Adaptive mapping for improved pitch accuracy on touch interfaces

Olivier Perrotin
LIMSI-CNRS
BP 133 - F91403
Orsay, France
olivier.perrotin@limsi.fr

Christophe d'Alessandro
LIMSI-CNRS
BP 133 - F91403
Orsay, France
cda@limsi.fr

ABSTRACT

Touch interfaces such as touchpad or pen tablet are often used for continuous pitch control of pitch in synthesis devices. Usually, pitch is set at the contact point on the interface, thus introducing possible pitch inaccuracies at the note onset. This paper proposes a new algorithm, based on an adaptive attraction mapping, for improving initial pitch accuracy with touche interfaces with continuous control. At each new contact on the interface, the algorithm adjusts the mapping to produce the most likely targeted note of the scale in the vicinity of the contact point. Then, pitch remains continuously adjustable as long as the contact is maintained, allowing for vibrato, portamento and other subtle melodic control. The results of experiments comparing the users' pitch accuracy with and without the help of the algorithm show that such a correction enables to play sharply in tune at the contact with the interface, regardless the musical background of the player. Therefore, the dynamic mapping algorithm allows for a clean and accurate attack when playing touch interfaces for controlling continuous pitch instruments like voice synthesizers.

Keywords

Sound synthesis control, touch interfaces, pen tablet, automatic correction, accuracy, precision

1. INTRODUCTION

Touch interfaces and especially pen tablets offer an accurate and precise control for a variety of synthesis parameters. This makes them attractive devices for controlling synthesizer, and particularly voice instruments [5]. For instance, the Cantor Digitalis, developed at LIMSI [2],[3] makes use of a Wacom tablet[1], to control a voice synthesizer. As in many systems, pitch is controlled by the position of the pen on the tablet (in Cantor Digitalis along the X-axis).

One of the main alternatives for gesture to pitch mapping, is the choice between discrete or continuous steps. In the "discrete steps" solution, pitch can take only discrete values, according to a predefined scale: the pitch value closest to the gesture position is selected, in a manner similar to selection of a note on a keyboard. On the "continuous step" solution, pitch can take any value, according to the gesture position, in a manner similar to a fretless string instrument like the violin, or human voice. On the one hand, the first one prevents the user to play out of tune. On the other hand, the second one offers more effects such as *glissandi*, or *vibrato*. Thus, the latter is often preferred for expressive playing, at the expense of a loss of accuracy. Indeed, playing a note in tune on a continuous board means striking the tablet at a very precise position. In Cantor Digitalis, depending on the size of the tablet, the average is 6 mm per semi-tone. Thus, a small deviation of only a few millimeters can lead to a clearly audible intonation deviation. Although it was shown in [1] that writing-like hand gestures are surprisingly accurate and precise for intonation control tasks, the position of initial contact can lead to inaccuracies in a musical task. To enhance the player's comfort we propose a software method which corrects every possible deviation of the user around the target position.

The dynamic mapping algorithm is presented in the second Section. In Section 3, we present the experiments carried out to assess the effects of this correction method on pitch accuracy and precision. Results are discussed in Section 4. Conclusions are presented in the final section.

2. ADAPTIVE MAPPING ALGORITHM

On a touch interface, pitch is set by the contact of hand-driven mobile object on a fixed device, like the finger on the finger board of a fretless string instrument. A small deviation in position is likely to produce some pitch inaccuracy. In acoustic instruments, this can be compensated with the help of fretting on the finger board (to the expense of subtle and expressive pitch modulations). In an electronic instrument fretting corresponds to rigid region mapping on the pitch axis. We propose a better solution allowing for accurate pitch setting at the contact instant but preserving the subtle pitch modulation capacities of a continuous control device : this adaptive mapping algorithm.

2.1 Principle and algorithm

Let's consider a touch interface like a tablet + stylus device. The position X of the stylus on the tablet along an axis must be mapped on the one dimensional Y pitch variable. Let's assume for the sake of simplicity that X is organized as a straight line. Let us consider relative coordinates, where $(0, 0)$ is the pen position needed to play a note in tune and $(0.5, 0.5)$ is the position to play a semitone (ST) higher. The mapping goes from theX position interval $[-0.5, 0.5]$ around the targeted pen position to the Y pitch interval $[-0.5, 0.5]$ around the targeted pitch. With a linear fixed mapping, a deviation of Δx in position corresponds to a deviation of ΔST. Figure 1 shows a linear mapping with in blue the linear curve, and in red, the related stylus position and pitch played.

