Computing melodic templates in oral music traditions

Sergey Bereg [a], José-Miguel Díaz-Báñez [b], Nadine Kroher [b,*], Inmaculada Ventura [b]

[a] *Department of Computer Science, University of Texas at Dallas, 800 West Campbell Road Richardson, TX 75080, USA*
[b] *Department of Applied Mathematics II, University of Seville, Camino de los Descubrimientos s/n, Seville 41092, Spain*

A R T I C L E   I N F O

A B S T R A C T

The term *melodic template or skeleton* refers to a basic melody which is subject to variation during a music performance. In many oral music traditions, these templates are implicitly passed throughout generations without ever being formalized in a score. In this work, we introduce a new geometric optimization problem, the *spanning tube problem*, to approximate a melodic template for a set of labeled performance transcriptions corresponding to a specific style in oral music traditions. Given a set of n piecewise linear functions, we solve the problem of finding a continuous function, f^*, and a minimum value, ε^*, such that, the vertical segment of length $2\varepsilon^*$ centered at $(x, f^*(x))$ intersects at least p functions ($p \le n$). The method explored here also provide a novel tool for quantitatively assess the amount of melodic variation which occurs across performances.

© 2018 Elsevier Inc. All rights reserved.

1. Introduction

There exists a long tradition of using mathematic concepts to understand, model and create music. Ever since ancient Greek mathematicians have first formulated musical acoustics from a mathematical standpoint [1], music and mathematics have gone hand in hand (see i.e. [2] for an overview). As a result, a number of challenging multi-disciplinary fields have opened up, including algorithmic composition [3], ethnomathematics [4] and automatic music analysis [5].

A challenging issue in the mathematical research of music is to understand the nature of music. A deep discussion about the use of mathematics and computer science in music can be found in [6]. Since computers allow to efficiently process musical data, mathematical research in music has also provided theoretical foundations for music technology, an interdisciplinary science of computational description, analysis and processing of music and music data. An important subfield of music technology is music information retrieval (MIR), the scientific discipline of retrieving high-level information from music recordings. In particular, several optimization problems have been explored in the context of music analysis, including the modeling of rhythmic structures [7–9], harmony improvisation [10], melodic similarity [11] and the detection of melodic patterns [12].

This paper addresses a geometric problem inspired by musical characteristics found in *flamenco* music. *Flamenco* is a rich oral music tradition from southern Spain, which attracts a growing community of enthusiasts around the globe. The computational analysis of *flamenco* melodies, which has mainly been carried out in the MIR community, has revealed algorithmic challenges due to its high degree of ornamentation and improvisation [13–15]. We propose a geometric method to approximate the melodic template inherent to a set of *flamenco* performances. The term *melodic template* refers to a basic melodic line, which undergoes variation, i.e. in the form of ornamentation and embellishment, during a music performance.

---

* Corresponding author.
E-mail addresses: besp@utdallas.edu (S. Bereg), dbanez@us.es (J.-M. Díaz-Báñez), nkroher@us.es (N. Kroher), iventura@us.es (I. Ventura).
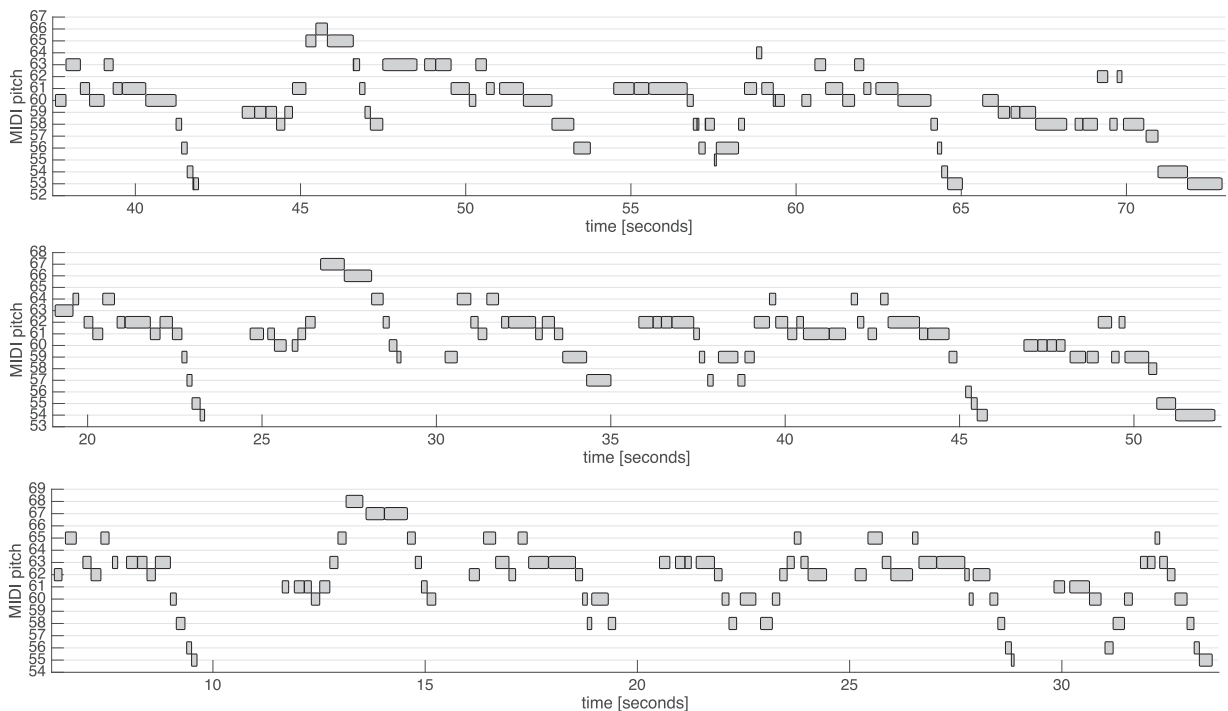
Fig. 1. MIDI transcriptions of three performances of the *fandango de Valverde* melody.

An example of MIDI transcriptions of the different interpretations of the same underlying template is depicted in Fig. 1. Similar concepts are found in *Arab-andalusian* [16], *Carnatic* [17] and *Shizou* [18] music.

Mathematically speaking, we are interested in approximating a set of n piecewise linear functions $f_i, i = 1, \ldots, n$ representing melodies in a temporal-pitch domain by a polygonal curve $f*$ so that, in any time, $f*$ is similar to an acceptable amount of such melodies.

Related problems are polygonal approximation and fitting problems. The aim of polygonal approximation is to approximate a complex polygonal curve by a simpler polygonal curve so that total approximation error is minimized for a given error metric. This is an important task in computer vision, computer graphics, digital cartography, statistics, and data compression. Polygonal approximation is an old problem and it has been well-studied from various perspectives as geographic information systems [19]; digital image analysis [20,21]; and computational geometry [22]. On the other hand, the fitting problems addresses the approximation of a discrete set of points by a polygonal curve or step function with a bounded number of line segments minimizing a fitting measure [23,24].

Unlike polygonal approximation and fitting problems, the input of our problem is a set of polygonal curves instead of a single polygonal curve or a set of points, and the constraint of the output is given by the number of neighboring curves at any time in the time-pitch domain. To the best of our knowledge, this problem has not been considered in the literature and it is motivated by some potential applications in the automatic indexing and analysis of oral music traditions. In the sequel, we outline several such applications.

