Audio Engineering Society

Convention Paper

Presented at the 124th Convention
2008 May 17–20 Amsterdam, The Netherlands

Seeing the Inaudible. Descriptors for generating objective and reproducible data in real-time for musical instrument playing standard situations

Tobias Grosshauser[1], Diemo Schwarz[2], and Norbert Schnell[3]

[1] IRCAM, Paris, 75004, France
Tobias@Grosshauser.de

[2] IRCAM, Paris, 75004, France
Diemo.Schwarz@IRCAM.fr

[3] IRCAM, Paris, 75004, France
Norbert.Schnell@IRCAM.fr

ABSTRACT

This article describes a method to generate objective and reproducible data to assist instrument teaching and practicing. The method is based on using audio descriptors and their efficient visualisation that assist in the perception of musical parameters difficult to hear. To aid comparison, we defined and recorded a comprehensive database of positive and negative sound examples from the violin that encompasses frequent mistakes made by students and a wide variety of playing styles.

1. INTRODUCTION

The fine control of sound quality and articulation is a basic skill for all instruments and an important issue in instrumental teaching. Audio descriptors extracted from the sound captured from an instrument performance objectively and reproducibly represent certain aspects of the sound quality and their development over time.

Defined objective parameters, adjustable in their visualisation, show new possibilities in practising situations of students at home and alone, preventing malfunctions and ineffective practising by being unfocussed, or in distance or internet teaching environments with low audio quality. The combination of two or more descriptors can visualize parameters that are difficult to perceive, like the length or hardness of the attack of a certain note, different vibrato styles, or bowing changes.

Even though, regarding the current state of research on audio descriptors for the analysis of instrument performance, it is difficult to envisage any reasoning based on absolute values, the representation, rendering and comparison of audio descriptor values has immediate applications in music pedagogy. Already a simple graphical representation of audio descriptors over time permits to illustrate and compare particular characteristics of instrument performance in the context of teaching and learning.

The ability to link graphical representations of audio descriptor values to pertinent features of a recorded and analysed instrument performance can be supported by allowing for interactive exploration of the representation. The interactive exploration permits the listener to intuitively learn the correspondence of the morphology of the graphical representation with the perceived sound as well as the contained musical and sonic features judged to be relevant in a given context.

To study the correspondence of particular audio descriptors to pertinent characteristics of violin performance using the developed technology, a database of recorded violin sound examples has been assembled.

2. BACKGROUND

Brian Moore, the author of the fundamental book about psychoacoustics „An introduction to the Psychology of Hearing“[1] started the sound representation with his software „Loudaes“ , used mainly by violin manufacturers. The missing link to the musicians everyday problems is a real-time sound analysing software with objective and visual sound descriptors. Studio On-Line [2], Music Minus One [3], Vienna Symphonic Library [4] also represent huge sound databases, but only with perfect but rigid sounds.

2.1. Existing Products and Research

First attempts are described in [5], where several descriptors are explored by using recorded beginner and good player legato notes. Hämäläinen et al. developed singing aid in [6], which describes the use of pitch-based control of a game character by the user's voice. J. A. Charles [7] tries to implement a feedback system for computer aided exercising.

2.2. Goal of the research

The goal of this research, is to be able to visualise sound phenomena that are hard to hear and explain by words in teaching, learning or exercising situations. Not only on comparing single tones, but moreover setting up many realistic combinations of playing techniques and matching descriptors.

3. OBJECTIVE RESULTS

Sound and articulation is a parameter to be trained by every musician to develop his personality and individual expression. Certain ranges of expression are obligatory, to be able to fulfil the needs of playing techniques of all periods. This is the situation, where objective, but personally definable, and visualized data are useful.

