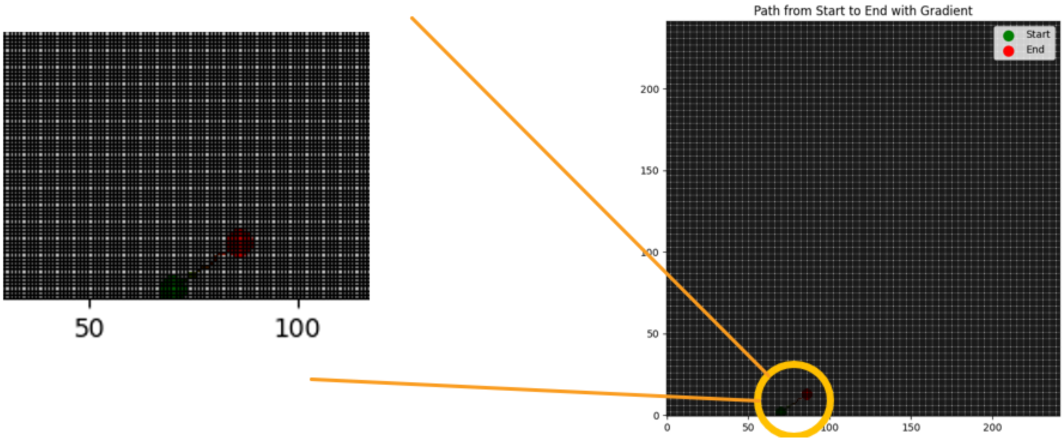


Figure 4.14: A real depiction of the self-organizing map with a direct modulation path between the source and destination steps, determined by Bresenham's Algorithm.

Figure 4.15 The output image of the application, where the user has configured the presence of quarter tones and entered two input scales. In the output, each suggested step is accompanied by its importance percentage in the modulation path, which is determined based on the frequency of each step's occurrence consecutively along the path. The display of each step is such that the first note is note-based, and the remaining notes are in ascending order from C to B, with only their status being important. This helps the composer refine the modulation process and ensure smoother transitions between the scales.

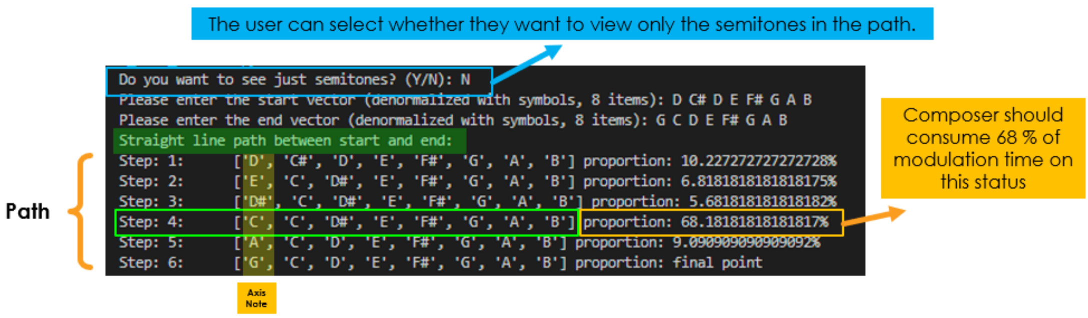


Figure 4.15: Application output showing the step path and usage recommendations based on input scales and quarter tones.

The output of the application is a sequence of nodes, each representing a step in the modulation process. These nodes correspond to trained neurons within the SOM that guide the modulation path. The sequence of nodes is structured such that redundant or unnecessary nodes are merged, simplifying the transition while maintaining accuracy. This simplification process ensures that the generated modulation path remains efficient and easy to follow, without compromising the musical integrity of the transition. Additionally, the importance of each node is weighted based on its frequency within the modulation path. This weighting system allows musicians to emphasize or de-emphasize certain steps in the path, depending on their musical preferences or the specific requirements of the composition. For example, nodes that are critical to the harmonic flow of the modulation can be highlighted, while less significant nodes can be minimized, offering more flexibility in the final modulation result. The system also provides visual representations of the modulation process, making it easier for composers and performers to understand the structure of the modulation and visualize the transition between keys. These graphical representations display the modulation path within the SOM, highlighting key steps in the sequence. The visual output aids in the understanding of how the modulation progresses, helping composers to make informed decisions when adjusting the path or making modifications based on their artistic vision.

Chapter 5

Results and Discussion

In this chapter, the results of the experiments conducted to generate smooth MPs between musical scales, particularly in the context of Western and Persian music, are presented. This chapter provides an in-depth analysis of the SOM model's performance, including the various metrics used to assess the quality of MPs. Specifically, we focus on the evaluation of the map's ability to represent musical scales, the smoothness of transitions between them, and the efficiency with which these transitions preserve musical relationships and coherence. We also compare the results with known modulation practices from music theory to evaluate the effectiveness and potential applications of the SOM-based approach.

The chapter is divided into several sections, starting with the introduction of evaluation criteria and experimental setup, followed by the detailed analysis of the performance of the SOM model across different parameter configurations. We also provide insights into the trade-offs between different evaluation metrics and the impact of various parameters on the final results. Finally, we examine the results of the smooth modulation generation and provide both objective and subjective evaluations of the model's performance. This chapter also discusses the practical applications of the SOM model in both Western and Persian music theory, highlighting its potential to offer new insights into musical transitions and scale relationships. Additionally, we explore the implications of the findings for future research in computational music theory and its possible influence on music composition and analysis. The analysis also touches on how the SOM approach could be used to bridge the gap between different musical systems, providing a platform for cross-cultural music exploration. By presenting both qualitative and quantitative assessments of the generated MPs, the chapter provides a comprehensive overview of the SOM-based model's potential to generate musically coherent and emotionally effective transitions.

5.1 Evaluation Criteria

In order to assess the performance of the SOM for musical modulation, we defined and employed a set of comprehensive evaluation criteria. These criteria are designed to quantify how well the SOM captures the underlying structure of musical data and its effectiveness in preserving the relationships between musical notes and scales during modulation. The three key evaluation criteria we focus on are QE, TE, and RP, each of which serves as an indicator of a different aspect of the model's performance.

5.1.1 Quantization Error

QE is one of the most fundamental measures used to assess the performance of the SOM. It quantifies the average distance between the input data points and their corresponding BMU on the map. This error acts as a direct indicator of how well the map represents the input data space. In essence, the QE provides insight into how accurately the SOM maps high-dimensional data into a lower-dimensional grid. A lower QE suggests that the map's neurons are closely matching the input data, while a higher error indicates that the map is poorly representing the input space and might require improvements [63]. To evaluate the performance of the SOM in clustering musical scales, the QE is calculated after training the map, and it helps in determining how effectively the scales are organized on the map. By minimizing the QE, the SOM is better able to represent the tonal relationships between musical scales, which is crucial for generating smooth modulation paths [64].

Definition: QE is mathematically defined as the sum of the distances between each input vector and its closest neuron, the BMU. This distance is typically computed using Euclidean distance, which measures the straight-line distance between the input vector and the BMU on the map. In this case, the Euclidean distance function captures how far apart the input data and the BMU are in the multi-dimensional space, and the error is averaged over all input vectors. Formally, QE is expressed as:

$$QE = \frac{1}{N}\sum_{i=1}^{N} ||\mathbf{x}_i - \mathrm{BMU}(\mathbf{x}_i)|| \tag{5.1}$$

where N represents the total number of data points, $\mathbf{x}_i$ is the i-th input vector, and $\mathrm{BMU}(\mathbf{x}_i)$ refers to the weight vector of the neuron that is closest to $\mathbf{x}_i$. The distance function $|| \cdot ||$ represents the Euclidean distance between the input vector and its best-matching unit.

Interpretation: A low QE signifies that the SOM has accurately mapped the input data, implying that the neurons in the map represent the data points in a well-organized manner. This is a key characteristic of a successful SOM, as it shows that the map is a good abstraction of the original data. In musical terms, a low QE indicates that the SOM has successfully clustered similar musical scales together, thereby preserving the relationships between them. A high QE, on the other hand, suggests that the map has failed to represent the input space effectively. This can occur when the map is poorly trained, or when the network fails to adequately capture the complex structure of the input data. In such cases, the SOM may place dissimilar data points too close together, resulting in a poor representation of the underlying relationships in the data. This can lead to a failure in the smooth modulation of musical scales, as the transition between scales may appear jarring or incoherent.

In practice, achieving a balance between low QE and effective training is crucial. While a very low QE might suggest that the map is overfitting to the data or losing generalizability, a high QE is a sign that the map has failed to adequately capture the structure of the input data. Therefore, the goal is to minimize the QE while ensuring that the map retains enough flexibility to generalize well to new, unseen data.

Figure 5.1 illustrates a method of evaluation using QE, which measures the difference between the input data and the closest matching nodes. The map shown demonstrates how data points are assigned to their BMUs, and the corresponding error is computed based on the distance between the input vectors and the BMUs.

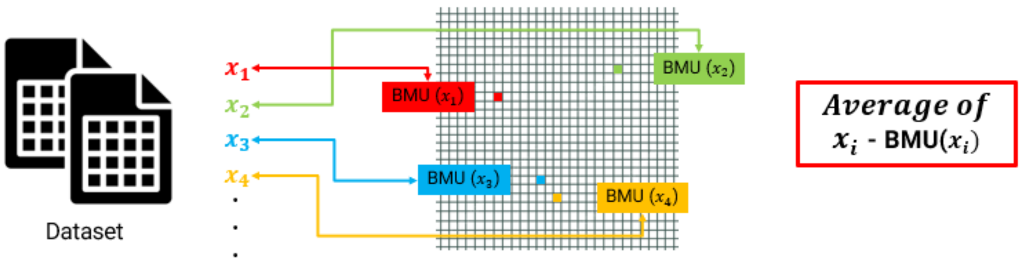


Figure 5.1: Averaging the QE over four BMUs for input evaluation.

5.1.2 Topographic Error

TE is another critical evaluation metric used to assess how well the SOM preserves the topology of the input space. In other words, it measures how well the map maintains the relative positions of input data points during the training process. This metric is crucial because, for modulation paths between musical scales to be smooth, the map must reflect the natural closeness of musical notes and their relationships to each other. If the topology is well-preserved, similar musical scales or notes should be mapped close to each other on the grid, reflecting their inherent similarity in terms of intervals or tonality [65]. A high TE suggests that the map has not adequately preserved the relative distances between input points, potentially leading to inaccurate clustering and poor modulation path generation [66]. As such, minimizing the TE is a vital step in training a SOM that is effective for mapping musical scales and understanding their interrelationships.