Comparing a performance to its underlying template, allows researchers to explore the creative process itself, model expressiveness and study the evolution of improvisation and interpretation [25]. While these aspects have been extensively studied in popular, jazz and classical music, existing approaches rely on the existence of the template as a musical score. However, in oral music traditions, scores are usually not available and rare manual transcriptions refer to a particular performance instead of the underlying melodic line. Being able to automatically approximate a template based on a set of performance transcriptions is a fundamental requirement for large-scale comparative performance analysis for oral music traditions

Another potential application for automatically generated melodic templates is melody classification, where the task is to assign a label to an unknown melody according to its underlying melodic template. In the context of flamenco music, due to the unavailability of the template, this problem has been previously addressed in a supervised k-nearest neighbor classification scenario [26]: An unknown melody is tagged based on the labels of the k most similar melodies in a manually annotated dataset. However, the necessary pair-wise comparisons with a number of melodies at runtime represent a major limitation for scalability. Comparing to automatically extracted templates instead, can decrease the runtime complexity from k to a single comparison.
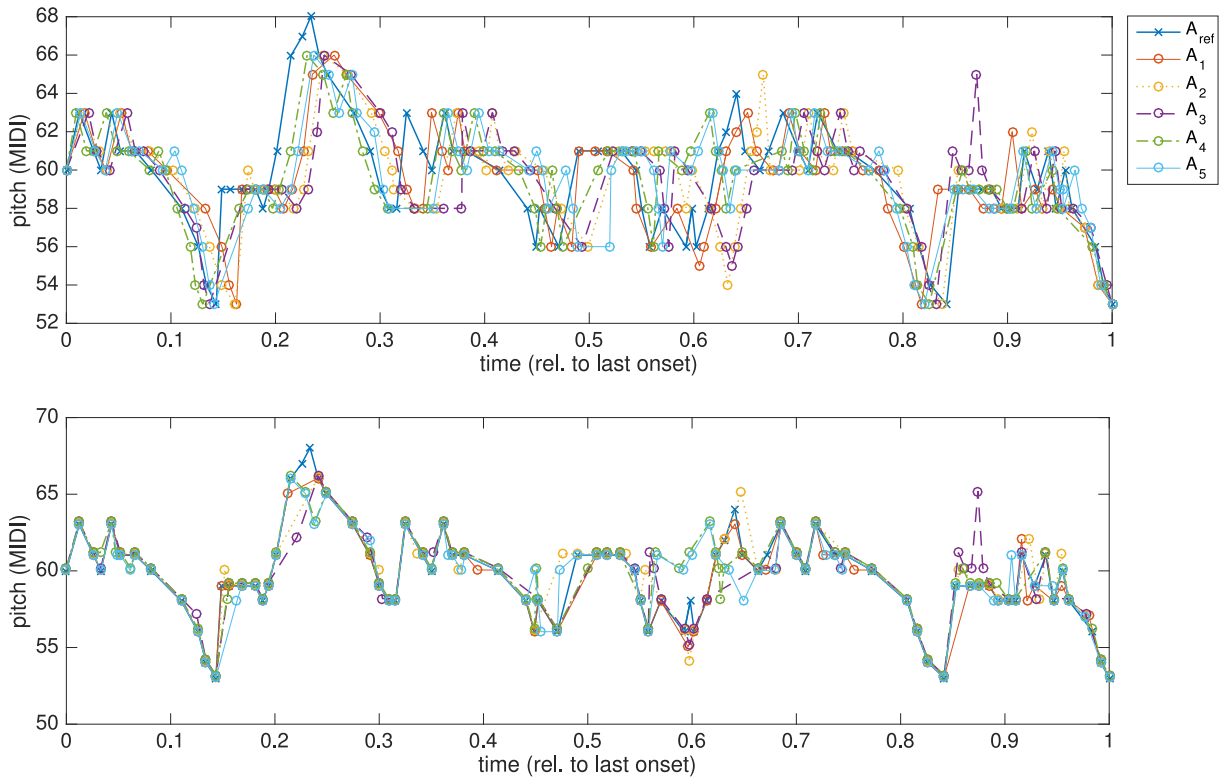
Fig. 2. Five performance transcriptions before (top) and after (bottom) temporal alignment.

The structure of the paper is as follows. We discuss the melody modeling and data preprocessing in Section 2. In Section 3 we introduce a new problem called the *Spanning Tube Problem* and develop a method for solving it. In Section 4 we apply our method for a comparative performance analysis of four *flamenco* styles which belong to the same style but differ in their melodic contour. Finally, we summarize our research and future directions in Section 5.

2. Melody modeling: dataset and preprocessing

In the scope of this study, we analyze a corpus of 40 flamenco commercial recordings belonging to 4 variants of the *fandango de Huelva* style. All interpretations of a variant are characterized by a common melodic skeleton which is subject to ornamentation and variation during performance. An automatic transcription tool [27] was used to extract a note representation of the singing voice melody and transcription errors were subsequently corrected by flamenco experts.

As a first preprocessing step, we convert each transcription, which is a discrete set of notes, described by their onset time, duration and MIDI pitch value, to a point set representation $A = \{a_1, a_2, \ldots, a_m\}$ where a note $a_i = \{(x_i, y_i)\}$ is characterized by its onset time relative to the last note onset, $x_i \in [0, 1]$, and its MIDI pitch value, y_i. Given that the transcription tool does not provide any rhythmic or metric quantization (which in flamenco, even when done manually, is a non-trivial task), the relative representation of the onset time is chosen to compensate for strong tempo variation among performances.

Furthermore, variants of the same melody may be performed in different keys. Singers usually select a key according to their individual pitch range. In order to meaningfully compare the pitch values of a set of performances, we therefore need to perform key normalization. Here, we apply a commonly used method for this task taken from [28]: One transcription is randomly selected as a reference and histogram of pitch occurrences h_{Ref} is computed. For each of the remaining transcriptions, we compute the pitch histogram in the same way and compute the correlation coefficient with h_{Ref} for different pitch shifts. Finally, for each recording, the pitch shift is applied, which yields the highest correlation coefficient.

