Violin Fingering Estimation According to Skill Level based on Hidden Markov Model

Wakana Nagata
Nagoya Institute of Technology
nagata@mmsp.nitech.ac.jp

Shinji Sako
Nagoya Institute of Technology
sako@mmsp.nitech.ac.jp

Tadashi Kitamura
Nagoya Institute of Technology
kitamura@nitech.ac.jp

ABSTRACT

This paper describes a method that estimates the appropriate violin fingering pattern according to the player's skill level. A violin can produce the same pitch for different fingering patterns, which generally vary depending on skill level. Our proposed method translates musical scores into suitable fingering patterns for the desired skill level by modeling a violin player's left hand based on a hidden Markov model. In this model, fingering is regarded as the hidden state and the output is the musical note in the score. We consider that differences in fingering patterns depend on skill level, which determines the prioritization between ease of playing and performance expression, and this priority is related to the output probability. Transition probability is defined by the appropriateness and ease of the transitions between states in the musical composition. Manually setting optimal model parameters for these probabilities is difficult because they are too numerous. Therefore, we decide on the parameters by training with textbook fingering. Experimental results show that fingering can be estimated for a skill level using the proposed method. The results of evaluations conducted of the method's fingering patterns for beginners indicate that they are as good as or better than textbook fingering patterns.

1. INTRODUCTION

In a violin, the same pitch can be produced by several fingering patterns, and violin players decide which fingering pattern to use. Fingering decision is difficult for beginners because they lack experience. However, even if the player has considerable experience, fingering decisions often require trial and error, because the optimum fingering pattern for a specific transition is not easy to determine after only one try. Thus, automatic fingering estimation can help players at various skill levels.

The optimum fingering differs according to a player's skill level. For low-skill players, fingering that is easily played is optimum, whereas for high-skill players, fingering that allows the best performance expression is optimum. Thus, violin fingering estimation must be based on skill level.

Some studies have focused on fingering estimation for a plucked or bowed string instrument [1–5] or for the piano [6–8]. The methods proposed in these studies estimate the easiest fingering and cannot recommend fingering patterns for various skill levels. Other studies have transcribed fingering from audio [9, 10] or video [11]. In these transcriptions, performance expression is considered because of differences in human performances. However, these methods transcribe the fingering only from a recording of the performance.

Our objective is to estimate the fingering patterns for musical compositions according to the skill level of the violin player. The violin player decides whether playing needs to be easy or whether performance expression is appropriate. We also realize that this priority is influenced by the note length. If the note is short, ease of play becomes a higher priority because playing a succession of short notes is more difficult. When the note is long, expression has a higher priority because playing longer notes is easier. Expression also has a higher priority when the skill level is high. From this point of view, we previously proposed a fingering estimation method [12]. However, in that method, the model parameters are set manually, which, in addition to the highly complex model structure, result in difficulty tuning the parameters of the model.

In this paper, we model violin fingering using the concept underlying the hidden Markov model (HMM). The difference between our proposed model and HMM lies in the fact that the transition probability depends on the observation sequence. We regard fingering as the hidden state and the notes in the musical composition as the output. We define the priority of performance expression based on note length and skill level, and this priority is used to determine the output probability. Because note length also influences ease of transition from one fingering pattern to another, we define the degree of change between fingering patterns based on note length, and this degree of change is related to transition probability. Empirical determination of the numerous parameters required by output and transition probabilities is extremely difficult; therefore, most parameters are estimated from textbook fingering patterns. Our new method eliminates the need for manually setting the model parameters.

2. VIOLIN FINGERING

2.1 Strings

A violin has four strings, each of which has a different tone color and pitch range because of differences in thickness and material. The four strings are normally tuned to G3, D4, A4, and E5 in descending order of thickness. Because the pitch ranges of different strings overlap, the same pitch can be produced by pressing different strings at different positions.

2.2 Fingers

The strings are pressed using all fingers except the thumb. The four fingers are numbered from one (index finger) to four (little finger). A string that is played without being pressed is called an *open string* and is numbered zero. An open string has a different tone color because none of its vibration is absorbed by a finger.

2.3 Difference by Skill Level

As skills develop, the player can choose many strings and use *vibrato*, a pulsating change in pitch by slightly shaking the left hand while a left finger presses the string. Choosing string and using vibrato are important in performance expression.

Easy fingering and fingering considering performance expression are often different. For example, open string is easy because no pressing is done; however, the player cannot use vibrato because it cannot be produced with an open string. Thus, players use different fingerings depending on skill level.