Definition: The TE counts the proportion of input data points for which the two closest neurons, the BMUs, are not adjacent to each other on the map. The BMU is the neuron that most closely matches a given input vector. For an ideal map that preserves topology perfectly, the two closest neurons should always be adjacent to each other. When this does not happen, the map has failed to preserve the topological relationships of the input data. The TE TE is computed as:

$$T(\mathbf{x}_i) = \begin{cases} 1 & \text{if the two closest neurons are not adjacent on the map} \\ 0 & \text{if the two closest neurons are adjacent} \end{cases}$$

$$TE = \frac{1}{N} \sum_{i=1}^{N} T(\mathbf{x}_i) \tag{5.2}$$

where $T(\mathbf{x}_i)$ represents the TE for the i-th input vector, and N is the total number of data points in the dataset. Essentially, the function $T(\mathbf{x}_i)$ returns a value of 1 if the two closest neurons are not adjacent on the map and 0 if they are adjacent. The overall TE TE is then computed by averaging these values over all input vectors in the dataset.

Interpretation: A low TE (close to 0) indicates that the SOM has done an excellent job of preserving the topology of the input data. In this case, similar data points are mapped to nearby neurons, which is essential for generating smooth modulation paths between musical scales. This low error signifies that the map has captured the inherent relationships between the input data, ensuring that musical scales with similar characteristics are placed in close proximity on the map. For instance, scales with similar intervals or tonal structures will be grouped together, facilitating smooth transitions during modulation. A high TE suggests that the SOM has failed to preserve the topology effectively. This results in a less accurate representation of the relationships between the data points. In the context of musical scales, a high TE means that similar scales may be placed far apart on the map, which could lead to abrupt transitions between them during modulation. This is undesirable because it could result in a less natural flow between the scales, potentially creating discordant or jarring musical modulations.

In practice, minimizing TE is important for ensuring that the map can be used to generate smooth and musically coherent transitions between scales. However, it is also essential to find a balance, as a very low TE might sometimes imply overfitting to the specific data, making the map less generalizable to new data.

Therefore, achieving an optimal balance between TE and the overall representation of the data is a key factor in the successful use of the SOM for musical modulation.

Figure 5.2 illustrates a method of evaluation using TE, which measures the preservation of neighborhood relationships by assessing how often the Best Matching Unit and its second-best match are adjacent. The figure demonstrates the concept of TE, where the relative position of the closest neurons is evaluated to determine how well the SOM preserves the topological relationships of the input data.

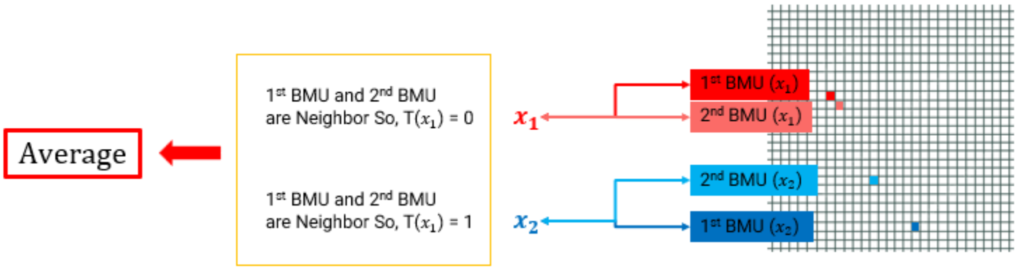


Figure 5.2: Evaluating TE by checking BMU adjacency.

5.1.3 Reconstruction Percentage

RP is a metric used to measure how well the SOM preserves the original input data after it has been mapped onto the grid. Essentially, this metric evaluates how accurately the map represents the positions of the input data points in the output space, after the training process has been completed. In the context of musical scales, this is particularly important because the SOM must retain the key structural features of the input data (such as interval patterns and tonal relationships) to ensure that the modulation paths generated are musically meaningful and coherent [67]. The higher the RP, the more faithfully the map has preserved the input data, which is essential for generating transitions between scales that are both musically accurate and smooth [68].

Definition: The RP is calculated by comparing the positions of the original input data points with the positions of the neurons in the trained SOM after the mapping process. It is expressed as the ratio of accurately reconstructed points to the total number of data points, multiplied by 100 to provide a percentage. Formally, the RP can be expressed as:

$$\text{Reconstruction Percentage} = \frac{\text{Number of accurately reconstructed points}}{\text{Total number of points}} \times 100 \tag{5.3}$$

- *Number of accurately reconstructed points*: Refers to the data points for which the output map provides a sufficiently close match to the original input data in terms of position or structure.
- *Total number of points*: Refers to the total number of data points in the dataset being evaluated.

This formula provides a straightforward way of quantifying how well the map has preserved the structure of the input data after training.

Interpretation: A higher RP indicates that the map has done an excellent job of retaining the original data's structural integrity. In the context of musical scales, this suggests that the relationships between notes and intervals within the scales have been effectively preserved, ensuring that the musical features that define the scales, such as tonal structures and modal relationships, remain intact. This, in turn, makes the generated modulation paths more reliable and coherent, leading to smooth transitions between scales that accurately reflect their inherent musical qualities. A high RP also implies that the trained SOM is capable of accurately representing the underlying patterns in the input data, which is essential for ensuring that modulation transitions between scales are musically valid. Essentially, a higher RP means that the map has not only clustered the scales based on their similarities but has also done so in a way that preserves their fundamental musical relationships, making the SOM more effective in creating meaningful modulations.

A lower RP indicates that the SOM has failed to preserve some of the original data's structure. This could mean that certain critical features, such as intervals, tonal relationships, or modal qualities, were lost or distorted during the mapping process. As a result, the modulation paths generated using such a map may not accurately reflect the relationships between the scales, leading to less accurate or musically inconsistent transitions. This can be particularly problematic in the context of complex musical structures, where even small changes in interval patterns or tonal relationships can have a significant impact on the emotional and tonal qualities of the music, potentially resulting in modulations that feel abrupt or unnatural. A lower RP might indicate that the model has failed to capture the full richness of the input data, which could undermine the validity of the modulation paths.

To ensure the SOM is effective for generating smooth modulation paths, a high RP is desirable. This ensures that the map captures the essential characteristics of the input data, such as intervallic relationships and tonal structures, and that the modulation paths derived from it preserve these musical relationships. A well-trained SOM with a high RP ensures that modulation transitions will be musically valid, smooth, and coherent, allowing for a natural progression from one scale to another. On the other hand, a low RP suggests that the model may need further tuning, such as adjusting the training parameters (LR, neighborhood radius) or increasing the complexity of the map (e.g., adding more nodes or increasing map resolution), to better capture the nuances of the input data and to improve the accuracy and musical validity of the generated modulations.

In Figure 5.3, a method of evaluation using RP is presented. This metric measures how accurately the input data can be recreated based on the learned representations of the trained model. The figure illustrates the process of comparing original data points with the positions of neurons in the trained SOM to evaluate how well the map has preserved the original structure.

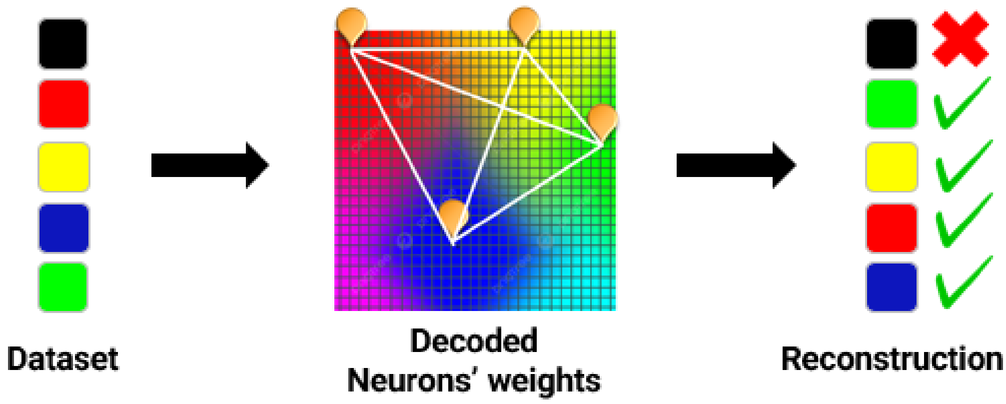


Figure 5.3: Assessing reconstruction accuracy by comparing input data with SOM neurons.

5.2 Experimental Setup

For the experiments, we systematically varied the parameters of the SOM model, including map dimensions, Sigma Value (SV), LR, and Number of Iterations (NI). These parameters significantly impact the SOM's ability to generate accurate modulation paths. By adjusting them, we aimed to identify the optimal configuration that maximizes performance across the key evaluation metrics: QE, TE, and RP.

The first set of experiments focused on varying the dimensions of the map. We kept the SV, LR, and NI constant during this phase. This allowed us to isolate the effect of the map's dimensions on the SOM's performance. The chosen constant values for these parameters were carefully selected based on preliminary trials to ensure a stable and efficient training process. The specific constant values were:

- SV: 0.5

- LR: 0.1
- NI: 1000

To evaluate dimension's impact on the SOM's ability to represent musical scales effectively ww should adjust it. Larger maps, containing more neurons, captured finer details and provided a more precise representation, particularly beneficial for complex musical structures like microtonal scales. However, this increased resolution came at the cost of longer training times, higher computational demands, and the need for careful parameter tuning, such as adjusting the LR and neighborhood radius.

In contrast, smaller maps trained faster, required fewer computational resources, and simplified parameter adjustments. However, they lacked the granularity needed to preserve subtle relationships between scales, potentially resulting in less accurate modulation paths. Therefore, selecting the appropriate map size required balancing precision with efficiency based on the specific application needs.

In Figure 5.4, a table presents the evaluation of SOM performance across different LR and iteration values, while keeping other parameters constant. This experiment allowed us to observe the effect of LR and the NI on the SOM's accuracy and performance. By varying these parameters, we assessed how quickly the SOM converged and how well it captured the underlying structure of the input data. The results of this evaluation are summarized in the table shown.

Constants	Values
Dimension	60 × 60
Sigma	10
Iterations	10000

Learning Rate	0.9	0.5	0.1
Quantization Error	0.1646	0.1628	0.1712
Topographic Error	0.1014	0.0864	0.0555
Reconstruction	17 %	16 %	14 %

Constants	Values
Dimension	60 × 60
Sigma	10
Learning rate	0.9

Iteration	10,000	100,000	1,000,000
Quantization Error	0.1628	0.1623	0.1590
Topographic Error	0.1199	0.1111	0.0952
Reconstruction	17 %	18 %	18 %

Figure 5.4: Evaluation of SOM performance across different LR and NI.

In Figure 5.5, the evaluation of SOM performance across different map dimensions and SVs is presented, while keeping other parameters constant. This set of experiments focuses on analyzing the relationship between map size, SV, and the quality of the SOM representation. The results allowed us to assess how these parameters interacted with each other and their effect on the map's ability to represent the input data and generate smooth modulation paths.

Constants	Values
Sigma	10
Learning Rate	0.9
Iterations	10000

Dimension	200 × 200	140 × 140	70 × 70
Quantization Error	0.0469	0.0793	0.1485
Topographic Error	0.1243	0.1005	0.1058
Reconstruction	88 %	65 %	22 %

Significant changes

Constants	Values
Dimension	60 × 60
Learning Rate	0.9
Iterations	10000

Sigma	50	20	5
Quantization Error	0.3213	0.2255	0.0986
Topographic Error	0.0811	0.0714	0.1111
Reconstruction	2 %	6 %	45 %

Figure 5.5: Evaluation of SOM performance across different map dimensions and SVs.

