A segmentation-based prototype to compute string instruments fingering

Daniele Radicioni  
Centro di Scienza Cognitiva, Università di Torino, Italy  
nadiccion@di.unito.it

Luca Anselma  
Dipartimento di Informatica, Università di Torino, Italy  
anselma@di.unito.it

Vincenzo Lombardo  
Dipartimento di Informatica, CSC, CIRMA, Università di Torino, Italy  
vincenzo@di.unito.it

In: R. Parnscott, A. Kessler & F. Zimmer (Eds.)  
Proceedings of the Conference on Interdisciplinary Musicology (CIM04)  
Graz/Austria, 15-18 April, 2004  
http://gewi.uni-graz.at/~cim04/

Background in automatic music performance. Several representational levels are involved in the performance process [Dannenberg, 1993]: a performance environment should be concerned at least in i) getting a score in input, ii) analyzing it, like a human performer would do, iii) modelling the constraints posed by body-instrument interaction, iv) manipulating sound parameters. While there exist some satisfactory solutions at level iv) (e.g. the guitar physical model described in [Cuzzucoli, Lombardo, 1999]), levels ii) and iii) still pose unanswered questions. In this paper we address the problem of fingering, that deeply affects the physical and expressive qualities of the sound being produced (cf. [Parnscott et al. 1997] and [Gillardino, 1975a, 1975b] in the case of keyboards and guitar, respectively).

Background in computing. Parnscott et al. (1997) propose an ergonomic model for short fragments of keyboard music. The algorithmic approach in [Sayegh, 1989] proposes a computationally efficient solution (based on a shortest path algorithm) to the fingering process in string instruments.

Aims. We aim to show that cognitive-based and suitable fingerings for string instruments can be achieved by a computationally efficient algorithm.

Main contribution. Our work might be regarded as a first step in extending Sayegh’s approach in a cognitive direction: while Sayegh computes the optimum (i.e. the least difficult one) fingering considering the whole piece, we take into account a number of features concerning the musical intentions, that underlie the fingering decision process. In this paper we focus on a specific feature, namely the segmentation. Musical phrase grouping is a well known feature of general music cognition (cf. [Temperley, 2001]). The score is manually pre-processed to extract the phrases that form it. And it is on this constituent structure that the model minimizes the difficulties for the fingering process, first within a phrase and then between subsequent phrases. It computes fingering for melodies without any length limitations, by estimating difficulties that come from the physical features of the instrument and the morphology of the hand.

Implications. The preliminary experiments with the prototype result to be more cognitively reliable and closer to a human expert's performance with respect to the ones provided by a global optimization approach like Sayegh’s. These results support an integration among computer science techniques, cognitive modelling, music analysis and theory for a promising approach to the automatic music performance.

Music performance involves the transformation of symbolic representations of musical thought into physical actions in order to operate a music instrument (Moore, 1990). A model of music performance involves the interpretation of music scores and the application of gestures to some sound synthesis device. Gesture modeling is particularly applicable in combination with a physical model approach to synthesis, since
the latter allows a natural representation of the interaction with the performer.

This paper presents a computational model of the fingering process for a string instrument: the model, which is interfaced with a physical model of the classical guitar (Cuzzucoli & Lombardo, 1999), takes in input a score and returns a suitable fingering for that score. We have also implemented a prototype that has produced encouraging preliminary results. The approach models the fingering problem as a shortest path traversal in a graph generated from the score; the segmentation of the piece into meaningful phrases improves the basic results, according to the comparisons made with the data provided by a human expert.

**The problem**

Fingering is the process that starts from a sequence of notes and assigns each note both a position on the guitar fingerboard and one finger of the left hand. It significantly affects the technical and expressive qualities of the sounds being produced, in an automatic performance environment as well, since it is concerned with the parameters of the synthesis model that affect the rendition of the timbre and its reliability (Roads, 1996). Nonetheless, both traditional and contemporary pieces often lack of fingering indications considered unnecessary or an execution choice (Allorto et al., 1990).

Fingering is a complex cognitive process involving several competencies (Gillardino, 1975a, 1975b), such as philological analysis, that interprets the sequence of notes, according to structural features, physical constraints, posed by the musical instrument, and biomechanical constraints, posed by the body parts of the musician that interact with the instrument.

The fingering process can be computationally intractable, since the number of possible fingerings for a piece of music grows exponentially with the number of notes: given \( n \) notes and four available fingers, \( 4^n \) different fingerings arise. Moreover, since for the whole family of string instruments the same note may be found on up to four positions, this number increases up to \( 16^n \).