In addition to key transposition, variants of the same melody may exhibit strong rhythmic distortions. We therefore perform a rhythmic normalization as follows. Given n transcriptions of a variant, we select a random transcription A_{ref} as a temporal reference to which we align all remaining transcriptions. This temporal alignment is realized via the *Needleman–Wunsch* algorithm [29] which finds an optimal matching with gaps among two sequences. We use linear interpolation to assign an onset time for notes which have not been matched in the procedure. Figure 2 shows an example of five transcriptions before and after the temporal alignment. This procedure runs in $O(kn^2)$ time.
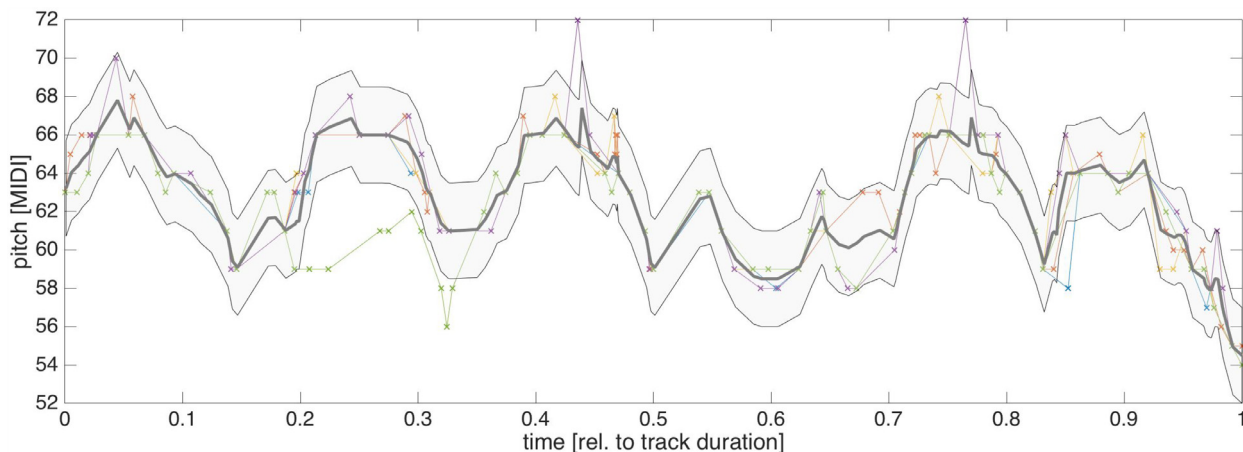
Fig. 3. The ε−tube for *Fandango de Calaña (n=5, p=4, ε = 2.5)*.

3. The geometrical problem

In order to estimate a melodic template, we aim to find a continuous function of time, which represents a set of n temporally aligned performance transcriptions. We consider that the data under this study may contain local outliers so some notes could be far from the template. Furthermore, our goal is to quantitatively assess the amount of melodic variation which occurs across performances. Geometrically, we view each of n temporally aligned performance transcriptions as a polygonal curve in the plane with the time axis and the pitch axis. Since they are monotone with time, we call them *x-monotone* curves. Typically, the performance transcriptions are discrete and we make an assumption that the curves are polygonal. The objectives and considerations give rise to a geometrical problem, which we call *Spanning Tube Problem*. In this Section we describe an approach for solving the spanning tube problem.

Definition 1. Given $a, b \in \mathbb{R}$ and a continuous function $f(x)$ with domain $[a, b]$, we define the ε-tube of f, $T(f, \varepsilon)$, as the locus of points (x, y) s.t. $(x, y) \in [a, b] \times [f(x) - \varepsilon, f(x) + \varepsilon]$. An example is illustrated in Fig. 3.

The Spanning Tube Problem (STP): *Let $a, b \in \mathbb{R}$, with $a < b$; let $n, m, p \in \mathbb{Z}^+$; and for $i = 1, \ldots, n$, let $f_i : [a, b] \to \mathbb{R}$ be a piecewise linear function with at most m links. Given $p \leq n$, find minimum $\varepsilon^* > 0$ such that there exists a continuous function $f^*(x)$ fulfilling that, for each $x \in [a, b]$ the vertical segment of length $2\varepsilon^*$ centered at $(x, f^*(x))$ intersects at least p functions.*

Remark 2. Note that the p intersected functions are not necessarily the same in any time and the template f^* has to be continuous and *x*-monotone.

Definition 3. Given a collection of n continuous functions f_i, let f_{up} be the upper and f_{low} be the lower envelope enclosing the n functions. Let f_{median} be the point-wise median of the f_i's.

Remark 4. For $n = p$, $\varepsilon^* = \max \frac{f_{up}(x) - f_{low}(x)}{2}$ and $f^*(x) = \frac{f_{up}(x) + f_{low}(x)}{2}$ for $x \in [a, b]$.

3.1. Solving the general problem

The following result can be obtained by using a left-to-right linear sweeping:

Theorem 5. *The decision problem for STP can be solved in $O(n^2 m \log n)$ time.*

Proof. We are given n functions $f_1(x), f_2(x), \ldots, f_n(x)$, $x \in [a, b]$ and two parameters $p \in \{1, 2, \ldots, n\}$ and $\varepsilon > 0$. Our task is to decide whether $\varepsilon < \varepsilon^*$, $\varepsilon = \varepsilon^*$ or $\varepsilon > \varepsilon^*$. This can be accomplished by sweeping the slab between lines with equation $x = a$ and $x = b$ by a vertical line as follows: we maintain two a sorted lists. Let L_f be the list of functions $f_i(x) \pm \varepsilon$ intersecting the sweep line and let L_e be the list of events where the sweep line stops. In the initial state, L_f is set to $f_i(a) \pm \varepsilon$ and L_e to $f_i(a)$. There are two types of events in L_e, see Fig. 4:

Vertex event. The sweeping line at a vertex event has an equation $x = c$ where one of the functions $f_i(x)$ changes from one linear function to another. Each given function contributes one vertex event to the list of events.

Crossing point event. At this event one of the functions $f_i(x) - \varepsilon$, $f_i(x) + \varepsilon$ intersects one of the functions $f_j(x) - \varepsilon$, $f_j(x) + \varepsilon$, for some $i < j$.

On the other hand, using L_f we store a number *Count(int)* for each interval *int* between two consecutive functions crossing the sweep line that indicates the number of intervals of form $[f_i(x) - \varepsilon, f_i(x) + \varepsilon]$ intersecting *int*. If *Count(int)* $\geq p$ we also
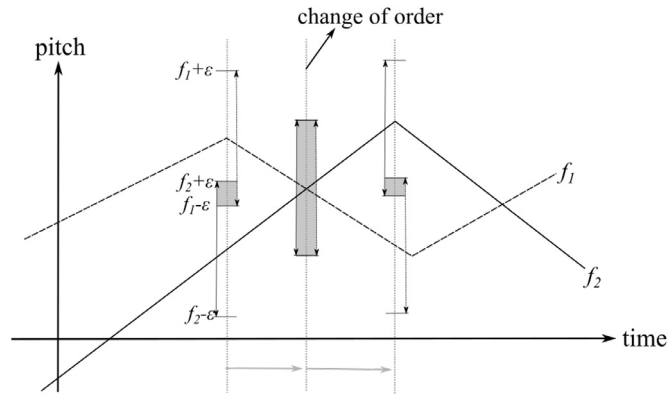
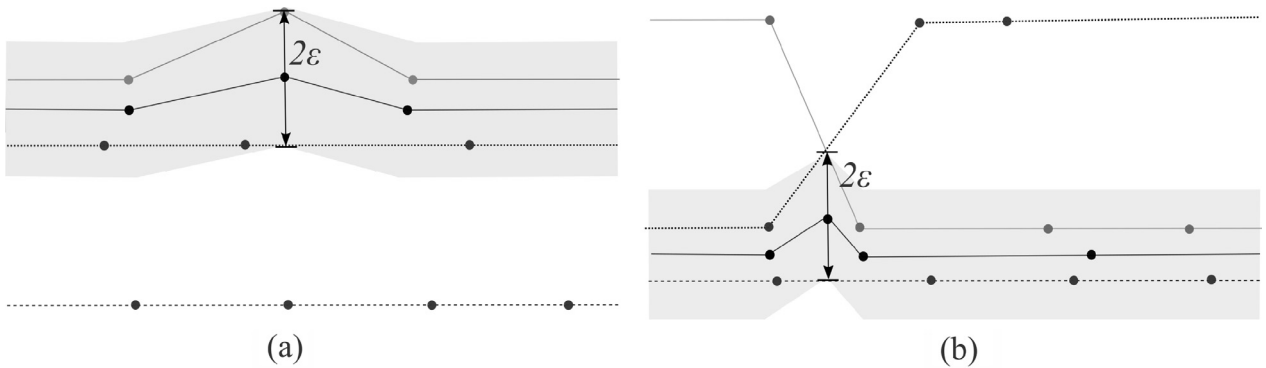**Fig. 4.** The events: Vertex and crossing point events.