3. VIOLIN FINGERING MODEL ACCORDING TO SKILL LEVEL

In this paper, we model violin fingering patterns similar to HMM, as shown in Figure 1. The model is somewhat different from HMM; the difference is that state transition depends on output sequence. The hidden state sequence $\boldsymbol{s}$ is the left hand state sequence, and output sequence $\boldsymbol{o}$ is the note and rest sequence in the score. We assume that the state changes for every note and that the state sequence is a Markov process.

To simplify problem, the model has the following restrictions: the score is monophonic, and only the factors pitch, note length, and rest length are considered by this model.

3.1 Hidden State

In order to ensure a unique correspondence between a single pitch and a state, we define the HMM's hidden state as the position of the left hand. The position of the left hand is described by the hand position and finger interval, as well as the string number and the finger number, which typically describe a fingering pattern. The hand position and finger interval should be steady; these elements are important when the appropriateness of the hand state in long span is being considered.

These four elements are represented by the following variables: the string number is x^{SN}, the finger number is x^{FN}, the hand position is x^{HP}, and the finger interval is x^{FI}. The hidden state is expressed as follows:

$$s_n = \{x_n^{\mathrm{SN}}, x_n^{\mathrm{FN}}, x_n^{\mathrm{HP}}, x_n^{\mathrm{FI}}\}. \tag{1}$$

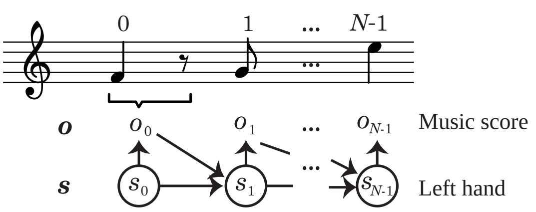


Figure 1. Model of a violin fingering pattern

3.1.1 String Number

We numbered the E string, A string, D string, and G string as 1, 2, 3, and 4, respectively.

3.1.2 Finger Number

We numbered the index finger, middle finger, ring finger, little finger, and open string as 1, 2, 3, 4, and 0, respectively.

3.1.3 Hand Position

The general position differs from the actual position depending on whether the note is natural, sharp, or flat. In this paper, we use the fret number of the index finger position under the assumption that a violin has frets. We assume that each string has 24 frets.

3.1.4 Finger Interval

We use the combination of the intervals between each finger. A finger not pressing a string does not have to be defined if only one note is played; however, we define all four fingers because even if a finger is not currently pressing a string, it is positioned according to the next note or on the basis of the previous note. We assume that the interval between each finger is a whole tone or a half tone. The number of combinations is 2^3.

3.2 Output Sequence

One hidden state corresponds to pitch p, note length l, and rest length r. In terms of fingering, note length l and rest length r define the priority of the performance expression (expressiveness e) and the ease of transition from the current fingering pattern to the next (changeableness c), as follows:

$$o_n = \{p_n, l_n, r_n\} = \{p_n, c_n, e_n\}. \tag{2}$$

3.2.1 Expressiveness

High-skill players can play with greater ease and expressiveness than low-skill players. However, even if skill level is high, expressiveness is relatively low when the note length is short. Conversely, even if skill level is low, expressiveness is high when the note length is long. Consequently, the same expressiveness can be achieved, even if the skill level is different, by changing the note length. Therefore, we determine expressiveness from both note length and skill level, as shown in Figure 2. Using this relation, we can estimate the appropriate fingering for any skill level in the unified framework.

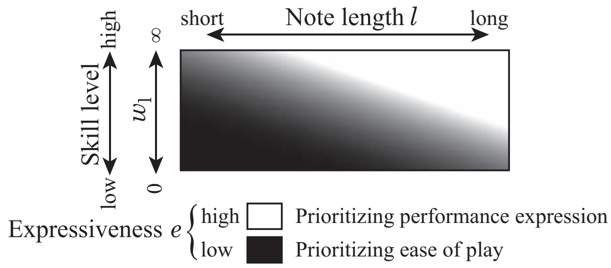


Figure 2. Relation among expressiveness, note length, and skill level

Based only on note length, the longer the note, the greater the ease and expressiveness it can be played with. We consider that variation of expressiveness is small when note length is overly long. Therefore, we use logarithms to describe the relation between note length and expressiveness.