**Related work**

There are a few approaches to the fingering problem and related topics in the literature.

Parncutt et al. (1997), with a refinement of Jacobs (2001), propose an ergonomic model for fingering with the keyboard, restricted to melodic fragments played with one hand. The model features a two-step control strategy: generation of all the possible fingerings and then a ranking phase based on the application of preference rules that encode ergonomic difficulties.

The case of fingering for string instruments is more complex than for keyboards, since we can find notes with the same fundamental frequency on different strings, up to four. Sayegh (1989) proposes a solution to the fingering problem for string instruments based on the *optimum path paradigm*, which models the fingering task as a graph traversal, that minimizes the difficulties (e.g. penalizing changes of positions and changes of string) expressed as a cost function that labels the edges between vertices.

While Sayegh’s approach is related to some intuitive observations on difficulties, a behavioral study of the complexity of left-hand movements by Heijink & Meulenbroek (2002) has provided evidence that the position of the hand on the guitar fingerboard, the need to reposition the hand within a tone sequence and the required finger span are three significant complexity factors. All of them contribute to the fingering decision.

**Approach**

Since there is a considerable overlapping of the tone range of guitar strings (the same note can be found on different strings), it is necessary to introduce the notion of *note position* on a string, which guarantees a unique identifier for the correspondence between the note event and the event production on the sound synthesis device.

The guitar fingerboard is represented as a matrix of positions, where the rows represent the strings and the columns represent the frets. Each matrix entry (*position*) is a pair `<string,fret>`. Strings are numbered from I to
VI, while frets range from 0 (standing for “unfretted”, open string) to 17. As each position can be pressed by one of four fingers, we define a fingered position a pair consisting of a position and a finger, i.e. <position,finger>, also in the flat representation <string,fret,finger>. It clears up that what we call the fingering process actually is a one-to-one matching between the sequence of notes and an equal-length sequence of fingered positions.

Given a music score, the program generates a graph that represents all the possible fingered position sequences (Figure 1): specifically, each vertex represents a fingered position for a note and each edge connects two adjacent fingered positions. Each edge is labelled with a weight representing the cost of the transition between two fingered positions.

The generation of the graph proceeds as follows. For each note from left to right in the score we generate all the possible fingered positions by combining the available triples <string,fret,finger> for that note. The availability is due to physical constraints on the instrument and bio-mechanical limitations of the left hand (finger crossing, chord aggregates, ...). For each fingered position (i.e. triple <string,fret,finger>) we create a vertex labeled with the triple, and link the vertex with all the vertices related to the immediately previous note. The weight assigned to each edge results from an estimation of the cost of the transition from one fingered position to another. Weights estimating difficulties are based on (Gilardino, 1975a, 1975b; Heijink & Meulienbroek, 2002) and on a performer’s intuitive introspection; then they have been quantitatively tuned while testing the prototype (see below).

This construction procedure guarantees that the graph has a layered structure, in that the vertices can be grouped in layers, and all the edges connect vertices of adjacent layers (see Figure 1). The number of layers is equal to the number of notes in the score.

The fingering process is modeled as a graph search problem: any path from the leftmost

![Figure 1. Graph generated for a three notes (F2-F4-F3) fragment. Each node represents a fingered position. In order to assign a fingering to the three notes, a shortest path algorithm is used: for lack of space weights on the edges are omitted.](image)
layer to the rightmost layer represents a fingering for the score. In this setting, the problem of finding a suitable fingering for a piece corresponds to the problem of finding a shortest path in the graph. The generated graph is a directed acyclic graph (Lawler, 1976), and the layered structure of the graph ensures a shortest path traversal in linear time with the number of notes in input.

On the base of the theory of Rosenbaum and colleagues (1995), we expect that performers will choose the bio-mechanically easiest solutions for fingering, provided that other overriding cognitive constraints do not exist, like, e.g., the musical intention to achieve smooth sounds, that is particularly critical while playing plucked string instruments (Pujol, 1960). In the preliminary test reported below, we have compared the results of the prototype with the data from a human expert asked to provide a fingering that would best suit the aim of the ease of execution.

In the next two sections we illustrate how we estimate the weights on the edges and how to incorporate the segmentation issues in the model above.