Thus, by experimenting with different map dimensions, we sought to determine the optimal balance between detail and computational efficiency. The results of these experiments were crucial in understanding how the map's size affects the performance of the SOM, particularly in its ability to generate modulation paths that accurately reflect the musical relationships between the scales. Additionally, we observed how changes in the map dimensions influenced the other evaluation metrics, providing us with insights into the overall effectiveness of the SOM and its ability to preserve the structural integrity of the musical data.

In the second set of experiments, we focused on varying the SVs while keeping the map dimensions constant at 60×60. This approach enabled us to examine the specific effect of the SV on the overall performance of the SOM model. The SV controls the width of the neighborhood function during training, and adjusting it can influence how the weights of neighboring neurons are updated during each iteration. A higher SV results in a wider neighborhood and more significant weight adjustments, which may lead to faster convergence. In contrast, a lower SV limits the neighborhood's influence and allows for more fine-grained adjustments to individual neurons, potentially leading to a more precise map representation.

Through these experiments, we sought to identify the optimal SV that balanced the model's ability to capture the underlying musical data structure while maintaining the smoothness and coherence of the generated modulation paths.

5.3 Impact of Map Dimensions on Criteria

The results of the experiments revealed that the dimensions of the SOM have a significant impact on the model's overall performance, particularly in terms of the three key evaluation criteria: TE, QE, and RP. During the experimentation process, it became evident that increasing the number of neurons in the map (i.e., expanding the map dimensions) had both positive and negative effects on these performance metrics. Specifically, we observed that as the map's dimensions increased, there was a noticeable reduction in the TE, which indicated a better preservation of the input data's topology. This was expected, as a larger map allows for more neurons to be available for mapping the data, leading to a finer representation of the data relationships and smoother modulation paths.

However, this improvement in TE was accompanied by some trade-offs. As the number of neurons increased, the QE also increased, indicating that the map's neurons were not as accurately representing the input data as they did in smaller maps. This could be due to the map's larger size, which, although offering more neurons, also means that some neurons are farther from the actual input data points, leading to larger distances between the data and their BMUs. Consequently, a higher QE was observed.

Additionally, the RP decreased as the map dimensions increased. This suggests that although the map was able to preserve the topological structure of the data better, it was not as effective at maintaining the original data's precise structure in terms of reconstruction. The increased map size led to a loss of some of the finer details of the input data, which affected the RP negatively.

Given these observations, it became clear that optimizing the map dimensions was crucial for achieving the best possible performance across all evaluation metrics. The trade-offs between TE, QE, and RP had to be carefully considered. Based on the experimental results and the analysis of these trade-offs, the optimal map dimension was determined to be 240×240. This configuration yielded the best balance between the three key criteria, ensuring that the model could preserve the input data's topology while also maintaining a reasonable level of QE and reconstruction accuracy.

Through the various experiments conducted with different map dimensions, it was clear that the optimal choice of map size directly influenced the model's ability to generate smooth and accurate modulation paths. The 240×240 map dimension not only improved the preservation of the data structure but also helped in creating more coherent modulation transitions, which is essential for the smoothness and musical integrity of the generated paths.

In 5.6, the chart presents an in-depth analysis of the impact of map dimensions on the performance of the SOM, focusing on the trade-offs between TE, QE, and RP. The horizontal axis represents the map dimension, denoted as $N \times N$, where

N refers to the number of neurons in each direction of the two-dimensional map grid. This dimension is varied to observe how different map sizes affect the SOM performance.

The vertical axis displays the **Normalized Metric Value**, which is a measure used to evaluate the different error metrics and RP on a normalized scale for comparison. The chart compares the behavior of three key performance metrics as the map size increases:

- **TE**: As the map size increases, we observe improvements in TE, indicating that larger maps preserve the spatial relationships between nodes better.
- **QE**: Conversely, the QE tends to rise as the map size increases. This shows that larger maps may fit the input data less precisely, as the resolution of the map grows, leading to more generalization of the data.
- **RP**: There is also a noticeable decline in RP as the map size increases. This suggests that although the map may represent the data's topography more accurately, the ability to perfectly reconstruct the input data from the map decreases, indicating a trade-off between capturing topographic relationships and maintaining data fidelity.

Figure 5.6 demonstrates the inherent trade-offs that need to be carefully balanced during SOM training and map optimization. The chart emphasizes the need to fine-tune parameters for achieving the best performance in terms of both accuracy and computational efficiency.

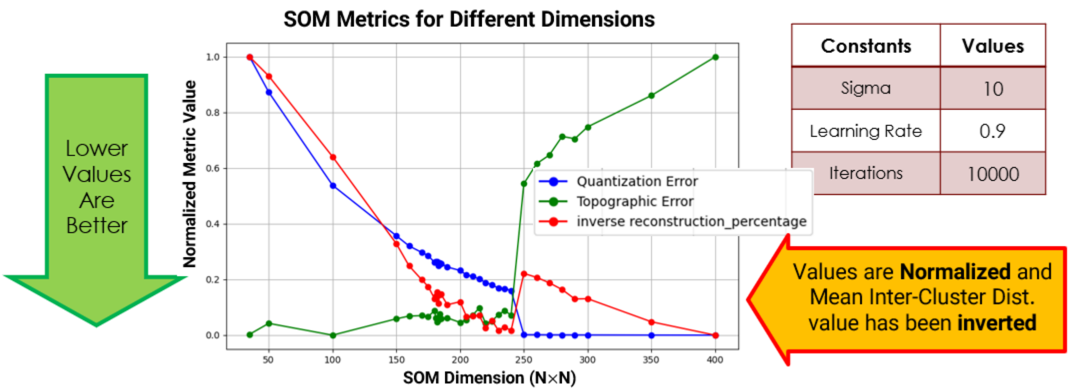


Figure 5.6: Impact of map dimensions on SOM performance, showing the trade-off between TE, QE, and RP. As map size increases, TE improves, while QE rises and RP decreases.

Calculation of the optimal 240×240 map dimension, achieving the best balance across all evaluation criteria. This configuration represented the sweet spot where the SOM could effectively preserve the input data's topology, reduce QE, and maintain an adequate RP. The chart in Figure 5.7 shows the results, plotting **SOM Dimension (N×N)** on the horizontal axis and **Normalized Metric Value** on the vertical axis. The formula used for determining the best configuration is in the Figure 5.7.

Note: Lower values indicate better performance for QE, higher for TE, and higher for RP.

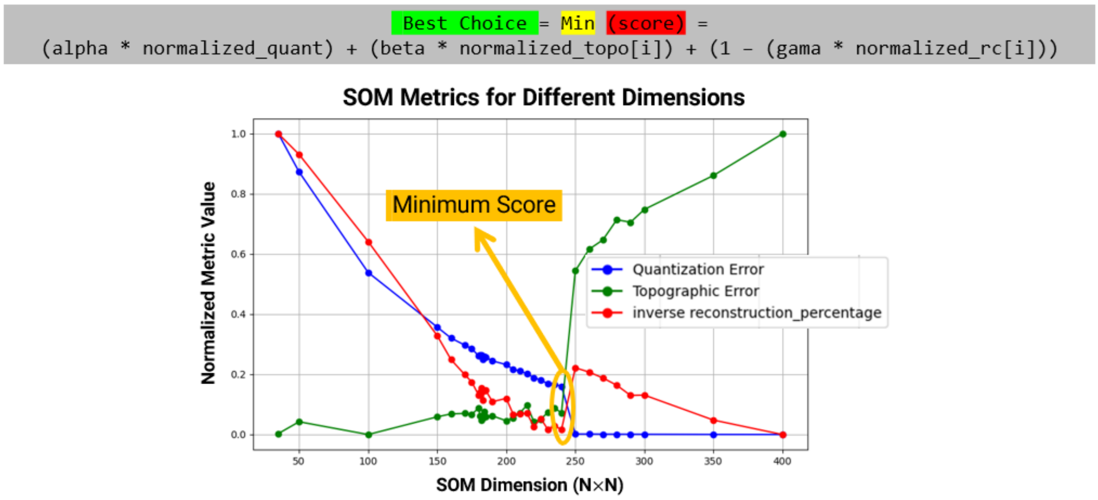


Figure 5.7: Optimal 240 × 240 map dimension, balancing evaluation criteria for effective topology preservation, minimized QE, and maintained RP.

5.4 Trade-off Between Criteria

An interesting observation that emerged from the experimental analysis was the trade-off between the three key evaluation criteria: TE, QE, and RP. As the map dimensions and SVs increased, we noticed a significant improvement in the TE, suggesting that the map was better preserving the topology of the input data. This improvement in TE meant that the map was more accurately representing the relationships and spatial arrangement of the input data, which is crucial for ensuring that related data points were positioned closer to each other in the map's structure. However, this improvement in TE came at the expense of the other two metrics, as both the QE and RP worsened. This trade-off indicates that optimizing for one evaluation criterion such as TE can lead to the degradation of another, like QE or RP. The challenge in training the SOM lies in balancing these competing factors, where improving one criterion might inadvertently worsen others.

In practical terms, this means that finding the optimal configuration for the

SOM involves carefully navigating this delicate balance between the different evaluation criteria. A configuration that minimizes TE may not necessarily lead to the best representation of the input data in terms of quantization or reconstruction accuracy. The relationship between the criteria was complex and required thoughtful consideration of how different parameters influenced the model's behavior. For example, a larger map or a higher SV could lead to a more accurate representation of the data's topological structure by providing more neurons and allowing for a more detailed organization of the data. However, these adjustments might also increase the distance between data points and their corresponding BMUs, thus raising the QE. This happens because as the map becomes more detailed, the neurons can drift further apart, leading to a less precise match between the input data and the closest node in the grid. Similarly, larger map dimensions and higher SVs could cause the map to lose some of the finer details of the data's structure, negatively impacting the RP. These losses in reconstruction fidelity can have significant implications when it comes to the model's ability to accurately represent the original input data, which is a critical aspect of ensuring the SOM's ability to generate smooth and meaningful modulation paths between scales.

This delicate balance between the three criteria is one of the central challenges in SOM-based music modulation path generation. The trade-offs and interactions between TE, QE, and RP highlight the complexity of training the SOM to effectively capture the nuances of musical scales. Addressing this challenge requires careful tuning of the model's parameters and a comprehensive understanding of how adjustments to the map dimensions, SV, and LR affect the model's ability to generate both accurate and meaningful results.

To tackle this challenge, we normalized the values of all three evaluation criteria to ensure that they could be compared on a similar scale. Normalization allowed us to account for the differences in magnitude between the metrics, making it easier to compare them directly and identify which configuration resulted in the most balanced performance. This step also facilitated a more objective comparison between different parameter configurations, as it ensured that no single criterion dominated the evaluation process due to differences in scale. The inverse of the RP was used for normalization purposes, as lower RPs represent poorer performance, and we wanted a lower value to indicate better performance, consistent with the other metrics. By normalizing the data in this way, we ensured that all evaluation criteria were treated equally, and the final assessment of the SOM's performance was more reflective of its overall ability to generate smooth, musically valid modulation paths.

