PARAMETRIC SYNTHESIS OF GLISSANDO NOTE TRANSITIONS – A USER STUDY IN A REAL-TIME APPLICATION

Henrik von Coler
Audio Communication Group
TU Berlin
Germany
voncoler@tu-berlin.de

Moritz Götz
Audio Communication Group
TU Berlin
Germany

Steffen Lepa
Audio Communication Group
TU Berlin
Germany
steffen.lepa@tu-berlin.de

ABSTRACT

This paper investigates the applicability of different mathematical models for the parametric synthesis of fundamental frequency trajectories in glissando note transitions. Hyperbolic tangent, cubic splines and Bézier curves were implemented in a real-time synthesis system. Within a user study, test subjects were presented two-note sequences with glissando transitions, which had to be re-synthesized using the three different trajectory models, employing a pure sine wave synthesizer. Resulting modeling errors and user feedback on the models were evaluated, indicating a significant disadvantage of the hyperbolic tangent in the modeling accuracy. Its reduced complexity and low number of parameters were however not rated to increase the usability.

1. INTRODUCTION

Note transitions are an essential part of articulation and thus of expressive musical performances. On instruments with continuous excitation and a continuous frequency scale, such as the violin or the singing voice, glissando note transitions are thus of particular interest. A so called *Glissando* or *Portamento* mode has thus been implemented in many analog and digital synthesizers since their early days. Most devices allow the tuning of the transition time, some offer the selection of different trajectory functions. The comparison of different parametric models presented in this work is considered a step towards an extension of this established concept.

The topic of modeling fundamental frequency trajectories has been addressed in the disciplines of speech and music analysis / synthesis in the past. The main features of these glissando transitions can be expressed in terms of the fundamental frequency f_0 and short-term energy trajectories (RMS). 't Hart [1] compared straight lines and parabolas for modeling the fundamental frequency of speech syllables using modulated pulse trains. Simple straight-line segments were indistinguishable from parabolic ones in a listening test. For modeling the prosody of speech utterances, Hirst et al. [2] applied quadratic spline functions.

Battey [3] used third order Bézier splines to model trajectories of f_0, amplitude and spectral centroid for musicological analysis but also referred to the application in *expressive computer rendering*. Barbot et al. [4] compared the modeling accuracy of cubic B-splines and natural cubic splines for f_0 trajectories of speech syllables. Using 4 support points each, the B-splines achieved a lower RMS error. B-Spline and spline models were compared by Lolive et al. [5] for the use of modeling fundamental frequency in speech synthesis systems. Within a sinusoidal modeling approach, Hahn et al. [6] used B-splines to model the temporal trajectories of partial parameters. Ardaillon et al. [7] evaluated a parametric f_0 model based on B-splines within a concatenative singing voice synthesis system through listening tests.

Although the qualities of different trajectory models in terms of modeling error and perception have been investigated thoroughly in the past, little is known about the usability of these models in real-time applications. The nature of parameters is individual for each model and an increasing number of parameters might decrease the intuitiveness. Modeling precision and usability are hypothesized to be opposed. The simpler the model, the larger the modeling error but the easier the control. This work thus focuses on the usability of trajectory models with parametric control in a real-time application. Hyperbolic tangent, cubic splines and Bézier curves will be compared in a user experiment. The hyperbolic tangent offers just one parameter, cubic splines have been implemented with two and Bézier curves with three control parameters [8].

The remainder of this paper is organized as follows: In Section 2 the implemented models will be introduced. Section 3 presents the user study, followed by the results in Section 4 and their discussion in Section 5. A conclusion is presented in Section 6.

2. GLISSANDO MODELING

Glissando note transitions are the segment between two adjacent notes of different pitch, in which the fundamental frequency trajectory and the energy trajectory are continuous. The glissando segment is defined as the region between the stationary segments of the pitches f_1 and f_2, as shown in the idealized model in Fig. 1. The idealized fundamental frequency trajectory (b) of these regions is closely related to sigmoid curves whereas the idealized energy trajectory (a) remains constant.