On the other hand, based only on skill level, the higher the skill level, the greater is the expressiveness. The relation between expressiveness and skill level is assumed to be linear.

Thus, considering both note length l and skill level w_{l}, expressiveness e is determined as follows:

$$e_n = w_{\mathrm{l}} \log(1 + l_n). \tag{3}$$

When w_{l} is zero, which is the level for the least skilled players, expressiveness is always zero, regardless of note length.

3.2.2 *Changeableness*

We consider that changeableness is high when note length or rest length is high. First, the hand can easily change to the fingering pattern of the next note when the current note or rest is long. Second, the possibility that the current musical phrase is ending is high when the current note or rest is long. A musical phrase needs a similar tone throughout and smooth transitions of the fingers; therefore, it does not often require a significant transition of the finger positions.

Thus, changeableness c_n is decided by the current note length l_n and rest length r_n. Because the degree of influence of l_n and r_n on c_n are not equal, c is defined with weight w_{r}, as follows:

$$c_n = l_n + w_{\mathrm{r}} r_n, \tag{4}$$

where $w_{\mathrm{r}} > 1$.

3.3 Estimation of Optimum Fingering

Assuming that the optimal fingering has the highest likelihood among all the state sequences considered from a note sequence, the optimal fingering is given as follows:

$$\hat{\boldsymbol{q}} = \arg\max_{\boldsymbol{q}} \pi_{q_0} \prod_{n=1}^{N-1} a_{q_{n-1},q_n}(o_n) \prod_{n=0}^{N-1} b_{q_n}(o_n), \tag{5}$$

where $\boldsymbol{q}$ is the state number sequence corresponding to the note sequence, π is initial probability, a is transition probability, and b is output probability. Because the number of state sequences increases exponentially with the number of notes, calculating the likelihood of all state sequences is realistically difficult. However, searching for the maximum likelihood state sequence is solved based on the Viterbi algorithm.

4. HMM'S PARAMETERS

HMM requires the following parameters: initial probability, transition probability, and output probability. These are usually estimated using the Baum-Welch algorithm; however, our study includes many states, which makes the Baum-Welch algorithm inappropriate. Therefore, we consider the probability of each element individually, and postulate the distribution of the transition probability, which has many state combinations.

4.1 Initial Probability

Initial probability π_i is the probability of being in state s_i when note number n is zero. We define initial probability by assuming that the elements are independent of each other.

Each probability of string number, finger number, hand position, and finger interval is defined as $P_{\mathrm{SN}}(x_i^{\mathrm{SN}})$, $P_{\mathrm{FN}}(x_i^{\mathrm{FN}})$, $P_{\mathrm{HP}}(x_i^{\mathrm{HP}})$, and $P_{\mathrm{FI}}(x_i^{\mathrm{FI}})$, respectively. Initial probability π_i is formulated as follows:

$$\pi_i = P_{\mathrm{SN}}(x_i^{\mathrm{SN}}) P_{\mathrm{FN}}(x_i^{\mathrm{FN}}) P_{\mathrm{HP}}(x_i^{\mathrm{HP}}) P_{\mathrm{FI}}(x_i^{\mathrm{FI}}). \tag{6}$$

4.1.1 *String Number*

We consider that initial probability does not depend on the string. Therefore, $P_{\mathrm{SN}}(x_i^{\mathrm{SN}})$ is uniformly distributed among the four strings; that is, $1/4$.

4.1.2 *Finger Number*

We define $P_{\mathrm{FN}}(x_i^{\mathrm{FN}})$ as the appearance probability of x_{FN} in the training data under the condition that $x^{\mathrm{FN}} = 0$ is weighted by the reciprocal of the appearance probability of the pitch that an open string makes. This is because $x^{\mathrm{FN}} = 0$ can only produce the pitch of an open string; therefore, the appearance probability of $x^{\mathrm{FN}} = 0$ is low.

4.1.3 *Hand Position and Finger Interval*

We define $P_{\mathrm{HP}}(x_i^{\mathrm{HP}})$ and $P_{\mathrm{FI}}(x_i^{\mathrm{FI}})$ as the appearance probability of x_{HP} and x_{FI} in the training data, respectively.

4.2 Transition Probability

Transition probability $a_{i,j}(o_n)$ is the probability of the state changing from s_i to s_j when the note number changes from n to $n+1$. Transition probability must consider the appearance probability of the destination state and the probability of state change. The appearance probability of the destination state is calculated in the same way as the initial probability. The probability of state change is defined by the postulate of the probability distribution of variation. The relation between output o and the state transition is influenced only by changeableness c. When changeableness is high, the transition probability cannot be based on variation. Therefore, we assume that the dispersion of the probability distribution of variation depends on changeableness.