We have defined, recorded, and labelled a database of violin sounds that, contrary to existing commercial and research databases, does not only focus on the different playing styles and techniques when played correctly, but presents many sorts of badly played notes with a wide range of errors a violin player can make. Mainly, playing parameters that affect a single aspect of sound quality were chosen and included in the database. An example would be: Too much bow speed, but correct bow position, pressure, angle.

3.1. Database containing typical basic characteristics of violin playing

The following tables give an overview of the sound database with the most used and nearly complete variations and possibilities of violin sounds. Table 1 shows the different parameters of playing modes, table 2 the parameters determining the timbral quality of sound, and table 3 special playing techniques.

Playing Modes		
Volume	high	low
Playing speed	slow bow movement slow/fast fingering	fast bow movement slow/fast fingering
Duration	short	long
Finger	open strings	fingers
Quantity	single notes	2 or more notes
Bowing Direction	up-bow	down-bow

Table 1 Playing Modes

Envelope and Timbre		
Attack	hard	soft
Sustain	stationary	developing
Release	hard	soft
Vibrato	with vibrato	without vibrato
Duration	short	long

Table 2 Envelope and Timbre

Playing techniques
Muted, Flagolett, FingerPizz, Sul Testo/Flauto, Double Stops, Warbler, Tremolo

Table 3 Playing Technics

3.2. Data Acquisition and Evaluation

The database is similar to existing ones, for instance in the work of Rasamimanana [8] or the IRCAM Studio On Line (SOL) database [2]. But there will be more playing technics included and defined “wrong” tones with one defined bad parameter.
All samples were recorded with the same microphones and recording hardware. Three different violins (O. Spidlen, Fiekker, anonymus), one bow and one violin player completed the set-up.

4. STRUCTURE OF THE DATA BASE

Table 4 and 5 shows the abreviations used in the data base:

Basic Description

Violin Used	Note,Fingers	String,Stage	Dyn <>, f, p
V1	H1	A1	f>p

Playing Mode	FullBow/HalfBow/Frog/Tip	SoundShaping
	V,]	
Det	FB	Vib

Table 5 Description of Abreviations

Violin used: V1 here e. g. O. Spidlen Violin
Note, Fingers: Fingers 1 to 4, Notes C to H
String, Stage: Strings GDAE, Fingers 1234
Dynamic: pp, p, mf, f, ff, sp, sf
Playing Mode: see table1
FullBow/HalfBow/Frog/Tip: FB Full Bow, HB HalfBow, only Frog, only Tip, V: Up- Bow,]:Down Bow
Sound Shaping: See table1
Too much pressure/less pressure: MP: too much pressure, LP: too less pressure
Too much/less bow speed: Mspeed: too much bow speed, Lspeed:too less bow speed
Bow Position, Distance to the bridge: C: too close, F: too far to/from the bridge
Bow Angle: R: Right/ N: Non 90 degree
Specials: Free space for new and future descriptions

File Name Example: V1_H1_A1_f>p_Det_FB_Vib_MP_-_-_-_-;

This sound file is played with: violin1_first finger_on string A_first stage_played from forte to piano_playing mode is detache_the full bow is used_it is played with vibrato_with too much bow pressure_the rest faults are not played.

5. CHOICE OF DESCRIPTORS:

Audio descriptors as well as techniques of extraction of descriptors from audio streams have been the subject of a large body of scientific work over the past years [9]. Fewer works have been dedicated to the use of audio descriptors in the framework of pedagogy.

Audio descriptors are a numerical representation of acoustic features [10] extracted from audio streams. Descriptors come in different forms: We distinguish instantaneous descriptors from global descriptors:

- *Instantaneous descriptors* take a value for each short frame of the audio signal at regularly spaced intervals. Typically, this gives a sequence of values every 20–40 ms. Examples of instantaneous descriptors are loudness, pitch, and spectral centroid.
- *Global descriptors* give a value for a longer segment of sound, for instance a note or phrase. Examples of global descriptors are attack time, presence of vibrato, amount of variation in loudness, etc. Note that, to derive musically meaningful segments, we would need a robust

segmentation of the audio signal, or limit the recording to one segment (note) only.