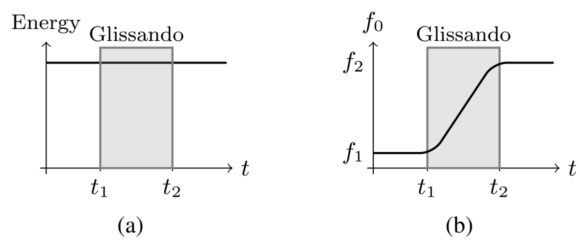

Figure 1: Idealized transition model for glissando articulation

For the calculation of the actual trajectories used in the experiment, an analysis of the fundamental frequency trajectory was performed with a hopsize of $L_{hop} = 256$ samples, respectively 2.7 ms, using the YIN algorithm [9].

The resulting trajectories were subsequently modeled using hyperbolic tangent, cubic splines and Bézier curves. The fundamentals of these models and the resulting parameters will be outlined in the remainder of this section.

2.1. Hyperbolic Tangent

The hyperbolic tangent is defined as:

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)} \tag{1}$$

In order to make this basic function applicable for different intervals $\Delta f = f_2 - f_1$ and durations $\Delta t = t_2 - t_1$, the following parameters are added:

$$T(t) = c + d \tanh\left(\frac{t-a}{b}\right), \quad t, a, b, c, d \in \mathbb{R} \tag{2}$$

For a transition between two values f_1 and f_2 the parameters c and d must be:

$$\begin{aligned} d &= \frac{|f_1 - f_2|}{2}, \\ c &= \min(f_1, f_2) + d. \end{aligned} \tag{3}$$

Parameter a is depending on the time values. For a transition between the first value t_0 and the last value t_1 parameter a must be:

$$a = \frac{|t_1 - t_2|}{2}. \tag{4}$$

The resulting single parameter b, presented to the user in the study, controls the slope of the function by time-scaling. In Figure 2, hyperbolic tangent curves are plotted with different values for b.

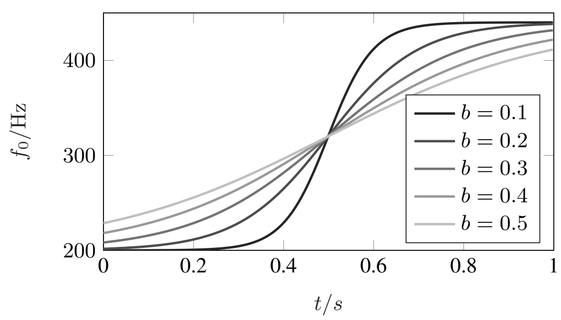

Figure 2: Hyperbolic tangent with different values for b

2.2. Cubic Splines

Splines are special functions for the piece wise interpolation by polynomials. A cubic spline S with n points $P_i = (x_i, y_i)$ is defined as:

$$\begin{aligned} S(x) &:= a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3, \\ & x \in [x_i, x_{i+1}], a_i, b_i, c_i.d_i \in \mathbb{R}, i = 0, 1, ..., n-1. \end{aligned} \tag{5}$$

Arbitrary points from the extracted f_0-trajectories can be used to get a polynomial representation of the curve. An equidistant 4-point model is used in the experiment. The x-values of the control points are thus fixed. In Figure 3, a natural cubic spline curve is plotted with four points. The outer points P_1 and P_2 are fully determined by the boundary conditions, so are the x-values for P_3 and P_4. Two remaining parameters – the y-values of P_3 and P_4 – are presented to the user in the experiment for controlling the trajectory.

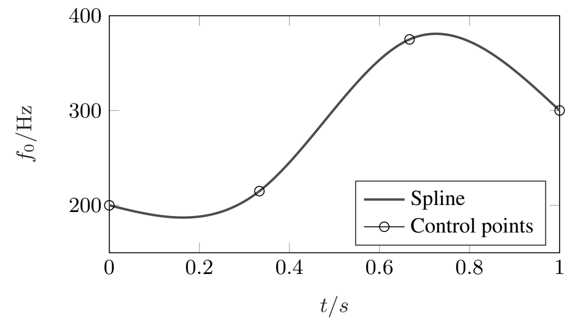

Figure 3: Example of a natural cubic spline with four control points

2.3. Bézier Curves

Bézier curves are controlled by a number of control points, of which only the start and end point lie on the curve itself. A Bézier curve $K(x)$ is defined by sum of Bernstein polynomials $B_i^n(x)$ and the control points P_i:

$$K(x) = \sum_{i=0}^{n} P_i B_i^n(x), n \in \mathbb{N} \tag{6}$$

For the application in the experiment the x-values of P_i have been set to be equally spaced:

$$P_{i,x} = \frac{i}{n}, \quad i = 0, 1, ..., n \tag{7}$$

Figure 4 shows an example of a Bézier curve with 5 control points. Since the outer points are fully determined by the boundary conditions (f_1 and f_2) and the x-values are predefined, the user is presented the three y-values of the inner control points as parameters in the experiment.