In spite of the high precision of hand gestures, it is likely

[1] http://www.wacom.com/products/pen-tablets/intuos

NIME'13, May 27 – 30, 2013, KAIST, Daejeon, Korea.

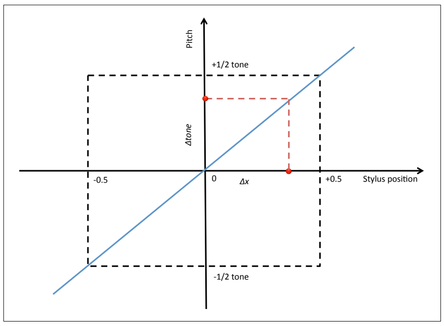

Figure 1: Linear mapping between pen position and pitch with relative coordinates

that the user touches the interface at a position with a small deviation from the target note, which produces a sound out of tune. This algorithm aims at adapting the mapping between pitch and stylus position on the interval of a tone around the target note. To make this algorithm possible, the adapted mapping has to obey the following rules : the new pitch value associated to the pen position at contact is 0 (in tune), and the mapping still goes through the points $(-0.5, -0.5)$ and $(+0.5, +0.5)$ to keep continuity with a fixed linear mapping out of the interval.

An example of adapted mapping is presented in figure 2. The solid blue line is the new mapping, non-linear in the interval [-0.5,0.5] and linear outside. The dot blue line is the previous linear mapping. In red are the related stylus position and pitch played

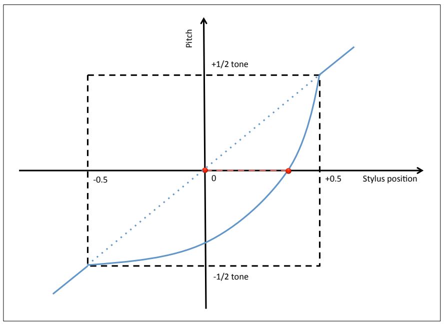

Figure 2: Non-Linear mapping between pen position and pitch with relative coordinates

Then, each time a new contact is made, the mapping is adapted on the interval of a tone to allow the user to play in tune but to allow also any pitch modulation. Contrary to the "keyboard" or "fretting" situation, pitch is not fixed. But similarly to the "keyboard" or "fretting" situation it follows a fixed scale at the instant of contact. While the user stays in the interval, the non-linear mapping is kept. If the user leaves this interval, a linear mapping is applied all along the tablet. The continuity at the points $(-0.5, 0.5)$ and $(+0.5, +0.5)$ enables to make a transition between non-linear mapping and linear mapping when the user slides the pen out of the interval around the targeted note. Although the curve is not smooth at these points, the transition is not audible at a scale of few millimeters. Likewise, with this scale, the non linear mapping is barely perceptible, unless the stylus is moved at a fast rate, more than one millimeter/second, which does occur in practice.

2.2 Analytic expression

An expression defined from $[-0.5, 0.5]$ to $[-0.5, 0.5]$ which verifies the rules given in section 2.1 can be given by :

$$g(y) = \frac{e^{(y+0.5)\gamma} - 1}{e^{\gamma} - 1} - 0.5 \tag{1}$$

Where $\gamma \neq 0$ is the curvature. It can be shown easily that $g(-0.5) = -0.5$ and $g(0.5) = 0.5$. This expression is bijective on $[-0.5, 0.5]$ and to simplify calculation, we use the inverse function $f = g^{-1}$ such as :

$$y = f(x) = \frac{1}{\gamma} \left[\log \left[(e^{\gamma} - 1)(x + 0.5) + 1 \right] \right] - 0.5 \tag{2}$$

With x the pen position, and y the pitch, we then have to choose the curvature γ_0 such as $f(x_0, \gamma_0) = 0$, where x_0 is the initial pen position. It leads to the expression :

$$\gamma_0 = 2 \log \left(\frac{1 - 2x_0}{1 + 2x_0} \right) \tag{3}$$

Therefore the final algorithm becomes :

1. When contact is done at the abscissa x_0, compute the curvature γ_0 with equation 3 needed to make the user play in tune
2. While the user stays in the interval $[-0.5, 0.5]$, compute the pitch y with equation 2, given $\gamma = \gamma_0$
3. If the user leaves this interval, apply a linear mapping all along the tablet, including the initial interval.
4. If a new contact is made, go back to 1.