Fig. 5. Two scenarios for determining the candidate values: (a) 2ε corresponds to the distance between a vertex of a function f_i and the value of another function f_j at the same time instance; (b) 2ε corresponds to the distance between the intersection of two functions f_i and f_k and the value of other f_j at the same time.

store a boolean $Ind(int)$ indicating whether there is a continuous function $f(x)$ such that $\forall x \in [a, x_0]$ there are at least p functions in $T(f, \varepsilon)$ where x_0 is determined by the current position of the sweep line.

We also maintain the ranges of form $[f_i(x) + s_i\varepsilon, f_j(x) + s_j\varepsilon]$, where $s_i, s_j \in \{-1, 1\}$, containing values of y such that, for current $x = c$ and any y in the range, there is a continuous function $f(x)$ with domain $[a, c]$ and $f(c) = y$.

Note that the continuity of the required function is guaranteed by computing the values of $Count(int)$ and $Ind(int)$.

Observe that L_e can be updated in $O(\log n)$ time and the order of functions $f_i(x) \pm \varepsilon$ is fixed between two events. For any fixed i and j, there are at most $O(m)$ crossing points events. The total number of events is $O(n^2 m)$ and the list of events has size $O(n)$. Since each event can be processed in $O(\log n)$ time, the running time is $O(n^2 m \log n)$. □

Based on the above result, we can apply bisection for computing an approximate solution. However, if we dispose of a discrete set of candidate values for ε^*, an exact solution can be found. The following lemma gives us the discrete set of candidate values for the optimization problem (STP).

Definition 6. We define a vertex of a piecewise linear function as a vertex-point and an intersection between two functions as an intersection-point.

Lemma 7. For $n > 2$ and $1 < p < n$, let ε^* be the optimal value for the optimization problem. Then, $2\varepsilon^*$ is equal to one of the values given by the vertical distance between a vertex- (Fig. 5 (a)) or an intersection-point (Fig. 5 (b)) and its $(p - 1)$ or p-nearest function.

Proof. Let ε^* be the minimum value for which there exists a continuous function $f^*(x)$ such that $\forall x \in [a, b]$ there are at least p functions in $T(f^*, \varepsilon^*)$. Then, there are two functions f_i and f_j in $T(f^*, \varepsilon^*)$ and $x_0 \in [a, b]$ such that $f_i(x_0) = f^*(x_0) + \varepsilon^*$ and $f_j(x_0) = f^*(x_0) - \varepsilon^*$. Otherwise, we can decrease the width of the tube contradicting optimality. It is easy to prove that the intersection points between the functions f_i and f_j and the boundary of the tube must have the same abscissa. There are three cases: the point $u = (x_0, f_i(x_0))$ is a vertex-point, a crossing point, or none of the above. Analogously for $v = (x_0, f_j(x_0))$.

Suppose that u and v are neither a vertex nor a crossing point. Then up to symmetry, we have two cases as in Fig. 6 (a) and (b). In both cases, the width of the tube can be slightly decreased maintaining continuity. This fact contradicts the
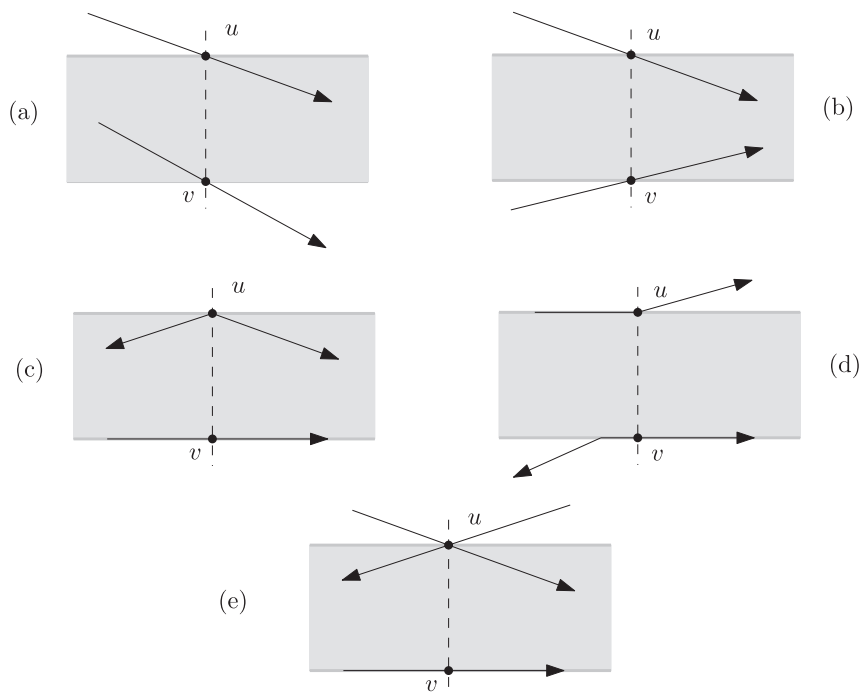
Fig. 6. (a) and (b): u and v are neither vertex nor crossing points. (c) and (d): u is a vertex point. (e): u is a crossing point.

optimality condition. In case shown in Fig. 6 (a) we can move the endpoints u and v on f_i and f_j respectively maintaining vertical the segment uv in the direction where the length of the vertical segment decreases. Then, there exists an instant in which a vertex or crossing point is found, so this case reduces to the vertex or crossing point event. Note that even when f_i and f_j are parallel, theses events are found if we maintain at least p functions inside the tube. A similar reasoning can be done in the second case (in Fig. 6 (b)): since the part of the slab to the left from the vertical segment uv must have at least p functions, then in the right part there are at least $p + 2$ functions so there exists a narrower tube containing at least p functions.

Consider now the case in which u is a vertex-point of $f_i(x)$ (w.l.o.g. u lies on the top boundary of the tube). Two cases arise, depending on the slopes of the neighboring segments, see (c) and (d) in Fig. 6.

Case 1. The function $f_i(x)$ is in the tube for x in some neighborhood of x_0 as in in Fig. 6 (c). Then $f_j(x)$ is the $p - 1$-nearest function of u.

Case 2. For $\delta > 0$ and $x \in (x_0, x_0 + \delta)$ or $x \in (x_0 - \delta, x_0)$, $f_i(x)$ is not in the tube as in Fig. 6(d). The function $f_i(x)$ on the top boundary comes out the tube at time x_0 and the function $f_j(x)$ is the p-nearest function.