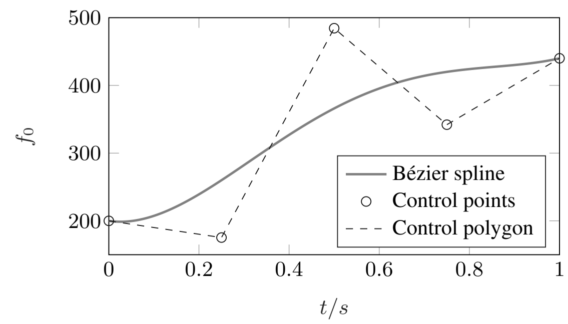

Figure 4: Bézier curve with five control points and control polygon

2.4. Modeling Accuracy

For evaluating the numerical modeling qualities, all 96 two-note sequences from the *TU-Note Violin Sample Library* [10, 11] with glissando transitions were used. This selection contains upward and downward glissandi at different positions and dynamics. The mean absolute error was applied to evaluate the deviation between original trajectories x_i and model estimates $\tilde{x}_i$ of length $N, \tilde{x}_i \in \mathbb{R}, n \in \mathbb{N}$:

$$\bar{\delta}_x := \frac{1}{N} \sum_{i=0}^{N-1} |x_i - \tilde{x}_i|. \tag{8}$$

Using $\bar{\delta}_x$, the best possible fit was calculated for all models, also evaluating different orders for splines and Bézier curves. The best parameter settings were found by calculating the model parameters related to a curve intersecting the original trajectory at the x-values of the control points. For the hyperbolic tangent this resulted in a modeling error of $\bar{\delta}_x = 0.083$. For the splines, an increase of the number of interpolation points lead to a monotonic decrease in modeling error (Table 1). For Bézier curves, a minimum error $\bar{\delta}_x = 0.0311$ was reached with 7 interpolation points (Table 2). A further increase lead to an increase of the error, which is likely to be caused by the fixed x-values of the control points and the method for finding the best parameter set. The numbers of control points chosen for the experiment are marked bold in Table 1 and 2.

Table 1: Minimum of mean absolute error for different spline orders

control points	mean of normalized $\bar{\delta}_x$	mean of $\bar{\delta}_x$ [Hz]
4	**0.0387**	**7.24**
5	0.0272	5.04
6	0.0205	3.66
7	0.0163	2.93
8	0.0145	2.59
9	0.0119	2.16
10	0.0109	1.98
11	0.0096	1.72
12	0.0094	1.67

Table 2: Minimum of mean absolute error for different Bézier orders

control points	mean of normalized $\bar{\delta}_x$	mean of $\bar{\delta}_x$ [Hz]
4	0.0539	10.21
5	**0.0394**	**7.26**
6	0.0358	6.61
7	0.0311	5.39
8	0.0325	5.74
9	0.0377	6.63
10	0.0379	6.67
11	0.0594	10.22
12	0.0805	13.73

3. USER STUDY

A user study was conducted to compare the usability of the three proposed trajectory models. Using a within-subject design, participants had to apply the three different models to reproduce seven sequences of two notes which are connected with a glissando. Errors between original and reproduction were evaluated alongside additional user feedback to obtain information on the real-time usability of the three models. The Bézier model was presented to the user with one tuning parameter, splines were used with two and Bézier curves with three parameters, respectively four and five support points.

3.1. Test System

The synthesis engine with the real-time trajectory modeling was programmed in C++, using the JACK API [12]. The runtime system was a Raspberry Pi 3 Model B Rev 1.2, running Raspbian GNU/Linux 9.1. A *Behringer U-Control UCA222* audio interface was used with a processing block size of 128 samples at a sampling rate of 48 kHz. A *Logilink* USB to MIDI Adapter was used for the MIDI input with a *Swissonic ControlKey 49*. Faders were routed to the parameters of the trajectory models to allow control by the participants. A pure sinusoidal synthesizer with fixed amplitude was implemented within the test system. A control surface for the user study which managed the handling of the trials, the input of the user data and configured the synthesis engine via MIDI was programmed in Pure Data [13].

