

# Evaluation of Audio-Based Feedback Technologies for Bow Learning Technique in Violin Beginners

Angel Blanco

Music and Machine Learning Lab, Universitat Pompeu  
Fabra  
C. de Roc Boronat, 138  
Barcelona, Spain 08018  
adavid.blanco@upf.edu

Rafael Ramirez

Music and Machine Learning Lab, Universitat Pompeu  
Fabra  
C. de Roc Boronat, 138  
Barcelona, Spain 08018  
rafael.ramirez@upf.edu

## ABSTRACT

We present a study of the effects of feedback technologies on the learning process of novice violin students. Twenty-one subjects participated in our experiment, divided into two groups: Beginners (participants with no prior violin playing experience,  $N=14$ ), and experts (participants with more than 6 years of violin playing experience,  $N=7$ ). The beginners group was further divided into two: a group of beginners learning with Youtube videos ( $N=7$ ), and a group of beginners with additional feedback related to the quality of their performance ( $N=7$ ). Participants were asked to perform a violin exercise during 21 trials while their audio was recorded and analyzed. Three different audio descriptors were extracted from each audio in order to evaluate the quality of the performance: Dynamic stability, pitch stability and aperiodicity. Beginners showed a significant improvement during the session (i.e. by comparing the beginning and the end of the session) in the quality of the sound recorded, while experts maintained their results. However, only the beginner group with feedback showed significant improvement between the middle and late part of the session, while the group without feedback remained stable.

## CCS CONCEPTS

•Applied computing → Computer-assisted instruction;

## KEYWORDS

Signal Processing, Audio descriptors, Violin, Learning

### ACM Reference format:

Angel Blanco and Rafael Ramirez. 2017. Evaluation of Audio-Based Feedback Technologies for Bow Learning Technique in Violin Beginners. In *Proceedings of 1st ACM SIGCHI International Workshop on Multimodal Interaction for Education, Glasgow, UK, November 13, 2017 (MIE'17)*, 3 pages. DOI: 10.1145/3139513.3139520

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

MIE'17, Glasgow, UK

© 2017 Copyright held by the owner/author(s). Publication rights licensed to ACM.  
978-1-4503-5557-5/17/11...\$15.00  
DOI: 10.1145/3139513.3139520

## 1 INTRODUCTION

Technology enhanced learning (TEL) or e-learning are terms which refer to a relatively new field characterized by the use of new technological tools and interfaces to improve the process of learning. In particular, TEL has been applied with great success to music theory learning. However, there is still a long way in the development of applications able to provide useful real-time audio and gesture feedback about music performance to students. Some efforts in this direction are the i-Maestro project[8], which aim was to improve performance learning with multimodal technology support using wireless modules and a sensor setup, whereas companies such as Riyaz [1] are helping people during the process of learning to sing through the use of instantaneous feedback using audio descriptors. In this work we aim to explore the effects of feedback technologies on the learning process of novice violin students. With this purpose, we explore the usefulness of audio descriptors (e.g. dynamic stability, pitch stability). Romani et al.[11] showed that there is a correlation between audio descriptors extracted from the recordings of professional musicians and their subjective scores about the quality of the generated sound. In this work we aim to evaluate the usefulness of real-time feedback based on audio descriptors in novice violin students.

## 2 MATERIALS AND METHODS

### 2.1 Participants

Twenty-one adults participated in the study. Participants granted their written consent and procedures were positively evaluated by the CIREP, Barcelona, Spain. Participants were asked to fill a form with questions regarding their level of musical studies and their primary instruments. Participants with ample experience in violin playing conformed the expert group (EG; 6 male, 1 female; mean age: 35.2(9.01); mean years studying violin: 7.6(2.19)) while participants with no prior violin playing (and neither any bowed string instrument, e.g. cello, double-bass) experience were included in the beginners group. The beginners group were divided into two different groups: a group of beginners learning with Youtube videos (BNF; 6 male, 1 female; mean age: 28.4 (3.12); all of them musicians with years of experience, mean: 10.4 (4.12)), and a group of beginners with additional feedback related to the quality of their performance (BF; 5 male, 2 female; mean age: 26.85 (4.90); all of them musicians with years of experience, mean: 9.3 (5.11)). Participants were recruited in person at the university campus. Before starting, participants received a written consent form and

were informed about their task, the experimental procedures and their right to withdraw from the experiment at any moment.