Finally, the case in which u is a intersection-point can be easily reduced to the vertex-point type, see Fig. 6(e). □

Theorem 8. *For $n > 2$ and $p > 1$, the optimization problem can be solved in $O(n^2 m \log n \log nm)$ time.*

Proof. The steps of the algorithm are: (I) Compute the candidate values; (II) Sort the candidates and (III) Perform binary search using the decision algorithm. Step (I) can be done by sweeping the arrangement of the n functions with a vertical line. In this status line we maintain the order of the intersection of the line with the functions (it can be vertex-points, intersection-points and functions) so that we can compute in $O(1)$ the corresponding $p - 1$- or p-nearest functions (one above and other below) for every candidate. The overall complexity of the sweep is $O(n^2 m \log n)$. Step (I) can be done in $O(n^2 m \log n)$ time, Step (II) requires $O(n^2 m \log nm)$ time and Step (III) in $O(n^2 m \log n \log nm)$ time. □

3.2. An efficient solution for a particular case

The following result reduces the number of candidate values for $n = 3$ and $p = 2$.

Lemma 9. *Set $n = 3$, $p = 2$ and let ε^* be the optimal value for the optimization problem. Then the value of $2\varepsilon^*$ is one of these values:*

- *a local maximum value of $f_{up} - f_{median}$ or*
- *a local maximum value of $f_{median} - f_{low}$ or*
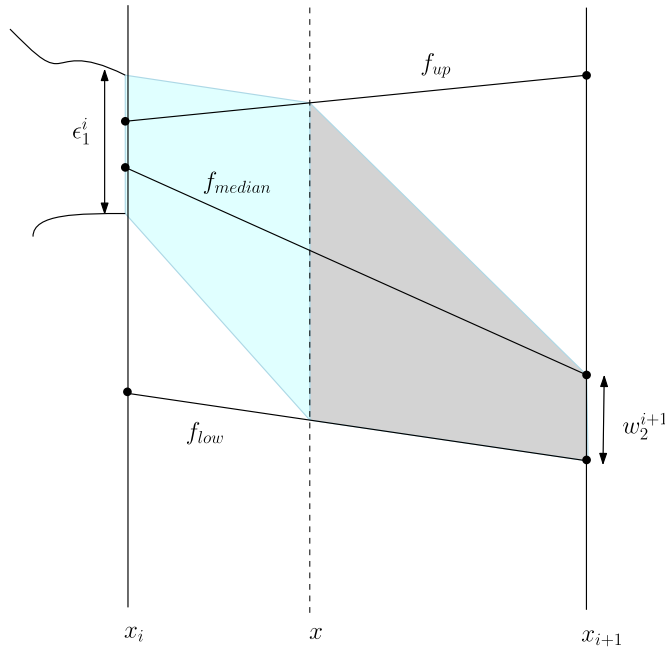- *a local minimum value of $f_{up} - f_{low}$.*

Fig. 7. Two trapezoids: The tube makes a transition in the ith slab.

Proof. For $\varepsilon > 0$, let $R_1(\varepsilon) = T(f_{up}, \varepsilon) \cap T(f_{mean}, \varepsilon)$ and let $R_2(\varepsilon) = T(f_{low}, \varepsilon) \cap T(f_{mean}, \varepsilon)$. Observe that, a solution exits for ε if and only if there exists a monotone function inside the region $R_1(\varepsilon) \cup R_2(\varepsilon)$. Now, a solution exists for $R_1(\varepsilon^*) \cup R_2(\varepsilon^*)$ but no within $R_1(\varepsilon) \cup R_2(\varepsilon)$ for any $\varepsilon < \varepsilon^*$. Note that $R_i(\varepsilon) \subset R_i(\varepsilon^*)$ for $i = 1, 2$. We have 3 cases:

- Two points in $R_1(\varepsilon^*)$ are connected by a function but are not connected in $R_1(\varepsilon)$. Then, for some $x^* \in (a, b)$, we have $f_{up}(x^*) - \varepsilon = f_{mean}(x^*) + \varepsilon$ and $f_{up}(x) - \varepsilon < f_{mean}(x) + \varepsilon$ for all $x \neq x^*$ in some neighborhood of x^*. Therefore $f_{up} - f_{mean}$ has a local maximum value at x^* with value $2\varepsilon^*$.
- Two points in $R_2(\varepsilon^*)$ are connected by a function but are not connected in $R_2(\varepsilon)$. Then, for some $x^* \in (a, b)$, we have $f_{mean}(x^*) - \varepsilon = f_{lower}(x^*) + \varepsilon$ and $f_{mean}(x) - \varepsilon < f_{low}(x) + \varepsilon$ for all $x \neq x^*$ in some neighborhood of x^*. Therefore $f_{mean} - f_{low}$ has a local maximum value at x^* with value $2\varepsilon^*$.
- Two points, one in $R_1(\varepsilon^*)$ and one in $R_2(\varepsilon^*)$, are connected in $R_1(\varepsilon^*) \cup R_2(\varepsilon^*)$ but are not within $R_1(\varepsilon) \cup R_2(\varepsilon)$. Then, for some $x^* \in (a, b)$, we have $f_{up}(x^*) - \varepsilon = f_{low}(x^*) + \varepsilon$ and $f_{up}(x) - \varepsilon > f_{low}(x) + \varepsilon$ for all $x \neq x^*$ in some neighborhood of x^*. Therefore $f_{up} - f_{low}$ has a local minimum value at x^* with value $2\varepsilon^*$. □

By Lemma 9, we have $O(m)$ candidates and performing binary search the optimal solution can be computed in $O(m \log m)$ time by using the decision algorithm. We will improve the running time by proving some properties first.

Definition 10. Let $x_0 = a, x_1, \ldots, x_t = b$ be the events defined by vertices and intersection points of the functions. Consider a slab between two events at x_i and at x_{i+1}, $i = 0, \ldots, t - 1$. Suppose that a spanning tube T in $[x_0, x_{i+1}]$ covers $f_{up}(x_i)$, $f_{median}(x_i)$ and $f_{low}(x_{i+1})$, $f_{median}(x_{i+1})$. We say that the tube makes a *transition HL* in the ith slab.

Definition 11. Let ϵ_1^i be the width of a spanning tube in $[x_0, x_{i+1}]$ covering $f_{up}(x_i)$ and $f_{median}(x_i)$ and let analogously ϵ_2^i be the width of a spanning tube covering $f_{median}(x_{i+1})$ and $f_{low}(x_{i+1})$, $i = 0, \ldots, t - 1$. We furthermore use w_1^i to denote the distance between $f_{up}(x_i)$ and $f_{median}(x_i)$ and w_2^i to denote the distance between $f_{median}(x_i)$ and $f_{low}(x_i)$ (see Fig. 7).

Lemma 12 (Two trapezoids). *Suppose a spanning tube T makes a transition HL in an i–th slab, $i = 0, \ldots, t - 1$. Then there exists $x \in [x_i, x_{i+1}]$ such that $f_{up}(x), f_{median}(x)$ and $f_{low}(x)$ are covered by T. Furthermore, the width of T is at least*

$$w = \max(\varepsilon_1^i, f_{up}(x) - f_{low}(x), w_2^{i+1}) \tag{1}$$

and there exist a tube of width w using two trapezoids in the slab, see Fig. 7.