Table 3: Stimuli employed in the seven tasks of the user study, stemming from the *TU-Note Violin Library* [10]

Item	note 1	note 2	length	direction
`TwoNote_DPA_18`	A3	D4	380 ms	up
`TwoNote_DPA_19`	D4	A3	320 ms	down
`TwoNote_DPA_65`	E5	B4	400 ms	down
`TwoNote_DPA_66`	B4	E5	485 ms	up
`TwoNote_DPA_113`	D4	G4	300 ms	up
`TwoNote_DPA_137`	A4	D5	700 ms	up
`TwoNote_DPA_186`	E6	B5	550 ms	down

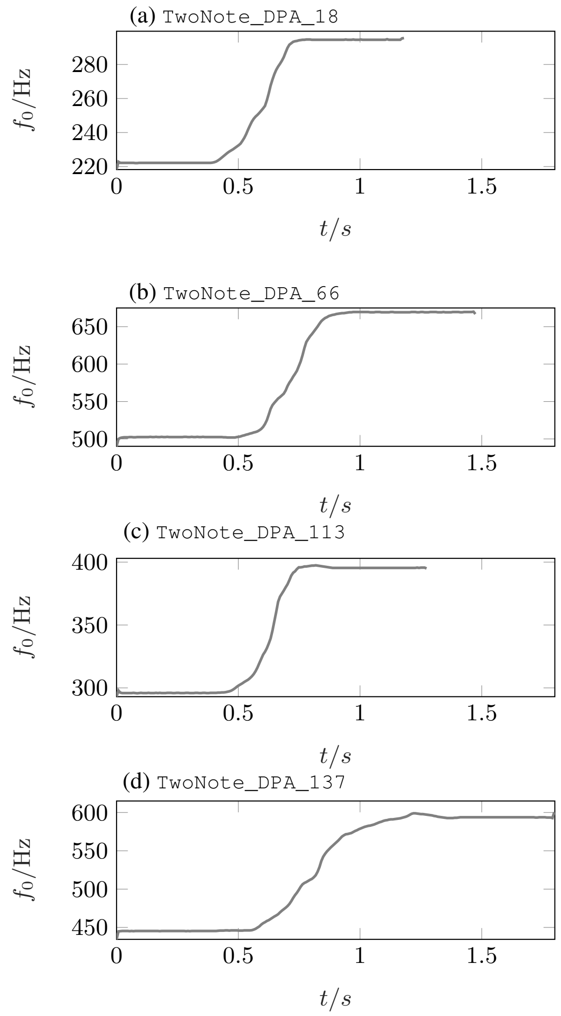

Figure 5: Fundamental frequency trajectories for upward glissando stimuli

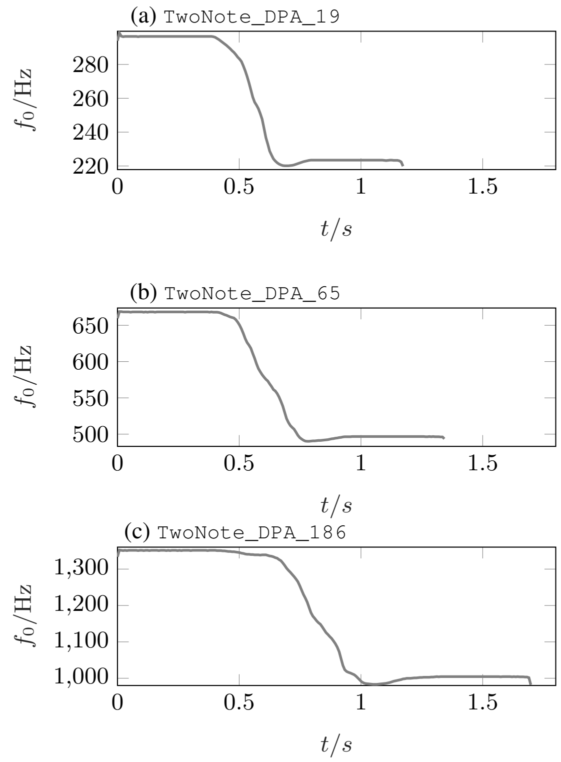

Figure 6: Fundamental frequency trajectories for downward glissando stimuli

3.2. Stimuli for Reproduction Tasks

Stimuli for the reproduction tasks were generated using the *TU-Note Violin Sample Library* [11, 10], which features two-note sequences with annotated glissando transitions. Four upward and three downward two-note sequences, listed in Table 3, were selected with different note frequencies, in order to cover the range of the instrument. The fundamental frequency trajectories of these seven sequences were extracted and are visualized in Figure 5 and Figure 6. These trajectories were then used to drive a simple sinusoidal synthesizer with a fixed amplitude, in order to exclude influences from features other than the fundamental frequency.