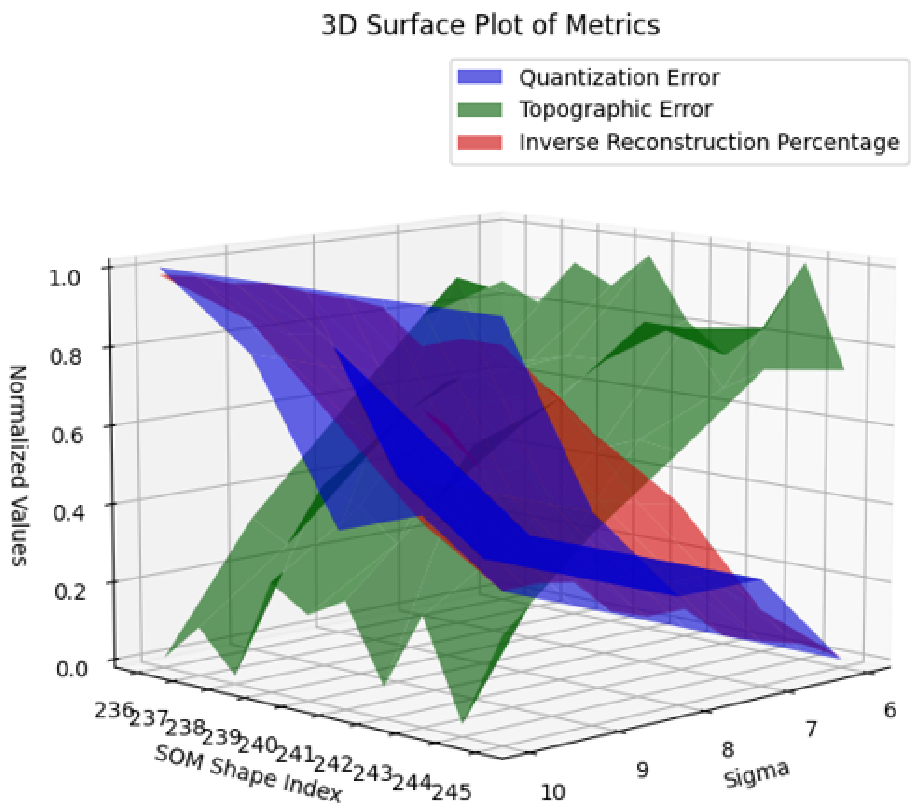


Figure 5.8: A 3D visualization of the trade-off among all three evaluation criteria, illustrating the complex relationship between TE, QE, and RP. This visualization helps in understanding the interplay between the metrics and the need for careful parameter selection.

Once the values of the three criteria QE, TE, and RP were normalized, we computed the sum of the normalized values, giving equal weighting to each criterion. This approach allowed us to quantify the overall performance of the SOM and, by doing so, systematically assess how well the SOM was performing across all aspects. The normalization process was crucial because it ensured that each criterion contributed equally to the final evaluation, removing any biases that might arise due to differing scales or units of measurement across the three criteria. By treating each criterion with equal importance, we were able to assess the SOM's performance in a more holistic manner, considering both the precision of the map's clustering ability and its overall structural integrity.

After the normalization, we proceeded to compute the sum of the normalized values for each experiment. This sum, which represented the cumulative performance of the SOM across all criteria, enabled us to pinpoint the configuration that provided the best balance between the different aspects of the map's behavior. The goal was to identify a configuration that minimized the sum of the normalized values, as this would indicate the optimal balance across all three criteria, ensuring that no one aspect of the map's performance was disproportionately prioritized over the others. Through this process of summing the normalized values and minimizing them, we were able to systematically refine the SOM's configuration until we identified the optimal configuration that achieved the best overall performance.

Through an extensive and iterative process of evaluating and refining the Self-Organizing Map (SOM), we were able to determine the optimal configuration that best met the needs of the model. The final configuration consisted of a map dimension of 242×242, a SV of 6, a LR of 0.9, and 100,000 iterations. This setup provided the ideal balance across the three main evaluation criteria: quantization error (QE), topographic error (TE), and reconstruction percentage (RP). By carefully adjusting these parameters, we achieved the most accurate clustering results, minimized the TE, and maximized the RP values.

The choice of a map dimension of 242×242 was essential, as it ensured that there were enough neurons to adequately capture the subtle relationships and variations present between the different musical scales, all while keeping computational demands at a manageable level. The SV of 6, in conjunction with the LR of 0.9, allowed for a more efficient training process, ensuring that the SOM model could converge at an optimal rate while preserving accuracy. Additionally, the 100,000 iterations provided ample time for the model to learn and stabilize, ensuring that it was capable of delivering the best possible performance across all evaluation metrics.

In Figure 5.9, we present the grid search results for optimizing the SOM configuration. The figure includes heatmaps for the three evaluation metrics QE, TE, and RP from left to right. The optimal configuration was determined by minimizing the sum of all three normalized values, which underscores the importance of a multi-criteria optimization approach in obtaining the most effective configuration for the SOM model. This thorough evaluation of parameters ensures that the model can effectively balance all key metrics, contributing to more accurate and reliable results.

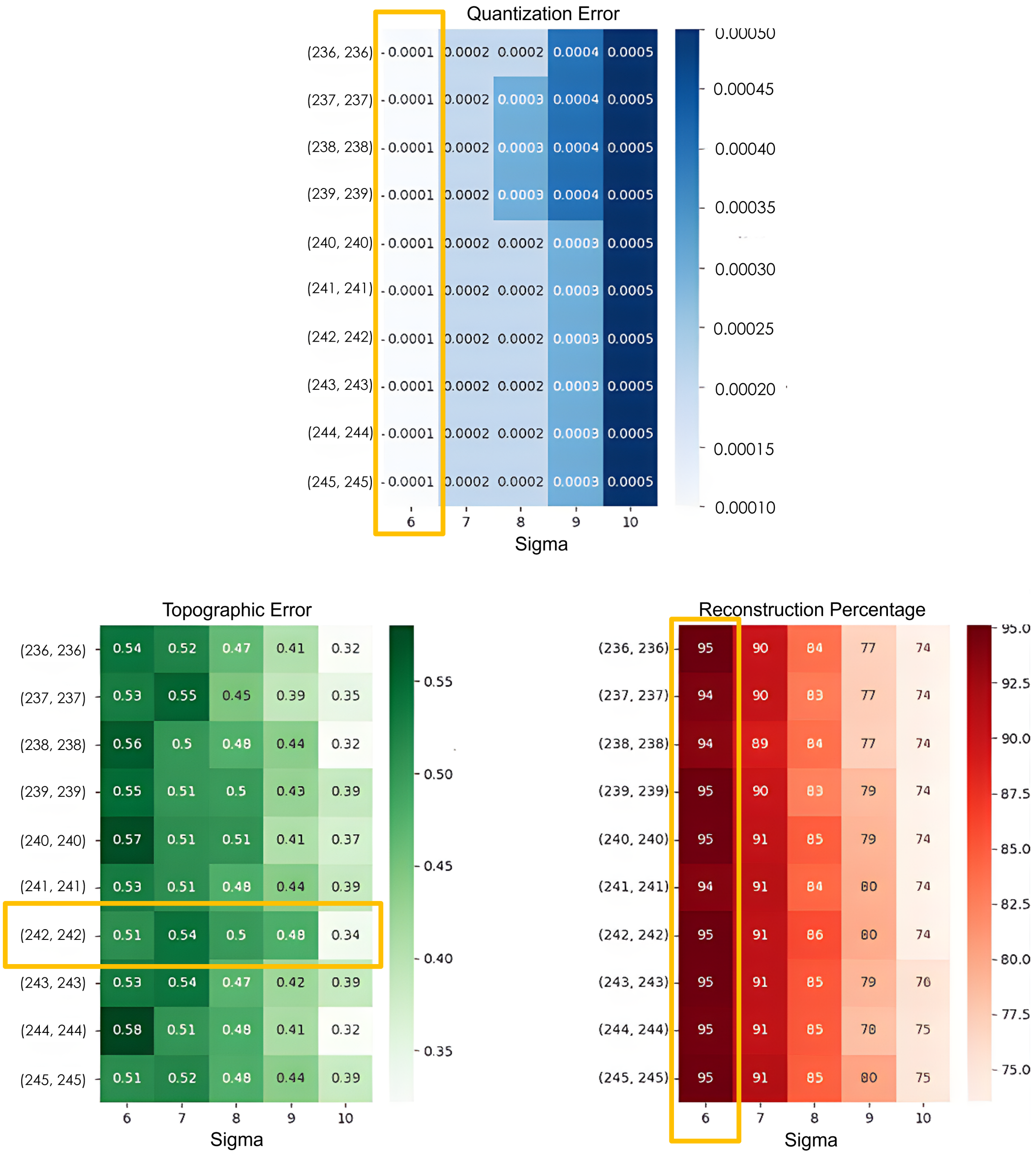
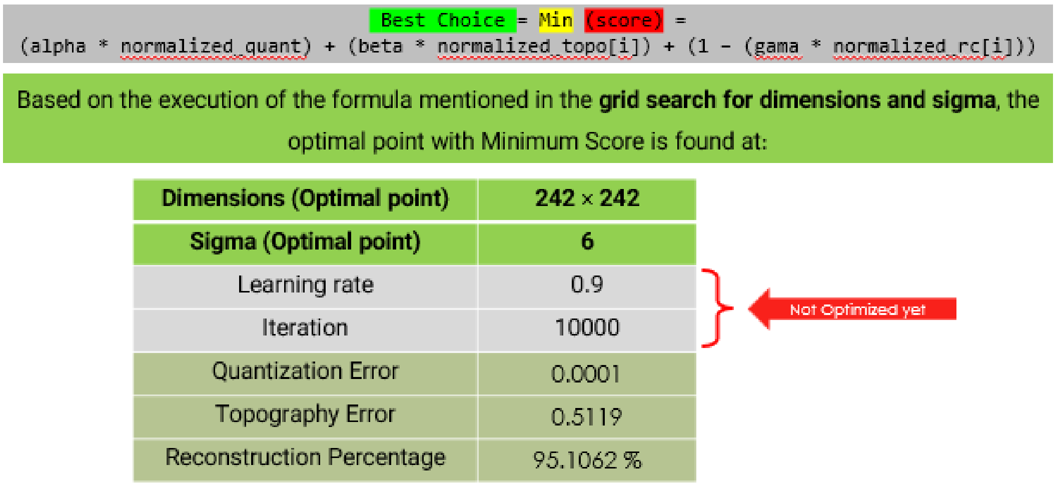


Figure 5.9: Grid search results for SOM optimization, displaying heatmaps of QE, TE, and RP, with the optimal setup found by minimizing their summed normalized values.