The following section gives an overview of the audio descriptors taken into consideration for highlighting sound quality in instrument pedagogy. For each presented descriptor, we will give a brief description of their technical definition, and the relation to which pedagogical concept. Details of their mathematical definition can be found in [11].

Effort:

The *effort* descriptor is defined as the time, the signal's energy goes from one predefined threshold value to the next threshold value, as illustrated in [fig 4] from [8].

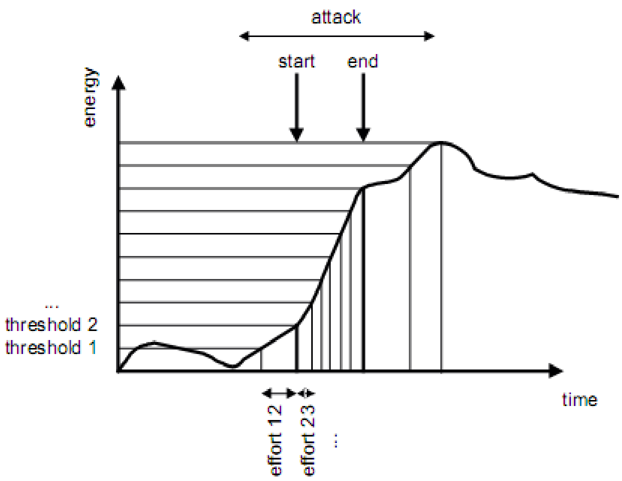

fig. 4 Attack Analysis [8]

In pedagogy, it is related to articulation in combination with different playing techniques. Its visualisation highlights the attack/articulation of a sound. As attack often is difficult to hear, because of the shortness of the event, it is visualised in fig. 5 by two descriptors Effort and Briliance, the lighter (blue) line represents the Effort (attack), the dark (black) line, Brilliance the rising brilliant sound after the decrease of the attack. This visualisation can raise awareness of a teacher or student about of the noise at the beginning and ending of the bowing.

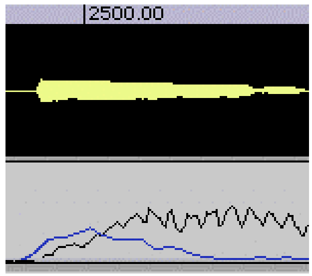

fig. 5 Bow Stroke with Attack (blue line) to Brilliance (black line) Change

Periodicity:

Periodicity measures the regularity of the audio signal, which is related to its harmonicity. It is low for inharmonic and noisy signals. It can be measured by the *Yin* algorithm as the correlation factor with a sine wave at the fundamental frequency, or by the zero-crossing rate (ZCR), defined as the number of times the sampled signal waveform crosses the zero level. Periodicity is higher for noisy segments of the signal. Therefore, it can be used to detect attacks or fricatives.

In pedagogy, *periodicity* it is useful in combination with *effort*, showing transition from attack to decay and sustain.

Pitch:

The *fundamental frequency* is the physical correlate of the perception of pitch. It can be given in Hz or on the logarithmic scale of MIDI note numbers.

Visualising the steadiness of the pitch and vibrato speed and amplitude makes it possible to see the shape of vibrato, amplitude, regularity and acceleration [fig. 6].

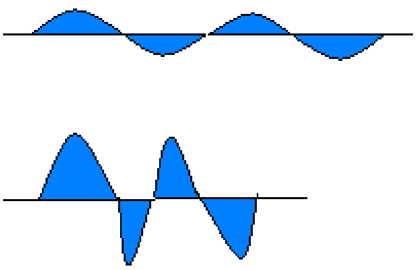

fig. 6 Regular versus irregular Vibrato

Brilliance:

The *spectral centroid*, or center of gravity of the short-time Fourier magnitude spectrum, is a correlate of the perception of brilliance. It offers the possibility to analyse the brilliance of certain parts of a piece, and, in comparison, to visualise the strength or sound quality of a note.