Proof. The existence of x can be seen using continuity of the tube. Since all 3 functions are linear in the slab, the first trapezoid covers f_{up} and f_{median}. Similarly the second trapezoid covers f_{median} and f_{low}. The lemma follows since the width of the tube formed by two trapezoids is w. □

Lemma 13 (Transition). *Let $w^i = f_{up}(x_i) - f_{low}(x_i), i = 0, \ldots, t$. The smallest width of a spanning tube T making a transition HL in an i–th slab is equal to $\max(\varepsilon_2^i, \min(w^i, w^{i+1}), w_1^{i+1})$. The instant $x \in [x_i, x_{i+1}]$ where the transition takes place can be decided by comparing w^i and w^{i+1}. Thus, there exists an optimal tube having all transitions at the events only.*
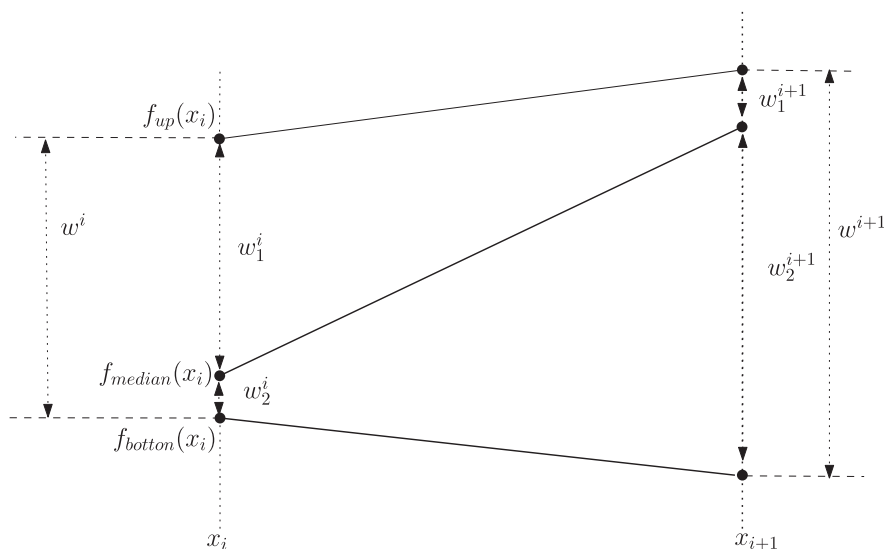
Fig. 8. An instance for updating the optimal tube in the slab between the events x_i and x_{i+1} for $n = 3$, $p = 2$.

Proof. This can be seen by changing x between x_i and x_{i+1}. The smallest value of the second term in Eq. (1) is $\min(w^i, w^{i+1})$. Thus, if $f_{up}(x)$ and $f_{low}(x)$ are not parallel, we can decide the line for transition ($x = x_i$ or $x = x_{i+1}$ using the smallest value of w^i and w^{i+1}). If they are parallel, one can choose line $x = x_i$ (or $x = x_{i+1}$) for the transition. □

Now, we are ready to show that we can solve the problem in linear time.

Theorem 14. *The optimization problem for $n = 3$ and $p = 2$ can be solved in $O(m)$ time.*

Proof. Let $x_0 = a, x_1, \ldots, x_t = b$ be the events defined by vertices and intersection points of the functions. For each event at $x = x_i$, we compute ε_1^i and ε_2^i for the spanning tubes in $[x_0, x_i]$ covering $[f_{up}(x_i), f_{median}(x_i)]$ and $[f_{median}(x_i), f_{low}(x_i)]$, respectively. At the beginning, $\varepsilon_j^0 = w_j^0$ for $j = 1, 2$. Then ε_1^{i+1} can be computed using two cases where the tube covers (i) $f_{up}(x_i)$ or (ii) $f_{low}(x_i)$. Then ε_1^{i+1} is the minimum of the two. The second value is computed using Lemma 13 and $w^i = f_{up}(x_i) - f_{low}(x_i)$, see Fig. 8. Thus,

$$\varepsilon_1^{i+1} = \min(\max(\varepsilon_1^i, w_1^{i+1}), \max(\varepsilon_2^i, \min(w^i, w^{i+1}), w_1^{i+1})). \tag{2}$$

Similarly,

$$\varepsilon_2^{i+1} = \min(\max(\varepsilon_2^i, w_2^{i+1}), \max(\varepsilon_1^i, \min(w^i, w^{i+1}), w_2^{i+1})). \tag{3}$$

Finally, $\varepsilon^* = \min(\varepsilon_1^t, \varepsilon_2^t)$.

Running time. For each event $i + 1$, the values of $w^{i+1}, w_1^{i+1}, w_1^{i+2}$ (see Fig. 8) and the values of $\varepsilon_1^{i+1}, \varepsilon_2^{i+1}$ (by Eqs. (2) and (3)) can be computed in constant time. Since we have $O(m)$ events, the overall running time is linear. □

Remark 15. Note that Theorem 14 improves the complexity obtained in Theorem 8 for $n = 3$.

4. Case study: Quantifying melodic variation

In this Section, we exemplify the use of the proposed method in a comparative performance analysis of the four *fandango* styles investigated in this study. In particular, we use the proposed system to quantify the amount of melodic variation the skeleton is subjected to during performance, compare across styles and analyze the local variation on a phrase-level.

From a music theoretic standpoint we know that among the four styles under study, the *Fandangos de Calaña* and the *Fandangos de Valverde* are close to their folkloric origin and performers tend to largely preserve the melodic skeleton during performance. In the *Fandangos Valientes de Huelva* and the *Fandangos Valientes de Alosno* on the other hand, performers tend to use heavy melodic ornamentation as an artistic asset, resulting in a more distorted skeleton.

The proposed method allows us to quantify the amount of melodic variation occurring in a set of performance transcriptions by fixing the parameter ε and determining in a decision problem the maximum percentage of performances $\frac{p}{n}$ which can be enclosed by the tube. Here, we compute this value for each of the four styles under study, varying ε between 1.0 and 3.0.

The largest differences among styles can be observed for $\varepsilon = 3.0$. Note, that in this case, the tube covers 6 semitones, which corresponds to half an octave. Consequently, melodic segments outside the tube correspond to relatively large deviations from the basic melodic contour. For this case, the results (Table 1) confirm the musicological considerations described

Table 1
Fraction of performances enclosed by the tube for different styles.