All implementations have been done on Max/MSP. We added our algorithm to the Cantor Digitalis patch [3].

3. EXPERIMENT

The aim of the experiment is to measure the improvement brought by the accuracy correction. Subjects are asked to reproduce simple patterns in different conditions : with or without audio feedback and with or without correction of accuracy, which give four combinations for each pattern. For each, a score with the name of notes is given to the subject. The latter has also the possibility to listen the pattern through a MIDI synthesizer as much as wished. It provides a piano sound with equal temperament and a4=440Hz. The patterns are given in random order.

Then the subject has 5 trials to reproduce the pattern at a given tempo of 120 b.p.m. provided by a metronome (sound and visual). To be able to use the correction algorithm, it is asked to the subject to raise the pen and make a new contact for each note played. However, the subject is not told wether the correction is activated or not. The tablet used for the experiment is a Wacom Intuos 4M. The linear mapping is 1.25cm per semitone on X-axis. X-coordinates and pen pressure are recorded for each subject. A training session presenting different patterns from the experiment but the same protocol was proposed first to the subjects to adapt with the tablet and the interface.

3.1 Subjects

The experiment were completed by 7 subjects. The average age is 28.4 years and the average musical training is 10.6 years. However, it is important to note half have a strong background whereas the others have few experience in music. None of the subjects presents known auditory impairment and all are right-handed. Details for each subjects are given in table 1.

Table 1: Subjects characteristics

Subject	Gender	Age (yr)	Music training (yr)
S1	M	22	17
S2	M	27	20
S3	M	27	10
S4	M	30	4
S5	F	30	4
S6	M	40	3
S7	F	23	16

3.2 Experiment

The set of melodies used in this experiment are 5-note patterns, inspired from Alberti basses. Usually found as piano accompaniment in the Classical era, this pattern plays the notes of a given chord in the following order : lowest, highest, middle, highest and is repeated.

Based on pattern in C major (c3, g3, e3, g3, c3) we choose to transpose the middle and highest notes. This way we have a changing interval between the c3 and highest note, and a fixed third between middle and highest notes. Then all these patterns are transposed a fifth above to give two times four sequences we use shown in figure 3. Such sequences offer the following advantages: Increasing intervals between first and second note, as all the patterns are played with the same tempo, the subjects have to adapt their speed, and may lose in accuracy for large intervals; alternate forward and backward hand movements; they give twice the same exercise (in C and in G) to double the number of experiments without giving the same sequence to play twice in the same conditions.

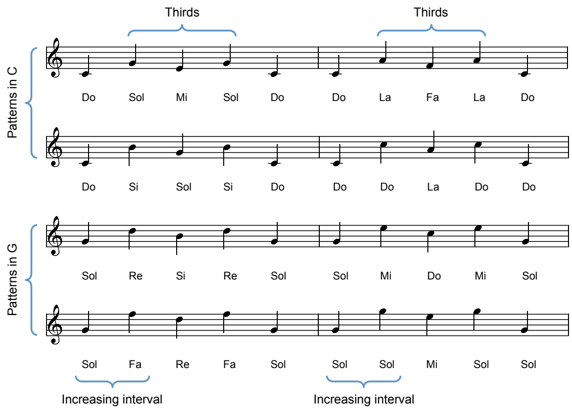

Figure 3: Patterns used for the experiment

3.3 Extraction of features

For each trial, X-coordinates and stylus pressure were recorded. A non-zero pressure indicates user is playing, and we extracted each notes from the whole recording this way using Matlab. For each note, two values are relevant :

- *The value at contact.* It is where the correction algorithm works.
- *The sustained value.* The user can slightly move the stylus between the contact and the final note, and this is the one we hear. After quantization of the pitch with a step of 0.5% of a tone, the final note is extracted taking the larger step.