**Estimating difficulties**

Underpinned by Heijink & Meulenkoek (2002), we assumed that moving hands horizontally -along the fingerboard- and vertically -across the fingerboard- causes qualitatively different amounts of difficulty. Moving along the neck, the performer’s hand is supposed to need a repositioning, whereas across the neck we only take into account finger displacements, that are considered less complex. This is due to the fact that movements accomplished while playing music are time constrained, and both hand repositioning and simple finger movements need to be performed in the same amount of time. Difficulties were arranged in two classes, estimating the *along the neck* (henceforth *ALONG*) and *across the neck* (*ACROSS*)

<table>
<thead>
<tr>
<th>MaxSpan</th>
<th>index</th>
<th>middle</th>
<th>ring</th>
<th>little</th>
</tr>
</thead>
<tbody>
<tr>
<td>little</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td>ring</td>
<td>3</td>
<td>0</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>middle</td>
<td>2</td>
<td>–</td>
<td></td>
<td>–</td>
</tr>
<tr>
<td>index</td>
<td>–</td>
<td></td>
<td></td>
<td>–</td>
</tr>
</tbody>
</table>

*Table 1. Maximum allowed distances between finger pairs: spans are expressed as number of frets.*

**Figure 2.** Difficulties are graduated between minimum/maximum span and comfortable span, while outside this interval the need of hand repositioning is verified.

*(ACROSS) difficulties.*

The difficulty between two fingered positions *i* and *j* is computed by means of:

\[
difficulty_{(i,j)} = ALONG + ACROSS
\]

that will label the corresponding edge in the graph.

For what pertains *ALONG*, a *comfortable*, a *minimum* and *maximum span* (Table 1) are provided for each pair of fingers, to estimate the difficulties inherent in a transition between two positions. Comfortable means that no effort is required to play these positions: it is expressed as the number of empty frets between two positions. In this case, no points are added at all to the score estimating the difficulty (of the fragment) computed so far. Minimum and maximum span are respectively the lower and upper bounds, expressed in frets, that an adult male performer would cover on a classical guitar with a given pair of fingers without repositioning his left hand. Within this range difficulties are graduated (Figure 2), by assigning heavier weights while getting near the extremities: if maximum or minimum span (that work as thresholds) is exceeded, a performer would need to reposition her/his hand. Since evidence was provided (Heijink & Meulenkoek, 2002) that hand repositioning increases the difficulty, it is consequently discouraged by penalizing such a transition.

Constraints grouped as *ACROSS* grasp difficulties originating from replacements of the left hand fingers on the fingerboard, and they are mainly concerned with avoiding unnatural postures, and with keeping the vertical finger span as close as possible.

The computation of the shortest path through the minimization of weights *ALONG* and *ACROSS*, yields the most comfortable way to play the notes in input: one might guess that
the fingering provided by such a strategy will often present empty strings (that are easier to play), and they also will be placed in preference around the first frets of the fingerboard.

**Segmentation**

A global optimization approach, such as the one outlined so far, is not cognitively reliable, since both human performers and listeners do break musical pieces into phrases, themes, motives, etc.. It is commonly accepted (Sloboda, 1985) that interpretations and expressive aspects of performance descend from the performer’s analysis of the musical pieces; furthermore the more salient metrical and grouping levels (namely *tactus* and phrases) override others, thus receiving emphasis throughout the performance (Palmer, 1997).

On the listener’s side, this segmentation was argued by Lerdhal & Jackendoff (1983), and then modeled in Temperley (2001), that implemented several preference rule systems evaluating possible structural analyses of a piece of music. On the performer’s side, the need of an analysis underlying the performance is both well present in the pedagogic literature (Bent, 1994), and confirmed in the respect of fingering by well-known virtuosi executors (Gillardino, 1975b). Performing a piece of music implies discerning and emphasizing structural features of the music such as structural boundaries; conventions such as slowing the tempo at the most important boundaries, are also proved in listener’s response (Sloboda, 2000).

Also studies in the skill acquisition in music performance (Drake & Palmer, 2000) reveal that long musical sequences are broken into short segments whose elements are taken into account at the same time, and that structural relationships such as phrase boundaries may be taken into account while segmenting the piece.

On these bases we conjectured that if performers aim at clarifying the score structure by means of segmentation, then they will likely employ fingering too in order to mark further those segment boundaries, being mainly concerned in “optimizing” fingering inside those margins.

Even though automatic analysis of music pieces is not a trivial task at all (see, e.g. Pardo & Birmingham, 1999), there have been a few attempts in the literature, such as Bod’s probabilistic parsing model (2001), or Maxwell’s expert system working essentially on harmonic textures (1992). At the current stage of development we produce a manual annotation of phrase boundaries.

