TU-NOTE VIOLIN SAMPLE LIBRARY – A DATABASE OF VIOLIN SOUNDS WITH SEGMENTATION GROUND TRUTH

Henrik von Coler

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

ABSTRACT

The presented sample library of violin sounds is designed as a tool for the research, development and testing of sound analysis/synthesis algorithms. The library features single sounds which cover the entire frequency range of the instrument in four dynamic levels, two-note sequences for the study of note transitions and vibrato, as well as solo pieces for performance analysis. All parts come with a hand-labeled segmentation ground truth which mark attack, release and transition/transient segments. Additional relevant information on the samples' properties is provided for single sounds and two-note sequences. Recordings took place in an anechoic chamber with a professional violinist and a recording engineer, using two microphone positions. This document describes the content and the recording setup in detail, alongside basic statistical properties of the data.

1. INTRODUCTION

Sample libraries for the use in music production are manifold. Ever since digital recording and storage technology made it possible, they have been created for most known instruments. Commercial products like the *Vienna Symphonic Library*[1] or *The EastWest Quantum Leap*[2] offer high quality samples with many additional techniques for expressive sample based synthesis. For several reasons, these libraries are not best suited for the use in research on sound analysis and synthesis. Many relevant details are subject to business secrets and thus not documented. Copyright issues may prevent a free use as desired in a scientific application. These libraries also lack annotation and metadata which is essential for research applications, if used for machine learning or sound analysis / synthesis tasks.

The audio research community has released several databases with single instrument sounds in the past, usually closely related to a specific aspect. Libraries like the *RWC* [1] or the *MUMS* [2] aim at genre or instrument classification and timbre analysis [3]. Databases for onset and transient detection which include hand labeled onset segments have been presented by Bello et al. [4] and von Coler et al. [5].

The presented library of violin sounds is designed as a tool for the research, development and testing of sound analysis/synthesis algorithms or machine learning tasks. The contained data is structured to enable the training of sinusoidal modeling systems which distinguish between stationary and transient segments. By design, the library allows the analysis of several performance aspects, such as different articulation styles, glissando [6] and vibrato. It features recordings of a violin in an anechoic chamber and consists of three parts:

1. single sounds
2. two-note sequences
3. solo (scales and compositions/excerpts)

For single sounds and two-note sequences, hand-labeled segmentation files are delivered with the data set. These files focus on the distinction between steady state and transient or transitional segments. The prepared audio files and the segmentation files are uploaded to a static repository with a DOI [7][3]. A *Creative Commons BY-ND 4.0* license ensures the unaltered distribution of the library.

The purpose of this paper is a more thorough introduction of the library. Section 2 will explain the composition of the content, followed by details on the recording setup and procedure in Section 3. The segmentation data will be introduced in Section 4. Section 5 presents selected statistical properties of the sample library. Final remarks are included in Section 6.

2. CONTENT DESCRIPTION

2.1. Single Sounds

Similar to libraries for sample based instruments, the single sounds capture the dynamic and frequency range of the violin, using sustained sounds. The violinist was instructed to play the sounds as long as possible, using just one bow, without any expression. Steady state segments, respectively the sustain parts, of these notes are thus as played as steady as possible. This task showed to be highly demanding and unusual, even for an experienced concert violinist.

On all of the four strings, the number of semitones listed in Table 1 was captured, each starting with the open string. This leads to a total of 84 positions. All positions are captured in four dynamic levels which were specified as **pp - mp - mf - ff** resulting in a total amount of 336 single sounds. According to Meyer [8], the dynamic interval interval of a violin covers a range from $58 \ldots 99\,\text{dB}$.

[1] www.vsl.co.at/

[2] http://www.soundsonline.com/symphonic-orchestra

[3] https://depositonce.tu-berlin.de//handle/11303/7527

Table 1: Number of positions on each string

String	Positions
G	18
D	18
A	18
E	30

Each item was recorded in several takes, until recording engineer, the author and the violinist agreed on success. Although all sounds were explicitly captured in both up- and down-stroke techniques, these modes have not been considered individually in the data set and thus appear randomly.