## 2.2 Materials

All participants used the same violin and the sound of each violin trial was recorded using a Zoom recorder [3] and processed later in Matlab [10] using the "Yin Algorithm" [5]. Yin is a commonly used pitch detection algorithm based upon autocorrelation to extract audio features in order to assess the quality of the audio produced by the subjects.

## 2.3 Method

Beginner participants (i.e. in BF and BNF) were shown a ten minute instructional video on stance and violin position [2]. All participants in the study were asked to play twenty-one trials consisting of four up and down bowing movements (playing the A open string), with the aim to achieve stable tone and dynamics. Participants were free to take as many breaks as they wanted through the experiment although, every five trials participants had the opportunity to review as many times as they wanted the reference expert video. The BF group was able to visualize after each trial their score in Dynamic Stability, Pitch Stability and Aperiodicity represented in a graph together with their past scores.

**2.3.1 Extraction of audio features.** Generated violin sound was recorded for each trial and processed independently. First, we extracted sound descriptors from the audio signal of each trial using the Yin algorithm implementation in Matlab [9]. This Matlab implementation first windows the signal, using a windows size which depends in the sample rate and the minimum frequency (30Hz by default), and for each windows it computes three different parameters which are: the fundamental frequency in octaves (reference: 440), the aperiodicity measure (the ratio of aperiodic to total power), and the period-smoothed instantaneous power. With those parameters we can compute the inverse values of sound descriptors such as Dynamic Stability or Pitch Stability by computing the standard deviation of both power and fundamental frequency, which can be used to assess sound quality [11]. Finally, the inverse values of Dynamic Stability and Pitch Stability together with Aperiodicity were normalized by subtracting the mean. The average value of the three descriptors for each trial conforms a unique descriptor called Sound Instability.

## 3 RESULTS

The results of Sound Instability along trials of the beginners compared with experts can be seen in Fig.1. As a first step, we did a classification task in Weka [6] to see if the features were appropriate to discriminate between beginners and experts. We used the OneR [7] classifier for each feature. The total number of instances was 522 (378 sounds from beginners sound and 144 sounds from experts). ZeroR was used as a baseline showing results of 72.41%. A summary of the classification accuracy for each feature can be seen in Table 1 and 2.

The number of trials were divided into three time periods: Early (trials 1-7), Middle (trials 7-14) and Late (trials 15-21). A two-way mixed ANOVA (3x3) was then performed in SPSS [4] with time as a within-subject factor (time at early, time at middle, and time

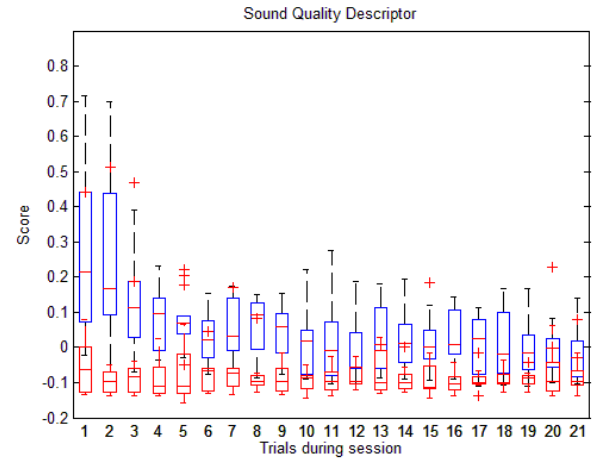
**Table 1: Table 1: Descriptors chosen to classify the sounds and their classification accuracy.\*Significant difference in values compared to baseline (ZeroR: 67.12%)**

Classifier	Dynamic Stability	Pith Stability
OneR	79.45%*	85.84%*

**Table 2: Table 2: Descriptors chosen to classify the sounds and their classification accuracy.\*Significant difference in values compared to baseline (ZeroR: 67.12%)**

Classifier	Aperiodicity	Sound Stability
OneR	82.19%*	82.87%*

at late) and Group as a between-subject factor (BF, BNF and EG). Before that, a Shapiro-Wilk test for normality was performed on the data showing not significant results and after, a Mauchly's Test of Sphericity showing a significance value and thus fulfilling the assumptions.