| Style	$\varepsilon = 1.0$	$\varepsilon = 2.0$	$\varepsilon = 3.0$
Fandango de Calaña	0.2	0.5	0.8
Fandango de Valverde	0.1	0.3	0.5
Fandango Valiente de Alosno	0.1	0.3	0.3
Fandango Valiente de Huelva	0.0	0.1	0.1
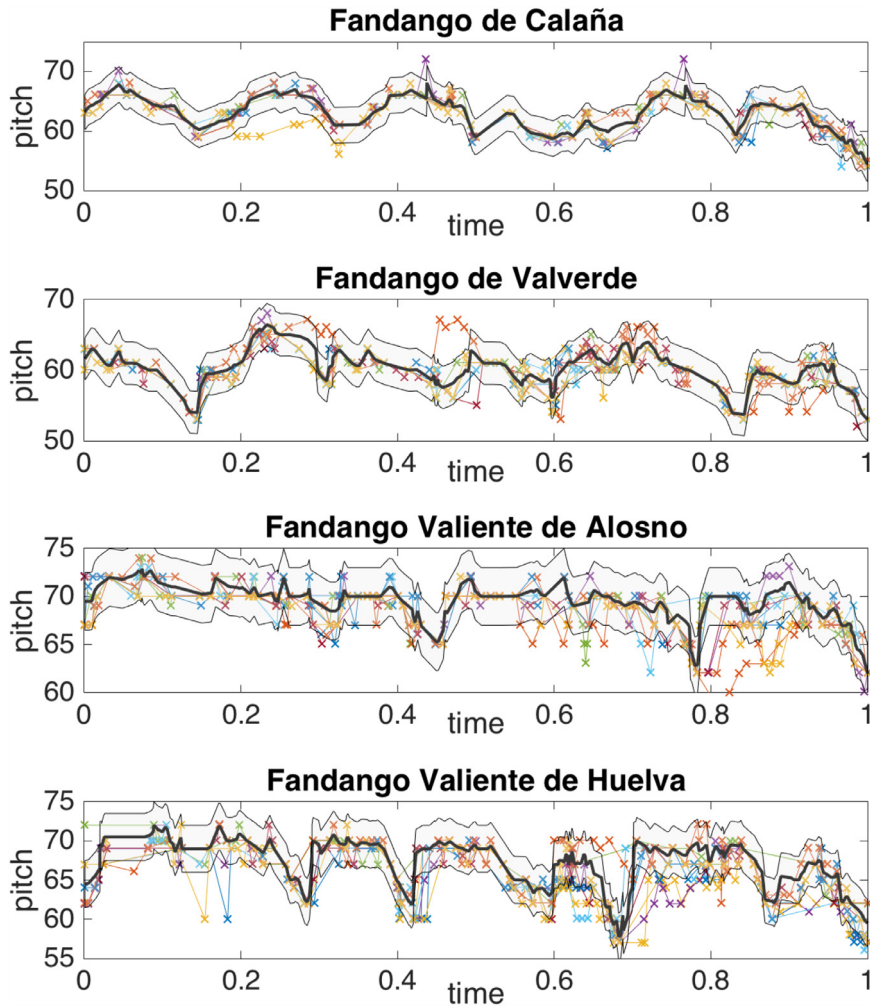
Fig. 9. Aligned transcriptions and computed skeleton for the four *fandango* styles under study for $\varepsilon = 3.0$.

above: For the *Fandangos de Calaña*, 80% of the analyzed performance transcriptions can be enclosed by the tube with $\varepsilon = 3.0$, indicating that performers largely follow the underlying skeleton. The value for the *Fandangos de Valverde* is slightly lower with $\frac{p}{n} = 0.5$. The two *valiente* styles show a significantly higher amount of melodic variation. For the *Fandangos Valientes de Huelva* only 10%, and for the *Fandangos Valientes de Alosno* only 30% of the transcriptions are enclosed in the tube. Fig. 9 shows the aligned transcriptions together with the computed templates using the values for $\frac{p}{N}$ from Table 1. A similar trend can be observed for smaller values of ε.

However, analysing the transcriptions in relation to the maximum spanning tube (Fig. 9), reveals that there exist local differences in the amount of occurring variation. For example, for the *Fandangos Valientes de Alosno*, all transcriptions are located inside the 3ε-tube from the beginning of the melody until approximately 0.3 on the relative time axis. In order to obtain a finer granularity of the amount of occurring variation, we therefore repeat the previous experiment on a phrase level. More precisely, each of the recordings contains 6 musical phrases, which we manually annotated. Fixing $\varepsilon = 3.0$, we solve the optimization problem and compute $\frac{p}{n}$ for each phrase separately.

Table 2
Fraction of performances enclosed by the tube per phrase for different styles and $\varepsilon = 3.0$.

Style	Phrase 1	Phrase 2	Phrase 3	Phrase 4	Phrase 5	Phrase 6
Fandango de Calaña	0.5	0.7	0.7	0.5	0.7	0.4
Fandango de Valverde	0.5	0.5	0.2	0.2	0.3	0.5
Fandango Valiente de Alosno	0.3	0.4	0.2	0.1	0.5	0.1
Fandango Valiente de Huelva	0.3	0.7	0.5	0.2	0.3	0.1
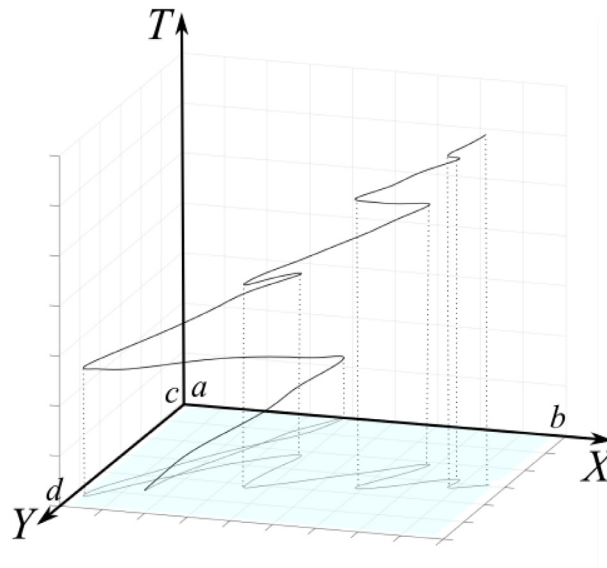
Fig. 10. The orthogonal projection of f onto the XY plane is the gesture. f is a T-monotone curve.

The results in Table 2 show, that with exception of the *Fandango de Valverde*, the last phrase tends to exhibit a high amount of variation compared to other phrases. This phrase, which is referred to as *caída* in flamenco jargon, represents the highlight of the flamenco performance, and consequently, it is likely that performers use a larger amount of ornamentation and variation as an expressive asset. For all styles, we furthermore observe a relatively high amount of variation for the fourth phrase.

These observations are of interesting from a musicological viewpoint and give rise to several lines of study related to melodic variation and expressiveness in flamenco music.

5. Conclusions and future research

Motivated by musical properties typically encountered in the analysis of oral traditions, we introduced in this paper a new geometric optimization problem, the Spanning Tube Problem (STP). We model n melodies as polygonal curves in the time-pitch space with at most m vertices. The aim is to compute a new polygonal curve, the template, which fits a fixed number p of similar items. The particular challenge here is, that the parameter p corresponds to an amount of melodies, but does not refer to a specific set of melodies. We solve the optimization problem by performing binary search on a discrete set of candidate values, for which we solve the corresponding decision problem. The obtained time complexity is $O(n^2 m \log n \log nm)$. We also prove that the particular case $n = 3$, $p = 2$ can be solved in linear time with a more elaborate approach. Finally, we perform an experimental study with flamenco melodies demonstrating how the resulting STP can be employed in a comparative performance analysis to quantify the amount of variation among a set of melodies.

There are several immediate suggestions for further research.

- The first issue is related to the complexity. Can the asymptotic complexity of the problem be improved? For large data sets, the time complexity of our approach is roughly cubic and we ask if a more detailed study allows us to improve the algorithm, as we did for the case $n = 3$ and $p = n - 1 = 2$. Since p could be close to n, it is even interesting to find a $O^*(n(n - p)m)$-time algorithm.
- In the STP, the width of the tube is the same at any point in time. However, we observed in the transcriptions that the melodic variability is not constant. Consequently, a future research could target a STP with variable width. Observe that the simple idea of intersecting the optimal tube with the polygon between the upper and lower hulls of the functions gives us a more accurate visualization of the variability.