The computational model has been updated by generating graphs for the individual phrases and finding the shortest path on each graph separately. Then, given the shortest paths that represent the fingering of two adjacent phrases A and B, the last vertex of the shortest path of A is connected with the first vertex of the shortest path of B.

**Preliminary test**

We examined the results of the fingering process as accomplished by a human performer and the prototype. The prototype was given twice the same score with phrasing annotated (I), and with no phrasing (II). On the basis of the cited literature, we make the following predictions.

Since phrase boundaries knowledge seems to be a relevant issue in music cognition both on the performer and the listener’s sides, we expect better results of the prototype in the case (I), i.e. the fingering produced by the prototype (I) should be closer than (II) to that provided by human expert, which is
assumed to be the “best” possible fingering. Differences should be explainable by higher order cognitive constraints, such as the non accounted musical intentions. In the second place, we expect that differences arising between (I) and (II) should be much salient on phrase boundaries, where path choices are less constrained.

Stimulus materials

The Rondò from the Sonata Op.6 n°2 of Francesco Molino (Götze, 1990) was simplified by authors (Figure 3) and reduced to a simple melody with no accompaniment, but preserving harmony and phrasing. The prototype yet does not account for chords. For the sake of clarity we have selected a piece from the late 18th century repertoire, because phrasing is neatly clarified by harmonic and melodic features. It is composed by three large phrases of 8 measures each, moving along a plain keynote – dominant – keynote schema. Phrases are articulated in 4+4 measures and each measure can be further divided into motives.

Human expert

One professional classical guitarist –also graduate in musicology of the Turin University- was interviewed. At first he was asked to examine the score, and then to highlight phrasing at various levels by means of slurs: the segmentation above mentioned (three phrases, starting with capital letters A, B, and A, each divided into subphrases and motives) was achieved. Except for what appears on the score, no further information concerning the music was given to the performer.

The following formula was read: please indicate by playing or writing on the score the best fingering you would use to perform this piece of music in a public execution. Next he was asked to assign a fingering to the phrase A either practicing on the guitar or directly on the score, as he liked: he started playing straight.

This sight-reading phase was video recorded as well, and it differs only slightly from the definitive one. In the end we transcribed the final fingering he produced. No time constraints were provided, but our explaining the task and his training till the last performance took less than half an hour.

Running the prototype

In the first run (I) the prototype processed the note sequence marked A on the score (of 38 note events), with the annotation of phrase boundaries.

In the second run (II) the same phrase A was fed to the prototype, without segmentation annotation.
Results
As evaluation metrics we record the differences accounting for positions \((p)\) \(<\text{string,fret}>\), and fingers \((f)\). A difference from the fingering suggested by the human expert has been marked when the finger involved was different but the position had been correctly located, while a greater difference was registered when the position was different, even if the finger was the same. Missing a position has been considered worse than missing a finger, since it implies the hand to lie elsewhere from the suggested position; given the standard tuning the guitar, positions corresponding to the same note can be found at least four-five frets away on close strings (and several frets if strings are not adjacent) one from another. Results are presented in Figure 4.

Fingering of the phrase A produced by (I) has minor differences with respect to the performer’s one, namely in a pair of finger swaps, thus providing a convincing fingering. (II) fell in seven finger swaps and two fully different positions. Also, all the errors in (I) and (II) occur at phrase boundaries.

To complete the assessment, it is remarkable that slight differences of (I) with the human fingering fall in spots were the human expert less easily decided for a fingering; those points (D, measure 1; F#, measure 3) fall at the beginning of or between motives (what we call caesurae), where motives are loosely coupled with previous ones.

Conclusion
In this paper we have addressed the fingering problem, that is a critical step for models of automatic music performance, especially those relying on physical modeling of sounds. A computational model based on a graph search representation of the problem has been presented: it minimizes the difficulties in order to ease the score execution. In a preliminary test, the model has produced similar results with respect to a human performer. Further research is needed to investigate strategies employed by human performers in order to explain how musical intentions, like smooth sounds, speed, phrasing fluency, are achieved.

References
keyboard fingerling for melodic fragments, Music Perception, 14(4), 341-382.


---

1 Fingers from index to little are used, since the left hand thumb works under the neck as counter-force opposed to the pressure of the fingers on the fretboard (Bobri, 1960).

2 It should be noticed, however, that the constant hidden in the linear complexity is given by the number of edges to be taken into account at each layer, in the worst case 256 (16²). This number is comparable with the length of the scores considered until now, and this result in an actual quadratic time.