3.3. Participants

15 participants were recruited through the mailing list for students of the audio communication group at TU Berlin. 14 of them were male and one was female. Participants' mean age was 27.4 years with a standard deviation of 5.6 years. The majority of the participants were musically skilled: 60 % played an instrument on a regular basis for more than 6 years and also 66.67 % had ear training for more than one year.

3.4. Procedure

After an introduction to the test system and a free play period with all three trajectory models, each participant went through 21 experimental trials: Each of the 7 task stimuli had to be re-synthesized by the users using each of the three models in a fully randomized order. In every trial, the task stimulus could be played back as often as desired. Additional information on the current trial was shown on the graphical user interface, which included a number referring to the currently active trajectory model (1,2,3) and the starting and the ending note of the sequence.

Participants were then instructed to reproduce the sequence using the MIDI keyboard and the real-time synthesis engine. The length of the glissando was fixed for each stimulus, but the parameters of each model could be adjusted. Once the participants were satisfied with their settings, three questions about the just employed model and its parameters had to be answered using vertical continuous sliders (ranging from 0-100) on the graphical user interface. In the study the questions were in German, hence a translated version is shown in Table 4.

4. RESULTS

Since the resulting data is not normal distributed and the amount of 15 participants may be considered small, the non-parametric

Table 4: Rating scales asked after each single task

Question	0	100
Parameter changes were ...	not audible	clearly audible
The model allowed an easy adjustment of the stimulus	completely disagree	completely agree
The number of parameters in this model is ...	too low	too high

Friedman test has been chosen to evaluate each dependent variable, separately. The independent factor is the trajectory model with three levels. The dependent variables are the mean absolute modeling error in the reproduction of the task stimulus $\bar{\delta}_x$ as well as the scores from the three rating scales. All dependent variables have been averaged across the seven presented tasks.

4.1. Modeling Error

Box plots in Figure 7 show a higher modeling error for the hyperbolic tangent than for splines and Bézier curves. The results show a statistically significant difference in modeling error depending on the trajectory model, $\chi^2 = 7.600, p = 0.000$. A post hoc analysis was conducted using Wilcoxon signed-rank tests. Bonferroni correction resulted in a significance level of p < 0.017. Median (IQR) modeling errors for the hyperbolic tangent, Spline and Bézier model were .3377 (.3270 to .3621), .1230 (.0957 to .2042) and .1266 (.0782 to .1722), respectively. There was no significant difference between the Bézier and the Spline model (Z = -.795 , p = .427). The Hyperbolic tangent model, however, showed a significantly higher modeling error than the the Spline (Z = -3.408, p = .001) and the Bézier model (Z = -3.408, p = 0.001).

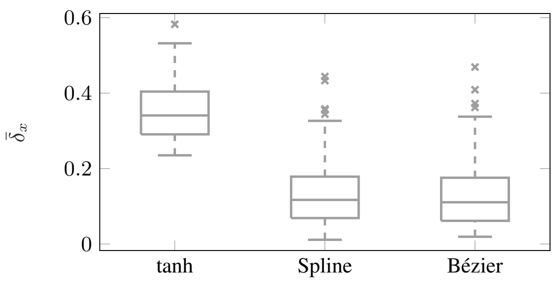

Figure 7: Boxplots for modeling error $\bar{\delta}_x$, averaged across tasks

4.2. Audibility of Parameter Changes

Results of the question whether parameter changes are audible are shown in Figure 8 as box plots, indicating a slightly better audibility for the hyperbolic tangent. Results of the Friedman test show a statistically significant difference in the audibility of parameter changes depending on the trajectory model, $\chi^2 = 6.218, p = .045$. Again, Wilcoxon signed-rank tests were used for a post hoc analysis with a Bonferroni correction, resulting in a significance level of p < 0.017. Median (IQR) of the rated audibility for the hyperbolic tangent, Spline and Bézier model were 86.1446 (77.9786 to 98.1354), 66.5328 (56.5978 to 87.6363) and 72.8342(56.3110 to 85.2744), respectively. The post hoc analysis, however, showed no significant difference between any of the models, neither between Bézier and the Spline model (Z = -.031 , p = .975) nor between Spline and hyperbolic tangent (Z = -2.166, p = .030) or Bézier and hyperbolic tangent (Z = -2.271, p = .023).