When the relation among elements is considered, the probability of the finger intervals depends on the hand position. Other elements are independent of each other. Therefore, the probabilities of the elements x^{SN}, x^{FN}, x^{HP}, and x^{FI} are defined as $P_{\mathrm{SN}}(x_j^{\mathrm{SN}}|s_i, c_n)$, $P_{\mathrm{FN}}(x_j^{\mathrm{FN}}|s_i, c_n)$, $P_{\mathrm{HP}}(x_j^{\mathrm{HP}}|s_i, c_n)$, and $P_{\mathrm{FI}}(x_j^{\mathrm{FI}}|s_i, c_n, x_j^{\mathrm{HP}})$, respectively. Transition probability $a_{i,j}(o_n)$ is formulated as follows:

$$\begin{aligned} a_{i,j}(o_n) \sim\ & P_{\mathrm{SN}}(x_j^{\mathrm{SN}}|s_i, c_n) \times P_{\mathrm{FN}}(x_j^{\mathrm{FN}}|s_i, c_n) \\ & \times P_{\mathrm{HP}}(x_j^{\mathrm{HP}}|s_i, c_n) \times P_{\mathrm{FI}}(x_j^{\mathrm{FI}}|s_i, c_n, x_j^{\mathrm{HP}}). \end{aligned} \tag{7}$$

The probability of each element is described as follows.

4.2.1 *String Number*

We consider that the probability of a string number depends only on movement range. In general, the string often remains the same for some notes. Therefore, distribution of transition of a string number is concentrated at

$x_i^{\mathrm{SN}} = x_j^{\mathrm{SN}}$, and we postulate that $P_{\mathrm{SN}}(x_j^{\mathrm{SN}}|s_i, c_n)$ is the following Laplace distribution:

$$P_{\mathrm{SN}}(x_j^{\mathrm{SN}}|s_i, c_n) \sim f_{\mathrm{Lap}}(x_j^{\mathrm{SN}}; x_i^{\mathrm{SN}}, k_1 c_n), \tag{8}$$

where f_{Lap} is the probability density function of the Laplace distribution as follows:

$$f_{\mathrm{Lap}}(x; \mu, \phi) = \frac{1}{2\phi} \exp\left(-\frac{|x-\mu|}{\phi}\right). \tag{9}$$

Because the dispersion of the Laplace distribution σ^2 is $2\phi^2$, we define k_1 to satisfy the following equation:

$$s^2 = 2(k_1 \bar{c})^2, \tag{10}$$

where s^2 is the dispersion of the training data's $x_n^{\mathrm{SN}} - x_{n-1}^{\mathrm{SN}}$ and $\bar{c}$ is the average of the training data's c.

4.2.2 Finger Number

We consider that the probability of a finger number depends only on the finger number of the target fingering pattern. Thus, $P_{\mathrm{FN}}(x_j^{\mathrm{FN}}|s_i, c_n)$ is defined as follows:

$$P_{\mathrm{FN}}(x_j^{\mathrm{FN}}|s_i, c_n) = P_{\mathrm{FN}}(x_j^{\mathrm{FN}}). \tag{11}$$

4.2.3 Hand Position

We consider that the probability of a hand position depends on both the hand position of the target fingering pattern and the movement range. We postulate that the probability of the movement range is a Laplace distribution for the same reason as that of the string number. Considering both the hand position of the target fingering pattern and the movement range, $P_{\mathrm{HP}}(x_j^{\mathrm{HP}}|s_i, c_n)$ is defined as follows with weight w_{HP}:

$$\begin{aligned} &P_{\mathrm{HP}}(x_j^{\mathrm{HP}}|s_i, c_n) \\ &\sim f_{\mathrm{Lap}}(x_j^{\mathrm{HP}}; x_i^{\mathrm{HP}}, k_2 c_n)^{w_{\mathrm{HP}}} P_{\mathrm{HP}}(x_j^{\mathrm{HP}})^{1-w_{\mathrm{HP}}}. \end{aligned} \tag{12}$$

Because the hand position change is easy when the current finger number is the open string ($x_i^{\mathrm{FN}} = 0$), k_2 is defined as follows:

$$k_2 = \begin{cases} k_{2,1} & (x_i^{\mathrm{FN}} = 0) \\ k_{2,2} & (x_i^{\mathrm{FN}} \neq 0) \end{cases}. \tag{13}$$

$k_{2,1}$ and $k_{2,2}$ are defined in the same way as k_1.