2.2. Two-Note Sequences

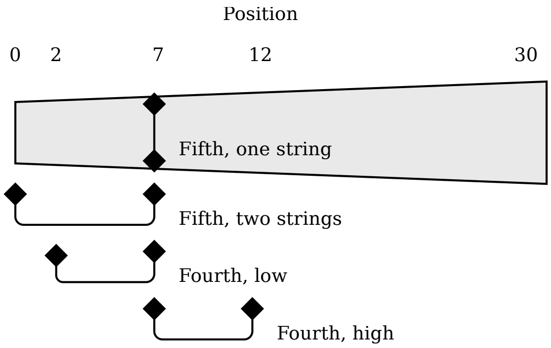


Figure 1: Violin board with positions for two-note sequences

For the study of basic articulation styles, a set of two-note sequences was recorded at different intervals, listed in Table 2. The respective positions on the board are visualized in Figure 1. All combinations were recorded at two dynamic levels **mp** and **ff**. Three different articulation styles (*detached, legato, glissando*) were used and some combinations were captured with additional vibrato. These combinations lead to a grand total of 344 two-note items.

5 semitones on one string were captured in 8 pairs with 24 versions (2 dynamic levels, 2 directions, with and without vibrato, 3 articulation styles): $2 \cdot 2 \cdot 3 = 24$.

Repeated tones were captured in 4 pairs with 6 versions (2 dynamic levels, legato and detached, the latter with and without vibrato): $2^2 + 2 = 6$

7 semitones on one string were captured in pairs with 20 versions (2 dynamic levels, two directions, detached only without vibrato, legato and glissando with and without vibrato): $2 \cdot 2 + 2^4 = 20$

7 semitones on two strings were captured in 3 pairs with 16 versions (2 dynamic levels, two directions, with and without vibrato and two articulation styles [legato, detached]):$2^4 = 16$

Table 2: All two-note pairs

5 semitones, one string						
Two-note	Note 1			Note 2		
item no.	ISO	Pos.	String	ISO	Pos.	String
01-24	D4	7	G	A3	2	1
25-48	A4	7	D	E4	2	2
49-72	E5	7	A	B4	2	3
73-96	B5	7	E	F#5	2	4
97-120	D4	7	G	G4	12	1
121-144	A4	7	D	D5	12	2
145-168	E5	7	A	A5	12	3
169-192	B	7	E	E6	13	4
Repeated tones						
Two-note	Note 1			Note 2		
item no.	ISO	Pos.	String	ISO	Pos.	String
193-198	D4	7	G	D4	7	G
199-204	A4	7	D	A4	7	D
205-210	E5	7	A	E5	7	A
211-216	B5	7	E	B5	7	E
7 semitones, one string						
Two-note	Note 1			Note 2		
item no.	ISO	Pos.	String	ISO	Pos.	String
217-236	D4	7	G	G3	0	G
237-256	A4	7	D	D4	0	D
257-276	E5	7	A	A4	0	A
277-296	B5	7	E	E5	0	E
7 semitones, two strings						
Two-note	Note 1			Note 2		
item no.	ISO	Pos.	String	ISO	Pos.	String
297-312	D4	7	G	A4	7	D
313-328	A4	7	D	E5	7	A
329-344	E5	7	A	B5	7	E

2.3. Solo: Scales and Compositions

Two scales – an ascending major scale and a descending minor scale – were each played in three interpretation styles, as listed in Table 3. The first style was plain, without any expressive gestures, followed by two expressive interpretations. Six solo pieces and excerpts, listed in Table 4 which mostly contain cantabile legato passages were recorded. All compositions were proposed by the violinist, ensuring familiarity with the material.