Our algorithm corrects pitch in an interval of $[-0.5, 0.5]$ tone around the target notes. Then, if the error is higher or lower than a semitone, the correction is useless, and such trials are removed from the results to focus on the effect of the algorithm. To compare the influence of the algorithm, we processed each of the four conditions of the experiment separately. The number of trials for each conditions are between 232 and 244 : T (tablet alone, without adaptive mapping): 244; T+A (tablet and audio, without adaptive mapping): 243; C+T (tablet alone, with adaptive mapping): 224; C+T+A (tablet and audio, with adaptive mapping): 232.

4. DISCUSSION

4.1 Definitions

Our performance analysis is based on "accuracy" and "precision", following the work on singing presented in [4] :

Accuracy is the average difference between the pitch played by a musician and the target pitch. With N the number of notes indexed by i in the melody, S_i a note played and T_i a target note, it can be computed with the following equation :

$$A = \frac{\sum_i^N (S_i - T_i)}{N} \tag{4}$$

Precision is the consistency of the pitch played in repeated occasions, which in statistics is the standard deviation of played notes relatively to the target notes. It is computed among all the trials independently on each note (pitch class PC) with the equation :

$$P_{PC} = \sqrt{\frac{\sum_i^{N_{PC}} (S_i - M_{PC})^2}{N_{PC}}} \tag{5}$$

Where N_{PC} is the number of notes with the same target within all trials. M_{PC} is the average of the N_{PC} pitches, and S_i is a note played. Then the global precision for a subject is computed averaging the precision on all pitch classes.

4.2 Analysis

We compute the accuracy and precision of each subject according to equations 4 and 5, for each experimental condition. These values are plotted in figure 4. Each box contains the accuracy or precision of the 7 subjects regarding the condition indicated. The red lines are the medians of all the means (or standard deviations) regarding the conditions. The blue boxes contain the 2nd and 3rd quartiles of the values. Values which are within 1.5 times the range of the cumulated 2nd and 3rd quartiles are between the black whiskers. The remaining outliers are the red crosses.

We first note relatively low dispersion. This means that the subjects show the same general ability for playing the tablet, regardless their musical background. Indeed the latter helps in music understanding, which was not required here. In this experiment, handling the tablet reflects actually the ability to use a pen, which is almost similar within the subjects.

The plots in the top show the accuracy and precision of subjects at contact. We showed in part 2.2 that the correction is supposed to provide an error equal to zero. However, due to real-time implementation issues, the algorithm computes the new pitch 10 ms after contact, whereas the pen could have moved slightly from the initial position. In the top left, the median mean error goes from -13.5 cents of semitones without correction to -0.61 cents of semitones with correction, regardless the audio. Moreover, all the values with correction are in a range smaller than 4 cents of semitone, which is the limit of perception of the ear to differentiate two pitches. Then it shows at contact, the algorithm corrects well the stylus position to provide a pitch