4.2.4 Finger Interval

The probability of the finger interval does not depend on the finger interval of the target fingering pattern if the hand position changes. Therefore, if $x_i^{\mathrm{HP}} \neq x_j^{\mathrm{HP}}$, $P_{\mathrm{FI}}(x_j^{\mathrm{FI}}|s_i, c_n, x_j^{\mathrm{HP}})$ is a uniform distribution; that is, $1/2^3$. If $x_i^{\mathrm{HP}} = x_j^{\mathrm{HP}}$, we consider that $P_{\mathrm{FI}}(x_j^{\mathrm{FI}}|s_i, c_n, x_j^{\mathrm{HP}})$ depends on both the finger interval of the target fingering pattern and the movement range. We define the movement range of a finger interval $M(x_i^{\mathrm{FI}}, x_j^{\mathrm{FI}})$ as the sum of the distance between the previous fret number and the current fret number for each finger, and postulate that the probability of the movement range is an exponential distribution because the distribution of transition of $M(x_i^{\mathrm{FI}}, x_j^{\mathrm{FI}})$ is concentrated at zero. Thus, $P_{\mathrm{FI}}(x_j^{\mathrm{FI}}|s_i, c_n, x_j^{\mathrm{HP}})$ is defined as follows:

$$\begin{aligned} &P_{\mathrm{FI}}(x_j^{\mathrm{FI}}|s_i, c_n, x_j^{\mathrm{HP}}) \\ &\sim \begin{cases} 1/2^3 & (x_i^{\mathrm{HP}} \neq x_j^{\mathrm{HP}}) \\ P_{\mathrm{Exp}}^{w_{\mathrm{FI}}} P_{\mathrm{FI}}(x_j^{\mathrm{FI}})^{1-w_{\mathrm{FI}}} & (x_i^{\mathrm{HP}} = x_j^{\mathrm{HP}}) \end{cases}, \end{aligned} \tag{14}$$

$$P_{\mathrm{Exp}} \sim \frac{1}{k_3 c_n} \exp\left(-\frac{M(x_i^{\mathrm{FI}}, x_j^{\mathrm{FI}})}{k_3 c_n}\right), \tag{15}$$

where k_3 is defined as follows for the same reason as that for the hand position:

$$k_3 = \begin{cases} k_{3,1} & (x_i^{\mathrm{FN}} = 0) \\ k_{3,2} & (x_i^{\mathrm{FN}} \neq 0) \end{cases}. \tag{16}$$

$k_{3,1}$ and $k_{3,2}$ are defined similar to k_1 for exponential distribution.

4.3 Output Probability

Output probability $b_i(o_n)$ is the probability that note o_n is outputted from state s_i. Pitch p and expressiveness e are independent of each other, and their probabilities are defined as $P_{\mathrm{p}}(p_n|s_i)$ and $P_{\mathrm{e}}(e_n|s_i)$, respectively. Output probability $b_i(o_n)$ is formulated as follows:

$$b_i(o_n) = P_{\mathrm{p}}(p_n|s_i) P_{\mathrm{e}}(e_n|s_i). \tag{17}$$

4.3.1 Pitch

$P_{\mathrm{p}}(p_n|s_i)$ is one if the output pitch from state s_i equals p_n and zero otherwise. Only one pitch results from a state.

4.3.2 Expressiveness

We consider that expressiveness relates only to the string number and the finger number. Because the string number x^{SN} and the finger number x^{FN} are independent of each other, $P_{\mathrm{e}}(e_n|s_i)$ is defined as follows:

$$P_{\mathrm{e}}(e_n|s_i) = \frac{P(e_n|x_i^{\mathrm{SN}}) P(e_n|x_i^{\mathrm{FN}})}{P(e_n)}. \tag{18}$$

The distributions of $P(e_n|x_i^{\mathrm{SN}})$ and $P(e_n|x_i^{\mathrm{FN}})$ are shown in Figure 3. Based on these distributions, we postulate that the distribution of expressiveness is a log-normal distribution as follows:

$$P(e_n|x_i^{\mathrm{X}}) \sim f_{\mathrm{ND}}(e_n; \mu_{\mathrm{X}, x_i^{\mathrm{X}}}, \sigma^2_{\mathrm{X}, x_i^{\mathrm{X}}}), \tag{19}$$

where X is SN or FN, and f_{ND} is the probability density function of the log-normal distribution as follows:

$$f_{\mathrm{ND}}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}x} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\}. \tag{20}$$