Table 3: Scales in the solo part

Item	Type	Interpretation
01	major, ascending	plain
02	major, ascending	expressive 1
03	major, ascending	expressive 2
04	minor, descending	plain
05	minor, descending	expressive 1
06	minor, descending	expressive 2

Table 4: Solo recordings

Item	Composition	Composer
07	Sonata in A major for Violin and Piano	César Franck
08	Violin Concerto in E minor, Op. 64, 2nd movement	Felix Mendelssohn
09	Méditation (Thaïs)	Jules Massenet
10	Chaconne in g minor	Tomaso Antonio Vitali
11	Violin Concerto in E minor, Op. 64, 3rd movement	Felix Mendelssohn
12	Violin Sonata no.5, Op.24, 12s movement	Ludwig van Beethoven

3. RECORDING SETUP

The recordings took place in the anechoic chamber at SIM[4], Berlin. Above a cutoff frequency of 100 Hz the room shows an attenuation coefficient of $\mu > 0.99$, hence the recordings are free of reverberation in the relevant frequency range. The recordings were conducted within two days, taking one day for the single sounds and the second day for two-note sequences and solo pieces. All material was captured with a sample-rate of of 96 kHz and a depth of 24 Bit.

Microphones

The following microphones were used:

- 1x DPA 4099 cardiod clip microphone
- 1x Brüel & Kjær 4006 omnidirectional small diaphragm microphone with free-field equalization, henceforth BuK

The DPA microphone was mounted as shown in Figure 2, above the lower end of the f-hole in 2 cm distance. Due to its fixed position, movements of the musician do not influence the recording. The B&K microphone was mounted in 1.5 m distance above the instrument, at an elevation angle of approximately 45°, as shown in Figure 3.

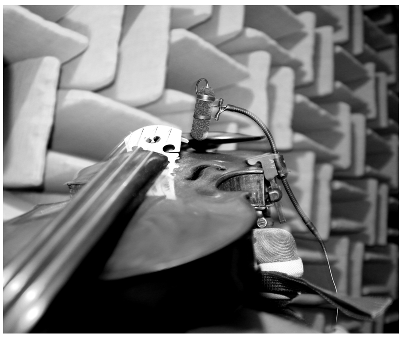


Figure 2: Position of the DPA microphone

[4] http://www.sim.spk-berlin.de/refelxionsarmer_raum_544.html

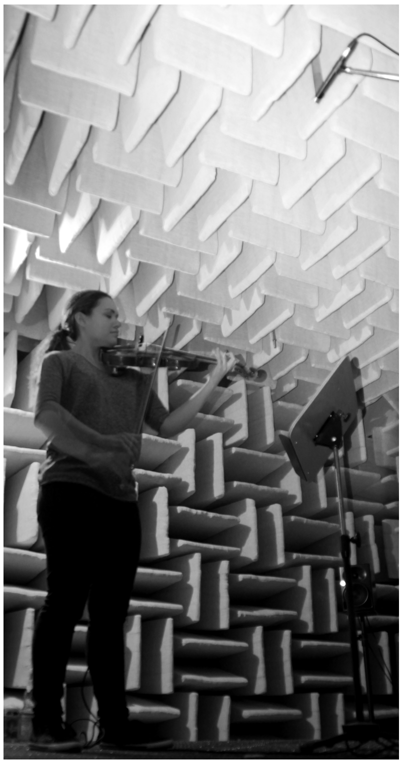
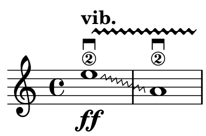


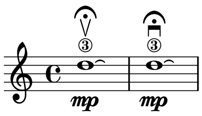
Figure 3: Position of the B&K microphone

Instructions

For each of the single-sound, two-note and scale items, a minimal score snippet was generated using *LilyPond* [9]. Examples for items' instructions are shown in Fig. 4. The resulting 63 page score was then used to guide the recordings. Although the isolated tasks may seem simple and unambiguous, this procedure ensured smooth recording sessions.