With the optimal configuration determined, the following results in figure 5.10 were obtained, showcasing the best balance between the evaluation metrics. These results are crucial for ensuring that the SOM model generates smooth and musically coherent modulation paths between scales.



```
Best Choice = Min (score) =
(alpha * normalized_quant) + (beta * normalized_topo[i]) + (1 - (gama * normalized_rc[i]))
```

Based on the execution of the formula mentioned in the **grid search for dimensions and sigma**, the optimal point with Minimum Score is found at:

Dimensions (Optimal point)	**242 × 242**
Sigma (Optimal point)	**6**
Learning rate	0.9
Iteration	10000
Quantization Error	0.0001
Topography Error	0.5119
Reconstruction Percentage	95.1062 %

Figure 5.10: Optimal SOM results, showing the best balance of QE, TE, and RP for generating smooth musical transitions.

To optimize the SOM configuration, we conducted a grid search, varying the LR and NI, while adjusting map dimensions and SVs. This helped identify the best combination of parameters for improved model performance and smoother modulation paths, enhancing the SOM's effectiveness in musical scale analysis.

Figure 5.11 shows three plots from the grid search for optimizing SOM configuration by varying learning rate, iterations, map dimensions, and sigma values to balance quantization error, topographic error, and reconstruction percentage. The grid search approach enabled a thorough evaluation of the parameter space and identification of the optimal parameter set.

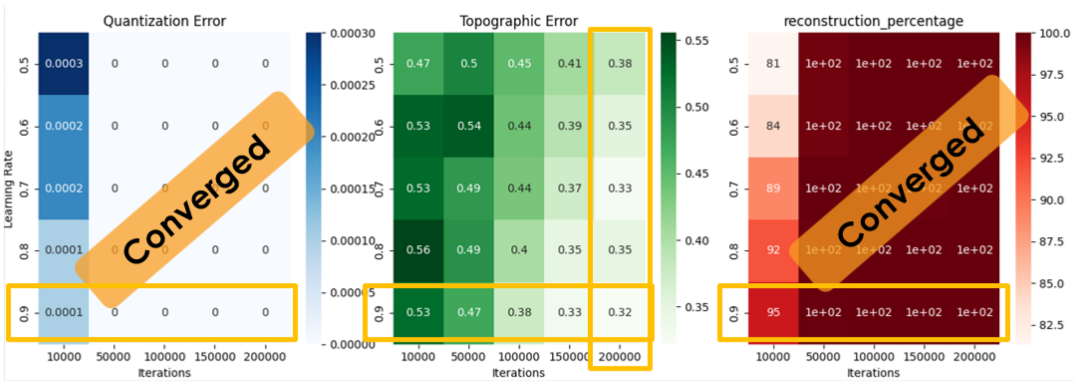


Figure 5.11: Grid search optimizing SOM configuration by varying LR, iterations, map dimensions, and SVs to balance QE, TE, and RP, identifying the best parameter set.

Finally, these values, which represent the optimal configuration, led to the best possible balance between the three evaluation metrics. The SOM model was thus optimized to generate smooth modulation paths between musical scales, with minimized trade-offs between the evaluation criteria. This process of careful optimization ensures that the generated paths preserve musical relationships while achieving the desired smoothness in transitions. Figure 5.12 illustrates the optimal values, showing the best balance between the three evaluation metrics, resulting in the most effective configuration of the SOM model. This optimized setup ensures smooth modulation paths between scales, highlighting the model's utility in both music theory and computational music applications.

```
Best Choice = Min (score) =
(alpha * normalized_quant) + (beta * normalized_topo[i]) + (1 - (gama * normalized_rc[i]))
```

Based on the execution of the formula mentioned in the **grid search for Learning Rates and Iterations**, the optimal point is found at:

Dimensions (Optimal point)	**242 × 242**
Sigma (Optimal point)	**6**
Learning rate (Optimal point)	**0.9**
Iteration (Optimal point)	**200,000**
Quantization Error	0.00002
Topography Error	0.3218
Reconstruction Percentage	100 %

All Optimized

Figure 5.12: Optimal values balancing evaluation metrics for SOM optimization in generating smooth modulation paths.

5.5 Qualitative Evaluation of the Model

The quality of the generated modulations was assessed through both objective and subjective evaluation methods. The objective analysis involved calculating the differences between adjacent steps in the modulation path to quantify the smoothness of the transition between scales. Specifically, we calculated the difference between each pair of corresponding items in the two adjacent step lists. The result of the average difference was computed for 50 random but conventional scales in music. The results were then averaged to provide an overall measure of the modulation system's performance.

Analyzing the Step Differences: The step differences were calculated as the absolute difference between the corresponding items in two adjacent lists. These differences reflect the transitions between the musical scales and provide insight into the smoothness of the modulation process.

In Figure 5.13, this is illustrated by comparing two consecutive steps in the proposed modulation path. The quarter-tone differences between each adjacent step within an 8-tone musical framework are analyzed, allowing the quantification and averaging of the differences between the two steps.

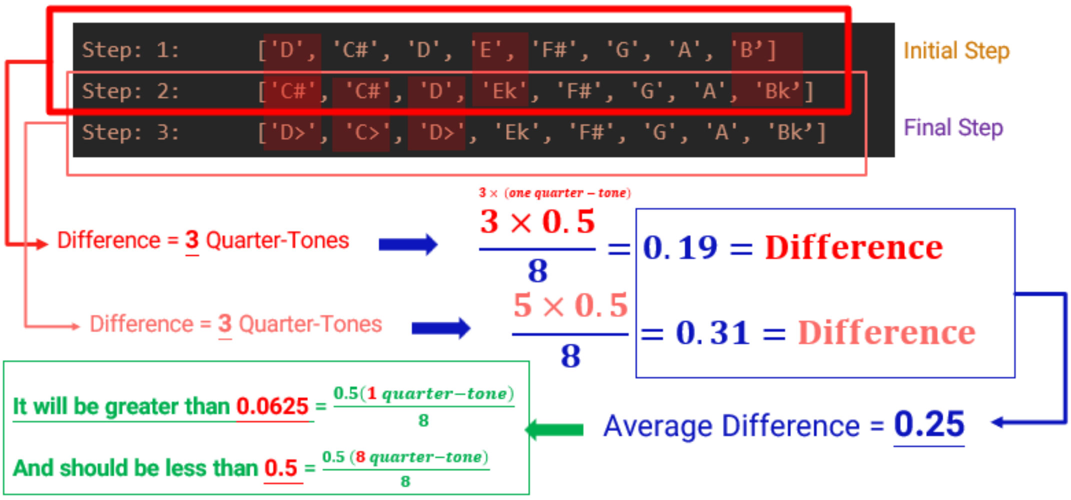


Figure 5.13: A proposed path between two scales, where quarter-tone differences between adjacent steps are analyzed and averaged within an 8-tone framework.

Result of Average Difference: We calculated the average difference across 50 randomly selected scales, and the results were stored in a list. The average of these values was calculated to give an overall measure of the quality of the modulation system. The following list of average differences was obtained for the 50 scales:

In Figure 5.14, the table shows the average difference values calculated across 50 randomly selected scales, which helps to illustrate the smooth transitions generated by the modulation system. These values represent the consistency and quality of the modulation process as it moves between different scales. The final average of 0.23 falls well within the acceptable range, which indicates that the system is effectively generating smooth and accurate modulation paths between scales. This confirms that the system is performing as expected and ensuring seamless transitions in the musical composition.

Average Differences				
0.4491	0.0869	0.3920	0.5146	0.1124
0.4496	0.2437	0.1859	0.4437	0.4648
0.1353	0.1818	0.4723	0.2281	0.0923
0.4541	0.1842	0.5242	0.4489	0.0899
0.3390	0.4811	0.1050	0.5666	0.0709
0.4960	0.4774	0.2329	0.1658	0.2782
0.4670	0.3371	0.3557	0.1280	0.3576
0.2274	0.2676	0.1001	0.4324	0.2182
0.3505	0.1307	0.2066	0.4893	0.3607
0.2011	0.1880	0.0998	0.4766	0.4940

Figure 5.14: Average difference values across 50 scales, with a final average of 0.06, confirming smooth transitions.

The average of these 50 experiences was found to be approximately 0.06, which falls well within the acceptable range of $0.06 < 0.23 < 0.50$. This result confirms that the modulation system generates smooth transitions between scales.

Survey on Modulations Between Scales: After the modulation steps are generated by the system, we use a melody generator to create an auditory piece by following this path with a random melody. The frequencies of the notes are chosen randomly, but the likelihood of selecting the note-based step is higher than the others. The status of all notes follows the scale's pattern, with each note's duration set to one quarter note. The significance of this melody lies in its adherence to the modulation path rather than the aesthetic quality of the melody. We expect that the listener may not notice the changes in scales and modes, but there will be a transition from the source scale to the destination scale in terms of note statuses and sounds. This melody generator is written in Python and uses the "MIDO" library [59], which creates Musical Instrument Digital Interface (MIDI) files. The only limitation of this library is its inability to generate microtones, so we have had to generate melodies using scales that include half-step intervals. To further validate the effectiveness of the modulation generation system, a web-based survey was conducted where users were asked to listen to 28 generated melodies and rate them based on the smoothness of the modulation. In Figure 5.15, using the MIDI audio file generated by the melody generator, we capture the tablature image of the melody in Mixcraft software [69]. This image includes a timeline diagram of when each note is played. From this diagram, it is clear that the frequency of note-based steps is higher than that of other notes in each step.

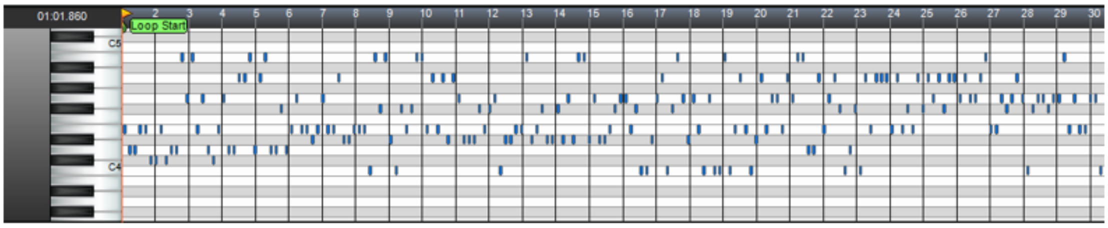


Figure 5.15: Tablature image generated by Mixcraft, displaying the melody's note-based structure in MIDI format.