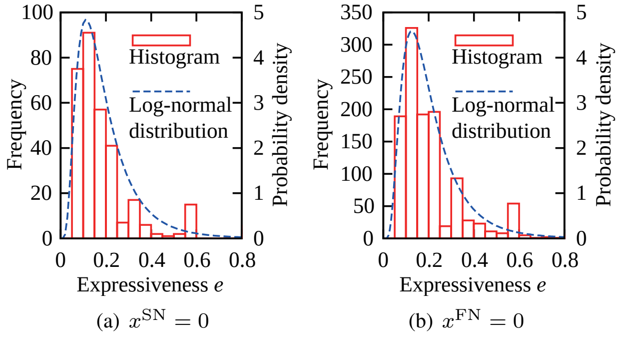


Figure 3. Examples of histogram and estimated log-normal distribution of expressiveness e

μ and σ^2 are defined by using maximum likelihood estimation as follows:

$$\mu_{\mathrm{X},k} = \frac{1}{|\boldsymbol{E}_{\mathrm{X},k}|} \sum_{e_i \in \boldsymbol{E}_{\mathrm{X},k}} \log e_i, \tag{21}$$

$$\sigma^2_{\mathrm{X},k} = \frac{1}{|\boldsymbol{E}_{\mathrm{X},k}|} \sum_{e_i \in \boldsymbol{E}_{\mathrm{X},k}} (\mu_{\mathrm{X},k} - \log e_i)^2, \tag{22}$$

where $\boldsymbol{E}_{\mathrm{X},k} = \{e_n | x_n^{\mathrm{X}} = k\}$, and $|\boldsymbol{E}_{\mathrm{X},k}|$ is the number of elements in set $\boldsymbol{E}_{\mathrm{X},k}$.

$P(e_n)$ depends only on output sequence $\boldsymbol{o}$ and does not influence the resulting state sequence $\boldsymbol{s}$. Therefore, we assume that $P(e_n)$ is an arbitrary positive fixed number, and disregard this number in the calculations.

5. EXPERIMENTS

We evaluated the proposed method based on two factors: the concordance rate between textbook fingering and estimated fingering, and subjective evaluation by violin players.

5.1 Experiment 1: Concordance Rate with Textbook

We observed how the concordance rate changes when w_{l} changes.

5.1.1 Conditions

The training data comprised 17 musical pieces (4,852 notes) from four textbooks for intermediate violin students. The test data contained two datasets: The beginner test dataset comprised 14 musical pieces (2,265 notes) from two textbooks for beginners. The intermediate test dataset comprised 14 musical pieces (5,086 notes) from four textbooks for intermediates. The test data did not overlap with the training data.

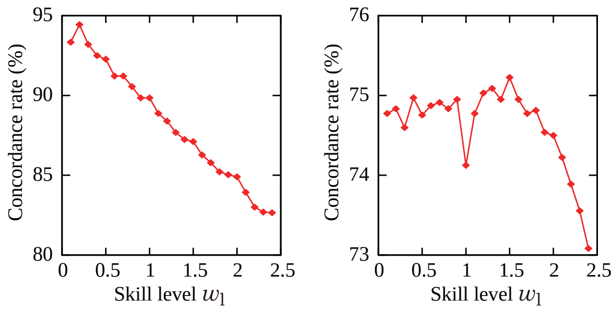
The concordance rate is the number of notes where the estimated fingering patterns match the textbook fingering pattern in both string number and finger number. We set w_{l} as 1.0 when training, and set w_{r} as 4.0. w_{HP} and w_{FI} were decided by using a grid search. The maximum concordance rate combination was searched from 121 combinations where each parameter was set to 0, 0.1, 0.2, ..., 1.0. As a result, $w_{\mathrm{HP}} = 0.6$ and $w_{\mathrm{FI}} = 0.1$.

5.1.2 Results

Figure 4 shows the concordance rate of each test dataset. In the beginner test dataset, the concordance rate is at maximum when $w_{\mathrm{l}} = 0.2$. In the intermediate test dataset, the concordance rate is at maximum when $w_{\mathrm{l}} = 1.5$. Thus, we conclude that w_{l} corresponds to the skill level. In the McNemar test between concordance rates of $w_{\mathrm{l}} = 0.2$ and $w_{\mathrm{l}} = 1.5$, we observed differences at a significance level of 5% in the beginner test dataset, but found no significant differences in the intermediate test dataset.