Visualisation of "Brilliance" [fig. 7] with two violin tones. First tone (A) with "bad", weakly sound quality, tone (B) with good sound quality:

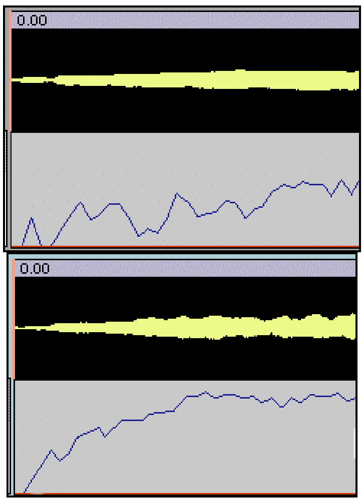

fig. 7 Two violin tones with different "Brilliance"

Although the wave form looks similar, the brilliance is quite different. In pedagogy this allows objective measurements in the need of the development of sound quality. In practicing situations of students paying alone, this could help to achive similar goals than in live teaching situations.

Loudness:

Loudness is related to the perception of the dynamic of a sound, modelling of its evolution along time of the dynamic of a sound can be achieved using a polynomial representation.

In pedagogy, it allows to work on the development and/or steadiness of the volume in a short section or a whole piece of music.

6. CONCLUSION AND FURTHER DEVELOPMENTS

Real violin sounds or instrument sounds in general have many independent parameters. Thus, furture work would consist in extending the existing database with further examples and parameters, in order to achieve a wider variety of results and comparison possibilities. However, fortunately we have an instrument efficient enough to deal with the visualisation: the human brain.

7. ACKNOWLEDGEMENTS

We would like to thank the members of the Real-time Music Interaction team (IMTR) at Ircam for their participation in this research, especially Norbert Schnell, Fabrice Guedy, Frederic Bevilacqua, Nicolas Rasamimanana, Jean-Philippe Lambert. This work is partially conducted in the framework of the I-MAESTRO project. The I-MAESTRO project is partially supported by the European Community under the Information Society Technologies (IST) priority of the 6th Framework Programme for R&D (IST-026883, www.i- maestro.org). Thanks to all I-MAESTRO project partners and participants, for their interest, contribution and collaboration.

8. REFERENCES

[1] Brian C. J. Moore, "An Introduction to the Psychology of Hearing", ISBN: 0125056281
[2] http://exlibris.ircam.fr/75.html?&L=1
[3] http://www.musicminus1.de/
[4] http://vsl.co.at/
[5] Development of a Computer-Based Violin Teaching Aid: ViTool, Charles, Jane; FitzGerald, Derry ; Coyle, Eugene, 118th AES Convention, Barcelona, Spain, 2005
[6] Hämäläinen, P., Mäki-Patola,T., Pulkki, V., Airas, M. 'Musical Computer Games Played by Singing', Proc. 7th Int. Conf. on Digital Audio Effects (DAFx'04), Naples, Oct. 5-8, 2004.
[7] Violin Timbre Space Features, Charles, Jane; FitzGerald, Derry ; Coyle, Eugene, Proceedings of the Irish Signals and Systems Conference, Dublin,Ireland, 2006
[8] N. H. Rasamimanana, Gesture Analysis of Violin Bow Strokes, 6th International Gesture Workshop, GW 2005, Gesture in Human-Computer Interaction and Simulation, Sylvie Gibet, Nicolas Courty, Jean-Francois Kamp (Eds.), Springer
[9] Peeters, McAdams, Herrera 2000
[10] Thom, Purnhagen, MPEG AudioSubgroup 1998
[11] Geoffroy Peeters, A large set of audio features for sound description (similarity and classification) in the Cuidado project, Ircam -- Centre Pompidou, 2004