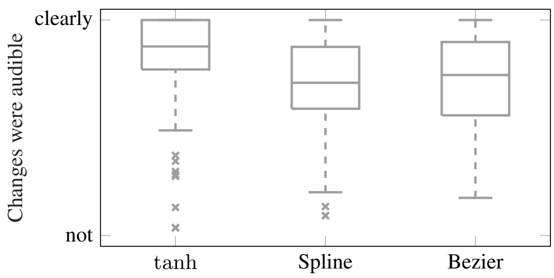

Figure 8: Box plots for question *audibility of parameter changes*

4.3. Ease of Adjustment

Figure 9 shows box plots for the responses to the question referring to the *ease of adjustment*. The Friedman Test showed no significant influence of the trajectory model on the perceived ease of adjustment, $\chi^2 = 2.533, p = .282$. Median (IQR) of the ease of adjustment for hyperbolic tangent, spline and Bézier model were 72.2318 (48.9673 to 93.2301), 77.3379 (57.8026 to 88.1813) and 71.7154 (45.1807 to 79.8336). No significant difference between any of the models, neither between Bézier and the Spline model (Z = -.909 , p = .363) nor between Spline and hyperbolic tangent (Z = -.057, p = .955) nor between Bézier and hyperbolic tangent (Z = -.795, p = .427).

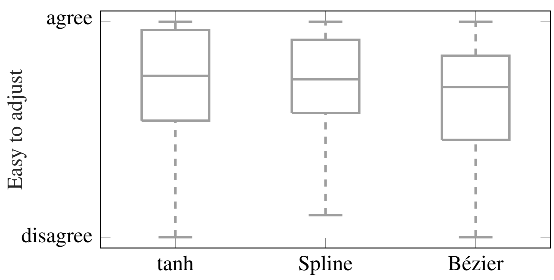

Figure 9: Box plots for the question *ease of adjustment*

4.4. Number of Parameters

Box plots for the question regarding the *number of parameters* are shown in Figure 10. There was a statistically significant difference in the rating whether the number of parameters was too low (0) or too high (100), depending on the number of provided parameters, $\chi^2 = 26.271, p = .000$. Median (IQR) of the response to the question for for hyperbolic tangent, spline and Bézier model were 29.1165 (19.8795 to 46.4429), 49.1968 (47.9920 to 50.4016) and 65.3758 (54.9340 to 67.6133). Results show a significant difference between one and two parameters (Z = -3.045 , p = .002) one and three parameters (Z = -3.408, p = .001) as well as two and three parameters (Z = -3.408, p = .001).

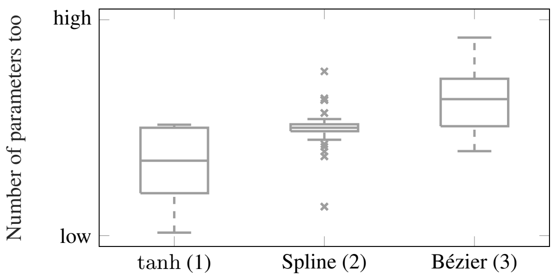

Figure 10: Box plots for question *Number of parameters*

5. DISCUSSION

The results show that the hyperbolic tangent leads to a larger modeling error than cubic splines and Bézier curves in the user experiment. Hence, the hyperbolic tangent is less suitable for synthesizing the glissando transitions presented in the sequences, regarding the mean absolute error. This relation could also be observed for the best model fits in the automated evaluation in Section 2.4, although the user experiment resulted in higher error rates.

Further, the results show a significant preference of two parameters, since this number is rated as neither too high, nor too low. This relation is presumably independent of the trajectory models and probably of a basic psychological nature, since two was the mean number of parameters presented to the users. Since the hyperbolic tangent was used with one, splines with two and Bézier with three parameters, these findings can not be interpreted, independently.