In the beginner test dataset, good fingering patterns can be estimated because the concordance rate is high, even when the training data used is at the intermediate level. Therefore, fingering can be estimated for a level that is different from the training data level by changing w_{l}.

On the other hand, the concordance rate using the intermediate test dataset was lower than that using the beginner test dataset. Furthermore, no significant differences were found between concordance rates of $w_{\mathrm{l}} = 0.2$ and $w_{\mathrm{l}} = 1.5$. This is because the fingering decisions considered performance expression, which has a high degree of freedom, thereby producing many optimum fingering patterns. Estimated fingering patterns of $w_{\mathrm{l}} = 0.2$ and $w_{\mathrm{l}} = 1.5$ using the intermediate test dataset are different in 720 notes (about 14%), although without significant differences in the concordance rate.



(a) Beginner test dataset (b) Intermediate test dataset

Figure 4. Rate of concordance with textbook fingering patterns

5.2 Experiment 2: Subjective Evaluation

We also verified whether w_{l} reflects the skill levels of the players.

5.2.1 Conditions

We performed a subjective evaluation experiment using the results of Experiment 1. Fingering patterns from $w_{\mathrm{l}} = 0.2$, $w_{\mathrm{l}} = 1.5$, and textbook in the first eight measures of each musical piece were evaluated. The first 10 measures of the musical score were shown to allow the subjects to evaluate the sequence of fingering patterns. The subjects were seven violin players (6–21 years of experience, average 15.0 years). The order of showing the musical piece and three fingering patterns was random, and they were not told which of the three fingering patterns the textbook pattern was. The following questions were asked:

Ease of transition (A) Whether the note on the musical score is played easily using the evaluated fingering pattern without considering performance expression. (difficult 1 – easy 5)

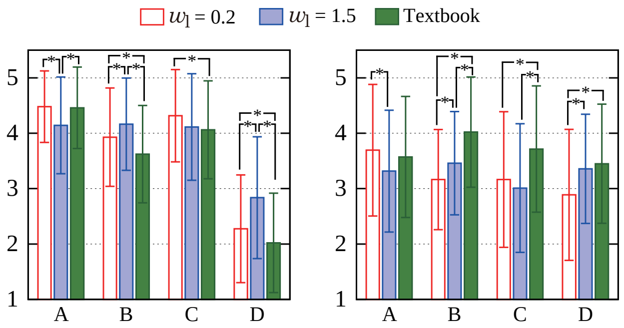
Expression (B) Whether performance expression using the evaluated fingering pattern is suitable. (unsuitable 1 – suitable 5)

Naturalness (C) Whether it is possible for a person to play the violin with the evaluated fingering pattern. (not possible 1 – possible 5)

Skill level (D) The skill level that the evaluated fingering pattern accords with. (low 1 – high 5)

5.2.2 Results

Figure 5 shows the average, standard deviation, and t-test results obtained from the subjective evaluation. When we compared ease of transition and expression, ease of transition was better in the beginner test dataset, whereas expression was greater in the intermediate test dataset. When we compared $w_{\mathrm{l}} = 0.2$ and $w_{\mathrm{l}} = 1.5$, ease of transition was better with $w_{\mathrm{l}} = 0.2$, and expression was greater



(a) Beginner test dataset (b) Intermediate test dataset

Figure 5. Average of the subjectivity evaluation (The error bar represents the standard deviation and the asterisk (*) represents the significance difference ($p < 0.05$) between the two evaluations).

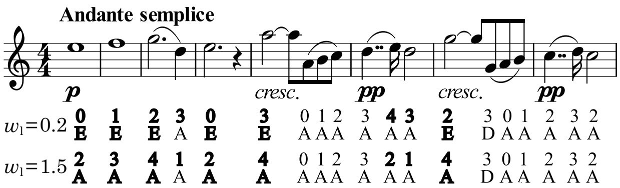


Figure 6. Examples of estimated fingering patterns (finger number and string) for $w_l = 0.2$ and $w_l = 1.5$ in the intermediate test dataset

with $w_l = 1.5$. In the example shown in Figure 6, although the fingering pattern of $w_l = 0.2$ uses open string, the fingering pattern $w_l = 1.5$ avoids using open string when the note length is long; therefore, the fingering pattern of $w_l = 1.5$ accords with higher skill levels than that of $w_l = 0.2$. The answers obtained to the skill level question also show the same relation between w_l and skill level. Thus, we verified that the assumption of relation between skill level and priority of performance expression is correct and that differences in w_l reflect the differences in skill level.