- Another possible variant is to restrict the number of vertices of the template. In fact, for highly ornamented melodies, the template appears to be very complex and a simpler prototype could better model the underlying melodic movement.
- Another interesting question is to efficiently solve the reverse problem: Given an $\varepsilon > 0$, compute the maximum number p^* of melodies that can be captured by an ε-tube. Note, that our algorithm solves this problem in $O(n^2 m \log n \log p)$ using binary search. Is it possible to improve the running time using a different approach?
- Finally, the extension of our problem to 3D leads to an interesting task: the recognition of user-defined temporal gestures from tactile interfaces. Given a set of n curves (gestures) f_i in the *XY* plane, we want to compute a new continuous curve approximating the set of curves f_i. Note that the gestures f_i can be represented by the orthogonal projection of three dimensional curves in the *XYT* space that are monotone on time. Consequently, this problem can be seen as a generalization of the Spanning Tube Problem, see Figure 10. The definition in 2D can be extended to 3D as follows:

The 3D STP: *Let $a, b, c, d, t \in \mathbb{R}$, with $a < b, c < d$ and $t > 0$; let $n, m, p \in \mathbb{Z}^+$; and for $i = 1, \ldots, n$, let $f_i : [0, t] \to \mathbb{R}^2$ be a T-monotone piecewise linear function with at most m links. Given $p \leq n$, find minimum $\varepsilon^* > 0$ such that there exists a T-monotone continuous function $f^*(t) = (f_1(t), f_2(t))$ fulfilling that, for each $t \in [0, T]$ the ball (for a L_p distance $(L_2, L_1, L_\infty, \ldots)$) of radius ε^* centered at $(t, f^*(t))$ intersects at least p functions.*

Acknowledgments

References

[1] R.L. Crocker, Pythagorean mathematics and music, J. Aesthet. Art Crit. 22 (2) (1963) 189–198.
[2] K. Vaughn, Music and mathematics: modest support for the oft-claimed relationship, J. Aesthet. Educ. 34 (3) (2000) 149–166.
[3] G. Nierhaus, Algorithmic Composition: Paradigms of Automated Music Generation, Springer Science & Business Media, 2009.
[4] M. Chemillier, Ethnomusicology, ethnomathematics. The logic underlying orally transmitted artistic practices, in: Mathematics and Music, Springer, 2002, pp. 161–183.
[5] C.B. Madden, Fractals in Music: Introductory Mathematics for Musical Analysis, High Art Press, 1999.
[6] A. Volk, A. Honingh, Mathematical and computational approaches to music: challenges in an interdisciplinary enterprise, J. Math. Music 6 (2) (2012) 73–81.
[7] G. Toussaint, Computational geometric aspects of rhythm, melody, and voice-leading, Comput. Geom. 43 (1) (2010) 2–22.
[8] O. Aichholzer, L.E. Caraballo, J.M. Díaz-Báñez, R. Fabila-Monroy, C. Ochoa, P. Nigsch, Characterization of extremal antipodal polygons, Graphs Comb. 31 (2) (2015) 321–333.
[9] L. Barba, L.E. Caraballo, J.M. Díaz-Báñez, R. Fabila-Monroy, E. Pérez-Castillo, Asymmetric polygons with maximum area, Eur. J. Oper. Res. 248 (3) (2016) 1123–1131.
[10] E. Valian, S. Tavakoli, S. Mohanna, An intelligent global harmony search approach to continuous optimization problems, Appl. Math. Comput. 232 (2014) 670–684.
[11] R. Typke, P. Giannopoulos, R.C. Veltkamp, F. Wiering, R.V. Oostrum, Using transportation distances for measuring melodic similarity, in: Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), 2003.
[12] A. Lubiw, L. Tanur, Pattern matching in polyphonic music as a weighted geometric translation problem, in: Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), 2004.
[13] F. Gómez, J. Mora, E. Gómez, J.M. Díaz-Báñez, Melodic contour and mid-level global features applied to the analysis of flamenco cantes, J. New Music Res. 45 (2) (2016) 145–159.
[14] A. Pikrakis, N. Kroher, J.M. Díaz-Báñez, Detection of melodic patterns in automatic flamenco transcriptions, in: Proceedings of the 6th International Workshop on Folk Music Analysis (FMA), 2016, pp. 14–17.
[15] N. Kroher, A. Pikrakis, J.M. Díaz-Báñez, Discovery of repeated melodic phrases in folk singing recordings, IEEE Trans. Multimed. 20 (6) (2018) 1291–1304.
[16] N. Kroher, A. Chaachoo, J. Díaz-Báñez, E. Gómez, F. Gómez-Martín, J. Mora, M. Sordo, Springer handbook of systematic musicology, in: Computational Ethnomusicology: A Study of Flamenco and Arab-Andalusian Vocal Music, Springer-Verlag, Berlin Heidelberg, 2018, pp. 885–897.
[17] P. Rao, J.C. Ross, K.K. Ganguli, V. Pandit, V. Ishwar, A. Bellur, H.A. Murthy, Classification of melodic motifs in raga music with time-series matching, J. New Music Res. 43 (1) (2014) 115–131.
[18] A.R. Thrasher, The melodic structure of jiangnan sizhu, Ethnomusicology 29 (2) (1985) 237–263.
[19] D.H. Douglas, T.K. Peucker, Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, Cartograph.: Int. J. Geogr. Inf. Geovisualization 10 (2) (1973) 112–122.
[20] H. Imai, M. Iri, Computational-geometric methods for polygonal approximations of a curve, Comput. Vis. Graph. Image Process. 36 (1) (1986) 31–41.
[21] J.D. Hobby, Polygonal approximations that minimize the number of inspections, in: Proceedings of the fourth annual ACM-SIAM Symposium on Discrete algorithms, 1993, pp. 93–102.
[22] W.S. Chan, F. Chin, Approximation of polygonal curves with minimum number of line segments or minimum error, Int. J. Comput. Geom. Appl. 6 (01) (1996) 59–77.
[23] J.M. Díaz-Báñez, J.A. Mesa, Fitting rectilinear polygonal curves to a set of points in the plane, Eur. J. Oper. Res. 130 (1) (2001) 214–222.
[24] H. Fournier, A. Vigneron, Fitting a step function to a point set, Algorithmica 60 (1) (2011) 95–109.
[25] G. Widmer, W. Goebl, Computational models of expressive music performance: the state of the art, J. New Music Res. 33 (3) (2004) 203–216.
[26] N. Kroher, J.M. Díaz-Báñez, Audio-based melody categorization: exploring signal representations and evaluation strategies, Comput. Music J. 41 (4) (2018) 64–82.
[27] N. Kroher, E. Gómez, Automatic transcription of flamenco singing from polyphonic music recordings, IEEE/ACM Trans. Audio Speech Lang. Process. 24 (5) (2016) 901–913.
[28] P. van Kranenburg, A computational approach to content-based retrieval of folk song melodies, Utrecht University, 2010 (Ph.d. thesis).
[29] S.B. Needleman, C.D. Wunsch, A general method applicable to the search for similarities in the amino acid sequence of two proteins, J. Mol. Biol. 48 (3) (1970) 443–453.