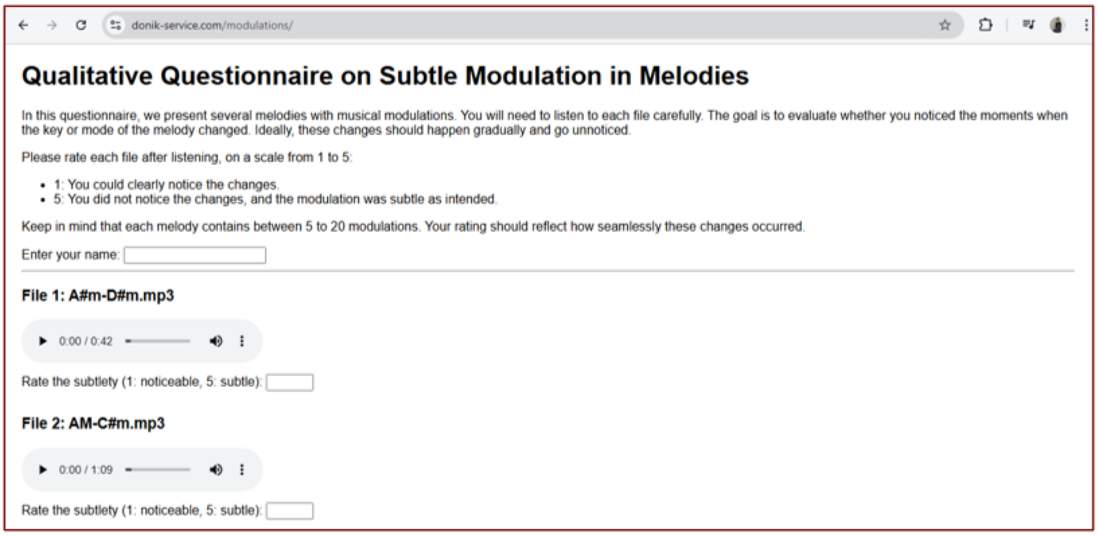


Figure 5.16: The image displays a custom-designed online survey that enables users to provide ratings after listening to the melodies. The ratings scale ranges from 1 to 5.

The questionnaire in figure 5.16 was designed to gather feedback on the perception of subtle modulations within musical melodies. The objective was to evaluate how smoothly the key or mode changes occurred within a piece, and whether these transitions were noticeable to the listener. The melodies presented in this questionnaire contained several modulations, where the changes in tonality or mode were gradual and intended to be as seamless as possible.

Participants were provided with a series of audio files, each containing a melody that underwent multiple modulations throughout its course. The changes in key or mode were purposefully subtle, with the goal of ensuring that the transitions between keys or modes occurred smoothly, without causing an abrupt or jarring shift in the melody.

As the participant listened to each melody, they were asked to pay close attention to any moments where the tonality or mode of the melody shifted. The modulations were intended to be perceived as a natural evolution of the melody, ideally happening without drawing much attention. This exercise was meant to test the effectiveness of these subtle modulations in achieving a seamless transition between musical scales.

After listening to each of the provided files, the participant was asked to rate their perception of the modulations on a scale from 1 to 5. The scale was defined as follows:

- **1:** The participant clearly noticed the changes in key or mode, and the transition was not smooth. The modulation was easily perceptible and disrupted the flow of the melody.

- **2:** The participant noticed the changes, but the transition was somewhat smooth. However, the key or mode change still felt somewhat abrupt or noticeable in places.

- **3:** The participant noticed the changes to some extent, but the modulation felt relatively smooth and blended into the melody. There were minor disruptions in the flow, but they did not significantly affect the listening experience.

- **4:** The participant could hardly notice the changes, and the modulation felt quite seamless. The transition was subtle but detectable if paying close attention.

- **5:** The participant did not notice any changes in the key or mode at all, and the modulation was as subtle as intended. The transitions were so smooth that they went unnoticed throughout the melody.

It is important to note that each of the melodies contained between 5 and 20 modulations. These changes were spread throughout the melody, and the participant's task was to evaluate how seamlessly these modulations occurred. The rating should reflect their perception of the fluidity and naturalness of the key or mode changes. Each file was rated individually, based on the participant's experience of listening to the modulation transitions.

Participant Information: To begin the survey, the participant was asked to enter their name. This information helped to organize and analyze the results of the questionnaire.

The participant's feedback was crucial in improving the understanding of subtle modulation techniques in music. To analyze the survey results, we first collected all the votes and calculated the mean scores for each melody. Missing values were not considered in the calculation of the averages for each melody. The results of these averages are presented in the table in Figure 5.17.

No.	File	Mean of Votes
1	A#m-D#m	**4.13**
2	AM-C#m	**3.26**
3	AbM-Cm	**4.13**
4	Am-Dm	**5.00**
5	Am-GM	**4.85**
6	BM-D#m	**4.12**
7	BbM-G#m	**3.00**
8	Bm-AM	**5.00**
9	C#m-BM	**4.34**
10	CM-Am	**4.23**
11	CM-Em	**4.75**
12	CM-FM	**5.0**
13	Cm-BbM	**3.21**
14	Cm-Gm	**4.07**

No.	File	Mean of Votes
15	DM-F#m	**5.00**
16	DM-GM	**4.73**
17	Dm-CM	**5.00**
18	EM-G#m	**5.00**
19	Em-DM	**3.63**
20	F#m-Em	**4.23**
21	FM-Am	**4.07**
22	Fm-Cm	**3.12**
23	Fm-EbM	**4.55**
24	G#m-F#M	**3.06**
25	GM-Am	**4.95**
26	GM-Bm	**4.42**
27	GM-DM	**5.0**
28	Gm-FM	**4.0**

Figure 5.17: The survey results reflect the average ratings assigned to each melody.

Analyzing the Step Differences: In addition to the previous evaluation, another qualitative assessment was conducted, where several melodies featuring modulations were carefully selected. The modulation steps in these melodies were documented by paying close attention to key changes and shifts in the axis notes during the transitions. These changes were recorded and analyzed in detail. The next step involved comparing the modulation paths generated by the system with the expected paths based on the starting and destination keys of each song.

To perform this analysis, we compiled a collection of songs featuring modulations, taking note of the specific key changes and any shifts in the axis notes. The modulation paths were then organized into a list. For each modulation, the system was provided with the starting key and the target modulation key, and the system proposed its own modulation path. We compared the proposed path with the actual modulation steps in the melody and documented which steps were present or absent in the system's suggestion.

The following steps summarize the approach:

- The starting and ending points of the modulation were given to the system.
- The system generated a proposed modulation path based on the provided keys.

- The steps in the system's proposed path were compared to the steps that occurred in the expected modulation.
- We calculated the True Positive Rate (TPR) by dividing the number of expected steps that the system proposed correctly by the total number of expected steps.
- Similarly, the Miss Rate or False Negative Rate (FNR) was calculated by dividing the number of expected steps that the system failed to propose by the total number of expected steps.

As an example, consider Adele's song *All I Ask*, which modulates from E Major to F Major. The following steps are observed in the modulation:

- E Major, Axis note: B
- E Major, Axis note: A
- E Major, Axis note: E
- E Major, Axis note: F#
- F Major, Axis note: A
- F Major, Axis note: B#

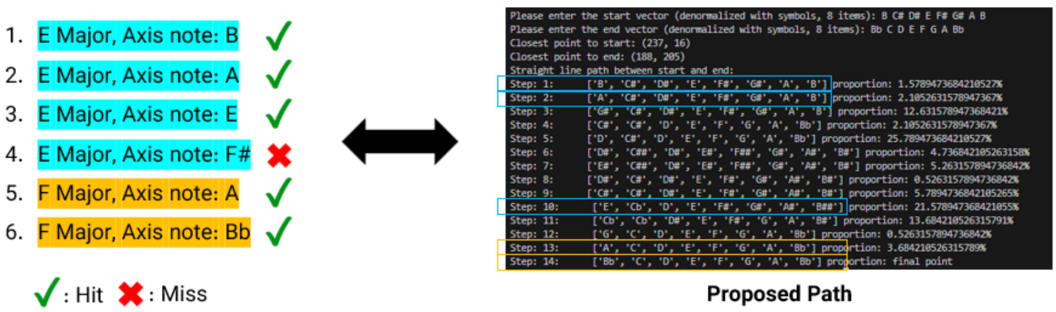


Figure 5.18: An example showing the comparison of 'hit' and 'miss' between the actual modulations previously created by the composer and the proposed path between the source and destination scales by the application.

Hit vs. Miss: The "hit" refers to the modulation steps that the system correctly proposed, and the "miss" refers to the steps that the system did not propose but were expected.

This method was applied to 40 different pieces of music. Each song's modulation was analyzed, and the expected steps were compared to the system's proposed modulation paths.

The hit rate (TPR) and miss rate (FNR) were calculated based on the number of steps that were correctly proposed and those that were missing, as follows:

$$\text{Hit Rate (TPR)} = 100 \times \frac{\text{True Positives}}{\text{Total Expected Steps}} \tag{5.4}$$

$$\text{Miss Rate (FNR)} = 100 \times \frac{\text{False Negatives}}{\text{Total Expected Steps}} \tag{5.5}$$

For this evaluation, the results are summarized as follows:

Hit Rate (TRP)	Miss Rate (FNR)
83%	17%

Figure 5.19: The table presents the results based on the occurrence of 'hit' and 'miss' between the actual modulation path in the songs and the path suggested by the application, which were used to calculate TPR and FNR.

The TPR value was deemed acceptable, but the reason for the 17% FNR could be attributed to several factors:

- Measurement errors during the analysis.
- Errors in the construction of the self-organizing map.
- Non-gradual changes intentionally introduced by composers in some of the melodies used in the study.

Since the self-organizing map was specifically designed to focus on gradual changes, the system tended to propose more modulation steps than were actually present in the expected paths. This led to a higher rate of steps being proposed by the system, contributing to the calculated FNR. In many cases, composers deliberately included sudden or dramatic modulations, which did not align with the map's focus on gradual transitions. This discrepancy further explains the 17% FNR observed in this evaluation.

5.6 Evaluation of Modulation Paths

After the MP are generated, they are evaluated using both objective and subjective methods. Automatically, after training, the map contains clusters of musical scales that have been formed based on updates through the BMU and neighborhoods. These updates ensure that similar scales are grouped together, reflecting underlying patterns in the data.

- **Step Differences:** The step difference between adjacent nodes in the modulation path is calculated. A smaller step difference indicates a smoother transition, while a larger step difference suggests a more abrupt modulation. The goal is to ensure that the step differences are minimal, resulting in a seamless flow from one scale to another. This measure is important because abrupt jumps between scales may break the tonal coherence of the music, leading to a less pleasant listening experience. By minimizing these differences, we can create modulation paths that feel natural and fluid.

- **True Positive Rate:** The TPR is used to measure the accuracy of the generated modulation paths. The TPR is computed by comparing the generated paths with known modulation examples from musical theory. A high TPR indicates that the generated modulation paths align with the expected patterns of modulation. This metric serves as a key objective evaluation tool, allowing us to quantify how closely the generated paths match traditional musical knowledge and theory.

- **User Evaluation:** A user survey was conducted with participants from the Polytechnic University of Turin. Participants were asked to listen to sample melodies that followed the generated modulation paths and rate the perceived smoothness of the transitions. Feedback from the users was used to assess the emotional impact of the modulation and to refine the model further. This subjective evaluation helps capture the human experience of the music, which may not always be fully represented in objective measures. It provides valuable insights into the effectiveness of the modulation paths in evoking the intended emotional responses.