Based on comparisons of our estimation and the textbook patterns, in terms of ease and expression using beginner test dataset, the evaluation of $w_l = 0.2$ was equal to or better than that of the textbook patterns. Therefore, our proposed method can suitably estimate fingering for beginners. On the other hand, in terms of expression and nature, using the intermediate test dataset, the evaluation of our estimation was worse than that of the textbook. The performance expression depends on slur, volume, and other factors; however, we consider only the note length in this paper, thereby reducing the estimation accuracy for expression. The main reason for the low evaluation in the nature question is that the same finger is continuously used, even when the pitches are different. This problem can be solved by introducing the relation between the pitch and the finger in transition probability.

6. CONCLUSION

In this paper, we proposed an HMM-based estimation method of violin fingering according to skill level. By prioritizing performance expression from note length and skill level, fingering patterns based on skill level can be estimated using only one model.

We tested the proposed method in two experiments based on its concordance with textbook fingering and based on subjective evaluation by violin players. However, performance expression, which depends on slur, volume, and other factors, could not be sufficiently tested. Therefore, we will consider these factors in future work.

Acknowledgments

This research was supported in part by JSPS KAKENHI (Grant-in-Aid for Scientific Research) Grant Number 26730182, and the Telecommunications Advancement Foundation (TAF).

7. REFERENCES

[1] A. Radisavljevic and P. Driessen, "Path difference learning for guitar fingering problem," in *Proc. the International Computer Music Conference. International Computer Music Association*, Miami, USA, 2004.

[2] M. Miura, I. Hirota, N. Hama, and M. Yanagida, "Constructing a system for finger-position determination and tablature generation for playing melodies on guitars," *System and Computers in Japan*, vol. 35, no. 6, pp. 10–19, 2004.

[3] D. R. Tuohy and W. D. Potter, "A genetic algorithm for the automatic generation of playable guitar tablature," in *Proc. the International Computer Music Conference*, Barcelona, Spain, 2005, pp. 499–502.

[4] D. Radicioni, C. D. Scienza, and V. Lombardo, "Guitar fingering for music performance," in *Proc. the International Computer Music Conference*, Barcelona, Spain, 2005, pp. 527–530.

[5] H. Gen, K. Hirokazu, and S. Shigeki, "Input-Output HMM Applied to Automatic Arrangement for Guitars," *Journal of information processing*, vol. 21, no. 2, pp. 264–271, 2013.

[6] M. Hart, R. Bosch, and E. Tsai, "Finding optimal piano fingerings," *The UMAP Journal*, vol. 21, no. 2, pp. 167–177, 2000.

[7] Y. Yonebayashi, H. Kameoka, and S. Sagayama, "Automatic decision of piano fingering based on hidden markov models," in *Proc. the 20th International Joint Conference on Artificial Intelligence*, San Francisco, CA, USA, 2007, pp. 2915–2921.

[8] A. A. Kasimi, E. Nichols, and C. Raphael, "A simple algorithm for automatic generation of polyphonic piano fingerings," in *Proc. the 8th International Conference on Music Information Retrieval*, Vienna, Austria, 2007, pp. 355–356.

[9] A. M. Barbancho, A. Klapuri, L. J. Tardon, and I. Barbancho, "Automatic transcription of guitar chords and fingering from audio," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 3, pp. 915–921, 2012.

[10] A. Maezawa, K. Itoyama, K. Komatani, T. Ogata, and H. G. Okuno, "Automated violin fingering transcription through analysis of an audio recording," *Computer Music Journal*, vol. 36, no. 3, pp. 57–72, 2012.

[11] A. M. Burns and M. M. Wanderley, "Visual methods for the retrieval of guitarist fingering," in *Proc. the 2006 Conference on New Interfaces for Musical Expression*, Paris, France, 2006, pp. 196–199.

[12] N. Wakana, S. Shinji, and K. Tadashi, "Violin Fingering Estimation Depending on Skill Level by Consideration of Note Length," *IEEJ Transactions on Electronics, Information and Systems*, vol. 134, no. 4, pp. 581–586, 2014. (In Japanese.)