It would be conceivable that the hyperbolic tangent was easier to adjust by the participants. The *ease of adjustment*, however, was not influenced by the model or by the number of parameters. This justifies the use of more complex models and rejects the initial hypothesis that they could be more difficult to use.

6. CONCLUSION

The presented study could deliver first insights on the usability of hyperbolic tangent, cubic splines and Bézier curves for glissando modeling in a real-time scenario. Using the hyperbolic tangent resulted in the largest modeling errors, whereas an increased number of parameters for the other models did not reduce the usability. Thus, the use of such models can be considered justified.

Several aspects of this study could be subject to further, more detailed experiments. It would be of interest to investigate the factor *number of parameters* independently of the trajectory model. For reasons of feasibility, these aspects have been mixed in this study.

Since the errors for the seven trajectory types in the tasks have been averaged, the individual features of the glissandi were not evaluated. Studies using the glissando type (up, down) as independent variable might reveal more differences between the trajectory models.

Future research should incorporate other instruments, additional musicians and different musical content. The glissando transitions of the violin in this user study were of rather smooth nature. They contained no overshoots, unlike for example the singing voice, which might be easier to synthesize with Bézier curves. Different instruments may require other models.

Finally, the mean absolute error may not the ideal measure to evaluate the performance. It was nevertheless chosen as a first step towards a procedure. In fact, the perceived modeling accuracy is a more important factor in musical re-synthesis tasks. Thus, a combination of the presented study with a listening test can deliver further results.

7. REFERENCES

[1] Johan 't Hart. "F0 stylization in speech: Straight lines versus parabolas". In: *The Journal of the Acoustical Society of America* 90.6 (1991), pp. 3368–3370.

[2] Daniel Hirst and Robert Espesser. "Automatic Modelling Of Fundamental Frequency Using A Quadratic Spline Function". In: Travaux de l'Institut de Phonétique d'Aix (1993), pp. 75–85.

[3] Bret Battey. "Bézier spline modeling of pitch-continuous melodic expression and ornamentation". In: *Computer Music Journal* 28.4 (2004), pp. 25–39.

[4] Nelly Barbot, Olivier Boëffard, and Damien Lolive. "F0 stylisation with a free-knot b-spline model and simulated-annealing optimization". In: *Proceedings of the 9th European Conference on Speech Communication and Technology (Eurospeech)*. Lisbon, Portugal, 2005.

[5] Damien Lolive, Nelly Barbot, and Olivier Boëffard. "Comparing B-Spline and Spline Models for F0 Modelling". In: *International Conference on Text, Speech and Dialogue*. Springer. Brno, Czech Republic, 2006, pp. 423–430.

[6] Henrik Hahn and Axel Röbel. "Extended Source-Filter Model for Harmonic Instruments for Expressive Control of Sound Synthesis and Transformation". In: *Proceedings of the 16th International Conference on Digital Audio Effects (DAFx-13)*. Maynooth, Ireland, 2013.

[7] Luc Ardaillon, Gilles Degottex, and Axel Roebel. "A multi-layer F0 model for singing voice synthesis using a B-spline representation with intuitive controls". In: *Interspeech 2015*. Dresden, Germany, 2015.

[8] Moritz Götz. "Analysis and Synthesis of Control Parameters in Note Transitions". MA thesis. TU-Berlin, 2018.

[9] Alain de Cheveigné and Hideki Kawahara. "YIN, a Fundamental Frequency Estimator for Speech and Music". In: *The Journal of the Acoustical Society of America* 111.4 (2002), pp. 1917–1930.

[10] Henrik von Coler. "TU-Note Violin Sample Library – A Database of Violin Sounds with Segmentation Ground Truth". In: *Proceedings of the 21st International Conference on Digital Audio Effects (DAFx-18)*. Aveiro, Portugal, 2018.

[11] Henrik von Coler, Jonas Margraf, and Paul Schuladen. *TU-Note Violin Sample Library*. TU-Berlin, 2018. DOI: 10.14279/depositonce-6747.

[12] *JACK API Website*. 2018. URL: http://jackaudio.org/.

[13] Miller S Puckette. "Pure Data." In: *Proceedings of the International Computer Music Conference (ICMC)*. San Francisco, 1996, pp. 224–227.