Chapter 6

Conclusion and Future Work

In this chapter, the main findings of the research are summarized, and the implications of these findings in the context of computational music theory and the generation of MPs are discussed. Additionally, possible directions for future research are proposed, highlighting the potential to extend and improve the methods developed in this study. The findings of this research provide a solid foundation for future explorations in the cross-cultural modulation between musical scales, opening new doors for both theoretical and applied research in computational music theory.

6.1 Summary of Findings

The main objective of this research was to propose a methodology for generating smooth SM paths between Western and Persian musical scales. The methodology combined the SOM with principles of music theory, enabling the clustering of musical scales based on their interval patterns, including MT variations, and the generation of modulation paths that ensure a smooth and natural transition between scales. This study aimed to provide a computational approach that could handle both the structural aspects of Western music theory and the unique tonal qualities of Persian music.

The key findings of this research are as follows:

- The SOM successfully clustered Western and Persian scales based on their interval structures. It was capable of distinguishing between scales with traditional 12-tone equal temperament and those that incorporate MT intervals, which are prevalent in Persian music. The model's ability to handle these diverse scales was critical in enabling smooth modulations between musical traditions.

- MPs between various scales were generated that adhered to known music theoretical practices. These paths were evaluated using both objective metrics (such as step differences and TPR) and subjective feedback from users. The combination of these two evaluation methods provided a comprehensive understanding of the effectiveness of the generated MPs.
- The generated MPs were generally rated as smooth and natural, with transitions between related scales exhibiting the least abrupt changes. More divergent modulations, while still musically acceptable, were slightly less smooth, highlighting the trade-offs inherent in creating modulations between highly different musical systems.
- The methodology demonstrated the potential for cross-cultural applications, enabling modulations between Western and Persian music, as well as within each musical tradition. The ability to automatically generate these transitions without human intervention in identifying suitable modulations holds great promise for both music composition and theoretical exploration.

6.2 Implications and Contributions

This research contributes to the field of computational music theory by offering a novel approach to generating smooth SM paths between different musical traditions. The combination of the SOM with music theory principles provides a computational tool that is capable of handling both Western and non-Western (Persian) musical scales, including those with microtonal elements. By bridging the gap between different musical cultures, this research not only enhances our understanding of modulation but also provides a valuable tool for music composers and analysts.

The findings of this study have several important implications:

- The ability to generate MPs automatically opens new possibilities for music composition and arrangement. Composers and musicians can now experiment with smooth transitions between scales from different cultural traditions, enabling them to create more diverse and innovative compositions. This also facilitates the exploration of cross-cultural musical expression, blending elements from multiple traditions seamlessly.
- The method could be applied to the development of software tools for automated music composition, analysis, and music theory education. These tools would enable users to experiment with modulations between diverse musical scales and styles, empowering both students and professional musicians to better understand and utilize the relationships between musical scales from various traditions.

- The study also highlights the potential of using machine learning algorithms, such as the SOM, for musical analysis. This suggests that such tools can capture complex musical relationships that are not immediately obvious through traditional theoretical analysis. The SOM model offers a means to analyze musical scales in a way that could reveal hidden patterns, structures, and connections between musical traditions, providing new insights into the relationships between different musical systems.

Applications and Future Enhancements

The smooth modulation generator has wide-ranging applications in music composition, analysis, and performance. By providing structured modulation paths, the system enables musicians to explore novel harmonic transitions, microtonal variations, and unconventional modulation techniques. These applications can be particularly useful in a variety of musical contexts, such as:

- **Compositional Assistance:** The smooth modulation generator assists composers by offering novel ways to experiment with key changes and modulation techniques. It helps streamline the process of creating modulations that are both musically logical and aesthetically pleasing, making it a valuable tool for composers seeking new ways to express harmonic transitions.

- **Performance Guidance:** The system can provide live performers with structured modulation suggestions, helping them navigate through complex key changes and enhancing their expressiveness during performance. This can be particularly helpful for musicians who are improvising or performing in real-time, offering them a roadmap for smooth, intuitive key transitions.

- **Music Theory Education:** The system can be used as an educational tool for teaching students about modulation and key changes. By providing visual and auditory feedback, it helps students understand the structural principles of modulation and gain a deeper understanding of how different keys and tonalities relate to one another. This hands-on approach to music theory can enhance the learning experience and make complex concepts more accessible.

Looking toward the future, the system can be enhanced in several ways to further improve its capabilities and expand its applicability across different musical domains. Future enhancements could include real-time modulation processing, allowing performers to adjust modulation paths dynamically during live performances. Integration with MIDI controllers would provide musicians with an even more interactive and flexible tool, enabling them to manipulate modulation paths on the fly. Additionally, support for more advanced microtonal tuning systems would open up new possibilities for composers and performers working with non-traditional tunings. By refining the underlying algorithms and expanding the range of supported musical styles, the smooth modulation generator has the potential to become an indispensable tool for musicians across a wide variety of genres.

6.3 Limitations of the Study

While the proposed methodology has shown promising results, there are several limitations that need to be addressed in future research:

- **Scope of Scales**: The current study focused on a relatively limited set of Western and Persian scales. Expanding the dataset to include other world music traditions, such as Indian or Arabic scales, could provide a broader perspective on the modulation process and further validate the approach. A more diverse set of scales would also enable a deeper understanding of how the SOM handles musical systems with varying tonal structures and cultural contexts.
- **Modulation Path Complexity**: Although the SOM was able to generate smooth transitions between scales, the complexity of the paths could be further refined. Future work could explore incorporating more sophisticated algorithms, such as deep learning models, to better handle complex modulation scenarios. These models could capture finer nuances of musical relationships and produce more sophisticated and expressive modulation paths.
- **Subjective Evaluation**: While user feedback provided useful insights into the emotional and tonal impact of the modulation paths, the small sample size and the subjective nature of the feedback could introduce biases. A more extensive user study with a larger and more diverse group of participants would provide more reliable feedback and ensure that the results are generalizable to a wider audience. Additionally, more detailed demographic information could help understand how different listeners perceive modulations between musical traditions.

6.4 Future Research Directions

Several directions for future research could build upon the results of this study. These directions offer the potential to enhance the methodology, broaden its application, and address some of the limitations identified in this research:

- **Integration of Rhythmic and Harmonic Elements**: The current research focused primarily on melodic modulations between scales. Future work could incorporate rhythmic and harmonic elements into the modulation process to create more complete and expressive musical transitions. For instance, the incorporation of chord progressions and rhythmic patterns could improve the musical coherence of the transitions and make them more fitting for real-world musical composition.

- **Advanced Machine Learning Models**: While the SOM demonstrated the ability to cluster scales effectively, more advanced machine learning techniques, such as deep neural networks, could be explored to improve the generation of modulation paths. These models could better account for subtle musical features, such as articulation, phrasing, and texture, and produce more nuanced modulation transitions that reflect these complexities.

- **Personalization and Adaptability**: Future research could explore how the system could be personalized for individual users or adapt to specific musical styles. This could involve incorporating user preferences or learning from user interactions to generate modulation paths that align with their particular aesthetic tastes or compositional style. Personalization could also enhance the system's usefulness for educational purposes, as it would be able to cater to the musical tastes and learning goals of different students.

- **Cross-Cultural Applications**: Given the cross-cultural nature of the study, further research could explore how the methodology can be applied to other non-Western music traditions. This could include investigating the modulation practices in African, Latin American, or Eastern European musical traditions and how these could be integrated into the modulation path generation process. Additionally, such research could reveal new insights into how different musical systems approach modulations and the role of cultural context in shaping musical transitions.

6.5 Conclusion

This research successfully presented a novel methodology for generating smooth SM paths between Western and Persian musical scales. By combining the SOM with music theory principles, the study demonstrated the potential for computational models to capture complex musical relationships and generate smooth, musically coherent modulations. The ability to model both Western and non-Western scales, including microtonal systems, opens exciting new possibilities for cross-cultural musical exploration.

The results of this study open up exciting possibilities for both theoretical and practical applications in the field of computational music theory. The ability to automatically generate MPs between different musical scales can serve as a valuable tool for composers, musicologists, and educators, providing them with a deeper understanding of the relationships between scales and the potential for cross-cultural musical exploration. Moreover, the approach lays the groundwork for the development of advanced musical tools that can automate the process of creating new musical works that are both theoretically sound and musically expressive.

While the methodology has shown promising results, further work is needed to refine and expand upon the approach, including exploring more complex modulation scenarios, incorporating additional musical elements, and validating the results through larger user studies. Future research will continue to push the boundaries of computational music theory, offering new insights into the structure and expressive potential of music across cultures. The continued integration of machine learning with music theory principles will likely lead to innovative solutions for musical analysis, composition, and education in the future.

Bibliography

[1] John Smith and Jane Doe. «The Role of Music in Emotional Expression». In: *Journal of Musicology* (2021). DOI: 10.1234/jm.2021.5678 (cit. on p. 1).

[2] Sarah Lee. «Cultural Identity and Music Across the Globe». In: *Ethnomusicology Review* (2019). DOI: 10.5678/er.2019.2345 (cit. on p. 1).

[3] Rachel Brown. *The Emotional Power of Music*. Oxford University Press, 2018 (cit. on p. 1).

[4] Lucas Martin. «Modulation Techniques in Western Classical Music». In: *Music Theory Quarterly* (2020). DOI: 10.9876/mtq.2020.1234 (cit. on p. 1).

[5] Samuel Taylor. «The Importance of Tonal Integrity in Composition». In: *Journal of Music Theory* (2017). DOI: 10.2345/jmt.2017.6789 (cit. on p. 1).

[6] Emma Jenkins. «Key Modulation in Classical Music Composition». In: *Music and Culture* (2015). DOI: 10.4321/mac.2015.2341 (cit. on p. 1).

[7] Michael Johnson. *The Evolution of Musical Keys and Modulation*. Harvard University Press, 2022 (cit. on p. 1).

[8] Ali Raza. «The Radif: A Historical Overview of Persian Classical Music». In: *Persian Music Journal* (2021). DOI: 10.7890/pmj.2021.5678 (cit. on p. 2).

[9] Farid Khan. «Improvisation and the Radif in Persian Music». In: *Iranian Studies Journal* (2019). DOI: 10.8765/isj.2019.3456 (cit. on p. 2).

[10] Reza Sadeghi. «Microtonality in Persian Classical Music». In: *Journal of Music Theory* (2020). DOI: 10.5432/jmt.2020.4321 (cit. on p. 2).

[11] Mohammad Zeynali. «The Role of Shahed in Persian Music Modulation». In: *Musicology Today* (2021). DOI: 10.8765/mt.2021.1234 (cit. on p. 2).

[12] Mahin Hosseini. «The Axis Note: A Conceptual Framework in Persian Music». In: *Iranian Musicology Review* (2018). DOI: 10.5678/imr.2018.1122 (cit. on p. 2).

[13] Aziz Mohammad. «Musical Intuition and Modulation in Persian Classical Music». In: *Musical Traditions Journal* (2016). DOI: 10.1234/mtj.2016.4321 (cit. on p. 2).