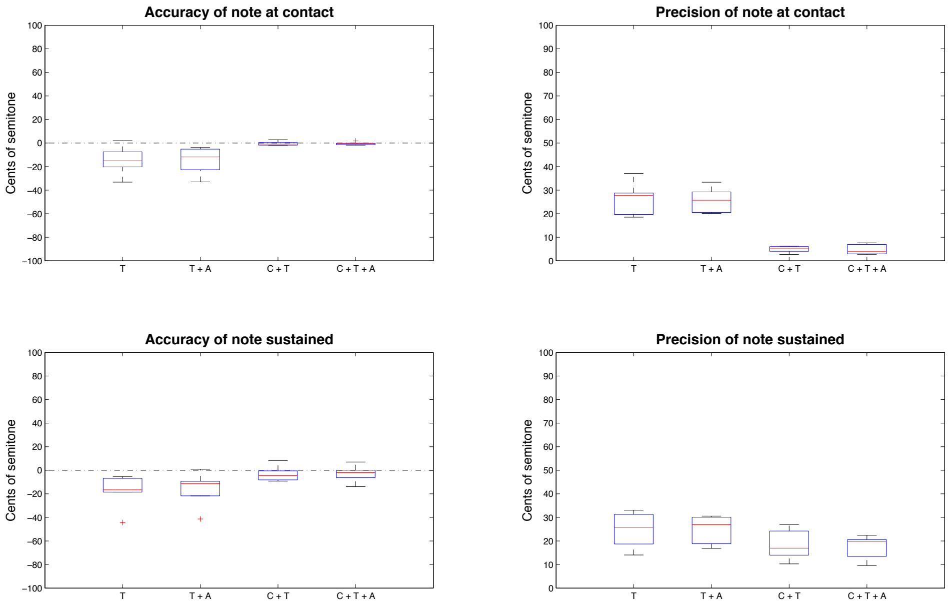

Figure 4: Accuracy (left) and precision (right) expressed in Cents for different conditions (from left to right in each pannel): T (tablet alone, without adaptive mapping), T+A (tablet and audio, without adaptive mapping), C+T (tablet alone, with adaptive mapping), C+T+A (tablet and audio, with adaptive mapping) . Top: at the contact instant. Bottom: when sustaining the tone..

which sounds in tune. This is confirmed with the standard deviations of the errors plotted in the top left corner. The median of standard deviation goes from 26.7 cents of semitones without correction to 4.63 cents of semitones with correction, regardless the audio. Then it is divided by about 6 when adding the correction.

The plots in the bottom show accuracy and precision of subjects during sustained notes. The median mean error goes from -14.1 cents of semitones without correction to -3.24 with correction. Although the medians are under the threshold of perception, the range of values is higher than at contact. It shows users move the stylus after contact to finally play slightly out of tune. This trend occurs whatever the condition. It is therefore a natural consequence of the user movement. However, it appears that this deviation is smaller when audio is activated. Thus an audio feedback enables the players to control better their gesture. It is also remarkable even the subject who plays really low without correction (the red crosses) is well corrected with the algorithm. Looking at the standard deviation, on the bottom right plot, the improvement seems less significant. The decrease of the median is smaller than the range of values. Then this plot shows that globally, in spite of the deviation of the stylus, the correction improves the mean of errors of sustained notes even if the standard deviation is still high.

Finally, these experiments show the effectiveness of the algorithm on note at contact, meaning the attack of the sound, regardless of the music player. Then, by a lack of control of the stylus, the latter deviates from the initial position to produce a sound slightly out of tune. However, it is important to note that even if the algorithm can't prevent this deviation, it compensates it slightly, letting the remaining error to the expertise of the player.

5. CONCLUSIONS

Based on simple mapping transformations, the adaptive algorithm presented in this paper is able to correct in real time the pitch played on a continuous interface such as a pen tablet. In the case a user touches the tablet in an interval of a tone around the target position, experiments showed the pitch is automatically adjusted to the target pitch, to provide a note perfectly in tune. Moreover, the corrected mapping is only applied on the interval around the target note, and the continuity with a linear mapping makes it imperceptible to listening. The main limit of this work, is the deviation of stylus after contact, which cannot be prevented by this algorithm. The experiments were made using a pen tablet, but the algorithm also works with touch interfaces such as the Magic Trackpad. It is all the more interesting that contact on touch interfaces with fingers are less precise than contacts with a pen. We can expect an even better result, but experiments have to be conducted to quantify the improvement of the algorithm on touch interfaces.

6. REFERENCES

[1] C. D'Alessandro, A. Rilliard, and S. Le Beux. Chironomic stylization of intonation. *Acoustical Society of America*, 129:1594–1604, 2011.

[2] L. Feugère, S. Le Beux, and C. D'Alessandro. Chorus digitalis: polyphonic gestural singing. *International Workshop on Performative Speech and Singing Synthesis*, March 14-15 2011.

[3] S. Le Beux, L. Feugère, and C. D'Alessandro. Chorus digitalis: Experiments in chironomic choir singing. *Proceedings of Interspeech*, August 28-31 2011.

[4] P. Q. Pfordresher, S. Brown, K. M. Meier, M. Belyk, and M. Liotti. Imprecise singing is widespread. *Acoustical Society of America*, pages 2182–2190, July 18 2010.

[5] M. Zbyszynski, M. Wright, A. Momeni, and D. Cullen. Ten years of tablet musical interfaces at cnmat. *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 100–105, June 2007.