**Figure 1: Boxplot of errors in mean phrase velocity predictions.**

Results revealed significant differences at time and at the interaction between time\*groups. Pairwise comparisons realized with the Bonferroni correction for the p-value (at  $p=0.05$ ) showed significant differences between experts and both beginners group, but not significance between beginners with and without feedback. Significant values were also found in time between the early and the middle and late period. No significant differences were seen between middle and late. Differences among means between the middle and late period of both groups of beginners (BF and BNF) showed how the group with feedback (BF) improved 27.87% more. Two paired t-test were performed between the middle and late period of time, one for each group. After adjusting the p-value for multiple comparison tests using the Bonferroni method for a number of three tests, no significant differences were observed in

the group without feedback ( $p = 0.603$ ) while significant ones were found for the group with feedback. ( $p = 0.024$ ).

## 4 CONCLUSIONS

In this work we have validated the use of dynamic stability, pitch stability and aperiodicity features to discriminate between good and bad sound confirming previous related work [11]. However, in this work we have shown how these descriptors were able not only to discriminate between sounds generated by beginners and experts but also, were able to track the learning process of the beginner group and therefore allowing us to study the impact of feedback technologies. Whereas experts did not seem to show any significant improvement along the session, both beginners group showed significant improvements between the early trials and the rest of the session. However, it is important to remark that only the beginner group with feedback (BF) showed improvement between the middle and late part of the session, while results of the group without feedback (BNF) remained stable. These results support the relevance of technology for providing feedback related to specific learning tasks. Low-cost technologies accessible to everyone (e.g. a microphone) are sufficient to allow tracking the learning process of students. This kind of technology may not be offering only additional information to the regular response of the instrument, but also motivation to students. Long-term effects of such technologies are out of the scope of this paper and should be further investigated.

## 5 ACKNOWLEDGEMENTS

This work has been partly sponsored by the Spanish TIN project TIMUL (TIN 2013-48152-C2-2-R), the European Union Horizon 2020 research and innovation programme under grant agreement No. 688269 (TELMi project), and the Spanish Ministry of Economy and Competitiveness under the Maria de Maeztu Units of Excellence Programme (MDM-2015-0502).

## REFERENCES

- [1] 2017-09. Riyaz project website. (2017-09). <https://riyazapp.com/>
- [2] 2017-09. Violin Master Class. Available online at: (2017-09). <http://www.violinmasterclass.com/en/masterclasses/stance-violin-position>.
- [3] 2017-09. Zoom Recorder. (2017-09). <https://www.zoom-na.com/es/products/field-video-recording/field-recording/zoom-h4n-handly-recorder>
- [4] IBM Corp. 2011. *Released 2011. IBM SPSS Statistics for Windows*. Version 20.0. Armonk, NY: IBM Corp.
- [5] Alain de Cheveigne and Hideki Kawahara. 2002. YIN, a fundamental frequency estimator for speech and music. *Acoustical Society of America*. (2002). DOI: <http://dx.doi.org/10.1121/1.1458024>
- [6] Mark A. Hall Eibe Frank and Ian H. Witten. 2016. *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"*. Morgan Kaufmann, Fourth Edition, 2016.
- [7] R. C. Holte. 1993. Very simple classification rules perform well on most commonly used datasets. *Machine Learning*. 11:63-91. (1993).
- [8] Paolo Nesi Kia Ng. 2008. i-Maestro: Technology-Enhanced Learning and Teaching for Music.. In *Proceedings of the 2008 Conference on New Interfaces for Musical Expression (NIME08)*, Genova, Italy: 225–228.
- [9] Quim Llimona. 2015. YIN pitch estimation toolbox. (2015). <https://github.com/lemonzi/matlab/tree/master/yin>
- [10] MATLAB. 2010. *version 7.10.0 (R2010a)*. The MathWorks Inc., Natick, Massachusetts.
- [11] Dabiri D. Tokuda H. Hariya W. Oishi K Roman Picas O., Parra Rodriguez H. and Serra X. 2015. A real-time system for measuring sound goodness in instrumental sounds. In *138th Audio Engineering Society Convention*.