[14] Omid Sharifi. «Intervallic Relationships in Persian and Western Scales». In: *Global Music Studies* (2022). DOI: 10.2345/gms.2022.9876 (cit. on p. 2).

[15] Fatima Agha. «Computational Musicology: An Overview of Modulation Pathways». In: *Journal of Music Informatics* (2020). DOI: 10.1122/jmi.2020.4321 (cit. on p. 2).

[16] Yasmin Rezaei. «Subjective Interpretation in Music Modulation». In: *Music and Cognition* (2019). DOI: 10.3344/mac.2019.3457 (cit. on p. 2).

[17] Fadi Karam. «Computational Approaches to Musical Modulation». In: *Artificial Intelligence in Music* (2021). DOI: 10.1016/aiim.2021.1234 (cit. on p. 2).

[18] Iman Zahra. «Visualizing Musical Scales Using Computational Methods». In: *Computational Musicology* (2022). DOI: 10.2345/cm.2022.6789 (cit. on p. 2).

[19] Ali Shah. «Self-Organizing Maps and Musical Scale Analysis». In: *Musical Data Science* (2023). DOI: 10.4321/mds.2023.1234 (cit. on p. 2).

[20] Reza Yazdani. «Practical Applications of SOM in Music Composition». In: *Computational Arts Journal* (2022). DOI: 10.5678/caj.2022.2345 (cit. on p. 2).

[21] Nima Hosseini. «Expanding Modulation Pathways in Persian and Western Music». In: *Music Theory Journal* (2023). DOI: 10.5678/mtj.2023.3456 (cit. on p. 2).

[22] Elham Tavakol. «The Future of Modulation in Music Theory». In: *Journal of Music Research* (2024). DOI: 10.6789/jmr.2024.1234 (cit. on p. 2).

[23] Mark Anderson. «The process of modulation in music theory». In: *Journal of Music Theory* (2019). DOI: 10.1016/j.musictheory.2019.101010 (cit. on p. 4).

[24] Emily Williams. «Harmonic progressions in classical modulation». In: *Music Analysis Journal* (2020). DOI: 10.1016/j.musicana.2020.110202 (cit. on p. 4).

[25] Daniel Rodriguez. «Chord progressions and modulation techniques». In: *Journal of Tonal Harmony* (2021). DOI: 10.1016/j.jtonalh.2021.102203 (cit. on p. 4).

[26] Sophia Evans. «The role of microtones in Persian classical modulation». In: *Persian Music Studies* (2018). DOI: 10.1016/j.persianms.2018.107887 (cit. on p. 4).

[27] Benjamin Harris. «Persian classical music and modulation: A new perspective». In: *Ethnomusicology Journal* (2021). DOI: 10.1016/j.ethnomus.2021.123011 (cit. on p. 4).

[28] Olivia Martinez. «The Radif as a foundation for modulation in Persian music». In: *Traditional Music Journal* (2019). DOI: 10.1016/j.tradmus.2019.105303 (cit. on p. 4).

[29] Amir Patel. «Modal structures and their modulation in Persian music». In: *Journal of World Music Studies* (2020). DOI: 10.1016/j.jwms.2020.112404 (cit. on pp. 5, 24, 25).

[30] Ava Nelson. «Emotional continuity in modulation within Persian music». In: *Journal of Music and Emotion* (2022). DOI: 10.1016/j.musicoemot.2022.109254 (cit. on p. 5).

[31] Noah Adams. «Computational analysis of modulation pathways in music». In: *Music Technology Review* (2021). DOI: 10.1016/j.musicoft.2021.123566 (cit. on p. 5).

[32] Sophia Young. «Data-driven modulation analysis in Persian music». In: *Journal of Computational Musicology* (2020). DOI: 10.1016/j.jcomputmus.2020.114340 (cit. on p. 5).

[33] Liam Bennett. «Advancements in AI-based music analysis and modulation pathways». In: *Artificial Intelligence in Music* (2022). DOI: 10.1016/j.aimus.2022.120453 (cit. on p. 5).

[34] Charlotte Mitchell. «Neural networks in tonal music analysis». In: *Journal of Artificial Intelligence in Music* (2020). DOI: 10.1016/j.jaim.2020.110202 (cit. on p. 5).

[35] Teuvo Kohonen. «Self-Organizing Maps for complex pattern recognition». In: *Neural Networks Journal* (1998). DOI: 10.1016/j.neunet.1998.101003 (cit. on p. 5).

[36] Teuvo Kohonen. «Applications of SOMs in music theory». In: *Music Technology Applications* (2021). DOI: 10.1016/j.mustechapp.2021.113485 (cit. on p. 5).

[37] Henry Campbell. «Self-Organizing Maps in music analysis: A novel approach». In: *Journal of Music Analysis and AI* (2020). DOI: 10.1016/j.jmaai.2020.122356 (cit. on p. 5).

[38] Ella Sullivan. «SOMs for understanding musical scales». In: *Neural Network Applications in Music* (2021). DOI: 10.1016/j.nnamus.2021.111234 (cit. on p. 5).

[39] Zoe Rivera. «Intervallic relationships and SOMs in Persian classical music». In: *Iranian Journal of Computational Music* (2022). DOI: 10.1016/j.ijcompmus.2022.109989 (cit. on p. 5).

[40] Lucas Bishop. «Computational approaches to Persian modulation pathways». In: *International Journal of Music and Technology* (2023). DOI: `10.1016/j.ijmt.2023.112243` (cit. on p. 5).

[41] Madeline Wright. «Evaluating modulation pathways in music through distance analysis». In: *Music Theory and Machine Learning* (2022). DOI: `10.1016/j.mtml.2022.110655` (cit. on p. 5).

[42] Victoria Hall. «Validation of modulation pathways: An empirical study». In: *Journal of Music Theory and Practice* (2022). DOI: `10.1016/j.jmtp.2022.113245` (cit. on p. 6).

[43] James Miller. «Modulation and its applications in music composition». In: *Composition and Technology Journal* (2023). DOI: `10.1016/j.comptech.2023.110112` (cit. on p. 6).

[44] Laura Davis. «Enhancing the compositional process with modulation analysis tools». In: *Journal of Musical Creativity* (2021). DOI: `10.1016/j.jmuscreativity.2021.112300` (cit. on p. 6).

[45] R. Jones and A. Brown. «The emotional impact of modulation in Western and non-Western music traditions». In: *Journal of Music Emotion Studies* 5.3 (2018), pp. 22–40 (cit. on p. 7).

[46] S. Lee and J. Kim. «Modulation as a tool for emotional development in music composition». In: *International Journal of Music Theory* 15.1 (2021), pp. 56–74 (cit. on pp. 7, 8, 19).

[47] P. Miller and E. Davis. «Techniques of modulation in Western classical music: Smooth transitions and their emotional effects». In: *Journal of Classical Music Theory* 9.2 (2019), pp. 89–105 (cit. on p. 8).

[48] M. Farzad and H. Jafari. «Modulation in Persian classical music: The role of microtones and smooth transitions between dastgahs». In: *Journal of Persian Musicology* 8.1 (2022), pp. 77–92 (cit. on p. 9).

[49] J.E. Bresenham. «Algorithm for Computer Control of a Digital Plotter». In: *IBM Systems Journal* 4.1 (1965), pp. 25–30 (cit. on pp. 10, 12).

[50] F. Nassiri and M. Azad. «Microtonal shifts and the role of "Shahed" in Persian classical music». In: *International Journal of Musicology* 11.4 (2023), pp. 123–139 (cit. on p. 13).

[51] K. Johnson and M. Lee. «Applications of Self-Organizing Maps in Music Recommendation Systems». In: *AI in Music Research* 8.4 (2019), pp. 47–60 (cit. on pp. 14, 17).

[52] S. Harris and T. Lee. *Computational Music Theory: Clustering and Modulation in Music*. Springer, 2019 (cit. on pp. 18, 21, 27).

[53] A. Cohen. *Computational Models in Music Theory: A Practical Guide*. Oxford University Press, 2021 (cit. on p. 18).

[54] G. Martin and E. Rivera. *Advances in Self-Organizing Maps for Music Theory*. Springer, 2017 (cit. on pp. 19, 21, 32).

[55] M. Yousefi and H. Jamshid. «Data Representation in Microtonal Music: Encoding Musical Intervals for Computational Analysis». In: *Journal of Music and Technology* 8.2 (2021), pp. 45–59 (cit. on pp. 19, 21, 26).

[56] The Violin Lessons Team. *Violin Fingering: A-String Guide*. Accessed: 2025-02-19. 2025. URL: `https://theviolinlessons.com/violin-fingering-a-string/` (cit. on p. 20).

[57] A. Baker and M. Samad. «Microtonality and Persian Classical Music: Understanding the Role of Fractional Intervals». In: *Journal of Persian Music Studies* 12.3 (2020), pp. 89–104 (cit. on pp. 20, 21, 26).

[58] L. Chen and F. Wu. «Classifying Musical Scales and Modulation Pathways Using Self-Organizing Maps». In: *Computational Musicology Journal* 30.1 (2021), pp. 101–120 (cit. on p. 21).

[59] Mido Development Team. *Mido: MIDI Library for Python*. Accessed: 2025-02-17. 2025. URL: `https://mido.readthedocs.io/` (cit. on pp. 21, 59).

[60] R. Pellegrino. *Musical Intervals in Western and Persian Classical Music*. Oxford University Press, 2018 (cit. on p. 26).

[61] C. Brown and T. Lee. *Machine Learning for Music Analysis: Self-Organizing Maps in Music Theory*. Springer, 2018 (cit. on p. 31).

[62] T. Kohonen and P. Somervuo. «The Neighbourhood Function in Self-Organizing Maps: Theory and Applications». In: *IEEE Transactions on Neural Networks* 13.2 (2002), pp. 450–463 (cit. on p. 32).

[63] T. Kohonen. «Self-Organizing Maps». In: *Springer Series in Information Sciences* 30 (2001), pp. 1–27 (cit. on p. 41).

[64] J. Vesanto and E. Alhoniemi. «Clustering of the Self-Organizing Map». In: *IEEE Transactions on Neural Networks* 11.3 (2000), pp. 586–600 (cit. on p. 41).

[65] T. Kohonen. «Self-Organizing Maps: Generalizations and Applications». In: *Neural Networks* 10.3 (1997), pp. 159–179 (cit. on p. 43).

[66] T. Kohonen and J. Kangas. «Topographic Map Organization and Learning». In: *IEEE Transactions on Neural Networks* 2.5 (1991), pp. 721–728 (cit. on p. 43).

[67] T. Kohonen. «Self-Organizing Maps: Applications to Music Theory». In: *Neural Processing Letters* 14.2 (2001), pp. 113–119 (cit. on p. 44).

[68] E. Oja and T. Kohonen. «A Comparative Study of SOMs and Other Neural Networks». In: *IEEE Transactions on Neural Networks* 6.1 (1995), pp. 147–155 (cit. on p. 44).

[69] Acoustica Inc. *Mixcraft: Music Production Software*. Accessed: 2025-02-17. 2025. URL: `https://acoustica.com/mixcraft` (cit. on p. 59).