(a) Two-note example with vibrato and glissando

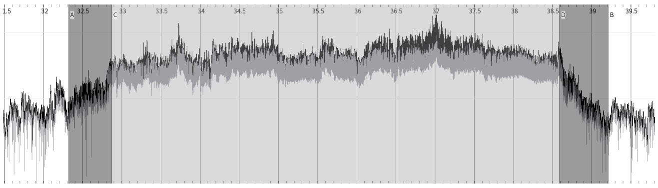


(b) Single-sound example with upbow and downbow

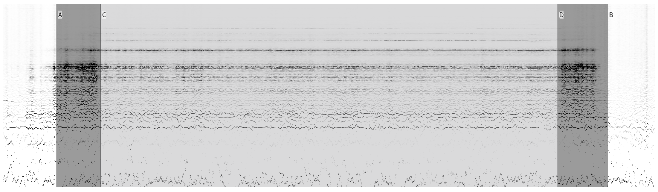
Figure 4: Instruction scores for two-note a and single-sound b

4. SEGMENTATION

The segmentation of a monophonic musical performance into notes, and even more into a note's subsegments is not trivial [10, 11]. During the labeling process, the best of the takes for each item was selected from the raw recordings and the manual segmentation scheme proposed by by von Coler et al. [5] was applied using *Sonic Visualiser* [12].



(a) Energy trajectory



(b) Peak frequency spectrogram

Figure 5: Sonic Visualiser setup for annotation of single sound 333

4.1. Single Sounds

Each single sound is divided into three segments, which are defined by four location markers in the segmentation files[5], as shown in Table 5. The first time instant (`A`) marks the beginning of the attack segment, the second instant (`C`) marks the end of the attack segment, respectively the beginning of the sustain part. The end of the sustain, which is also the beginning of the release segment, is labeled with the (`D`). The label (`B`) marks the end of the release portion and the complete sound. The left column holds the related time instants in seconds.

Table 5: Example for a single-sound segmentation file (`SampLib_DPA_01.txt`)

0.000000	A
0.940646	C
7.373000	D
8.730500	B

The definition of the attack segment is ambiguous in literature [13] and shall thus be specified for this context: Attack here refers to the actual *attack-transient*, the very first part of a sound with a significant inharmonic content and rapid fluctuations. In other contexts, the attack may be regarded the segment of rise in energy to the local maximum. Often, there is still a significant increase in energy after the attack-transient is finished. As the attack-transient is characterized by unsteady, evolving partials and low relative partial amplitudes, the manual segmentation process is performed using a temporal and a spectral representation. Figure 5 shows a typical Sonic Visualiser setup for the annotation of a single sound. The noisiness of the signal during attack and release can be seen in the spectral representation. How attack transient and rising slope may differ, is illustrated in Fig. 6. The gray area represents the labeled attack segment, which is finished before the end of the rising slope is reached.

Less ambiguous, the release part is labeled as the segment from the end of the excitation until the complete disappearance

[5]The segmentation files are part of the repository [7]

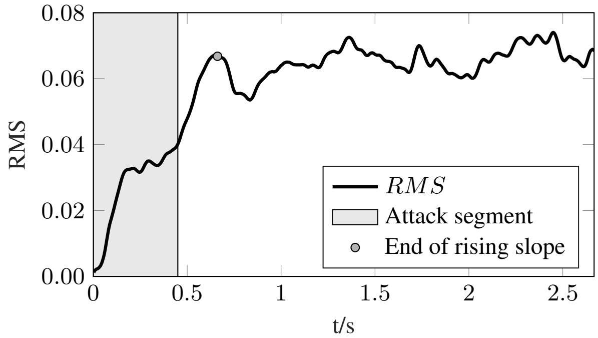


Figure 6: RMS trajectory of a note beginning with attack segment (gray) and end of the rising slope (single sound no. 19)

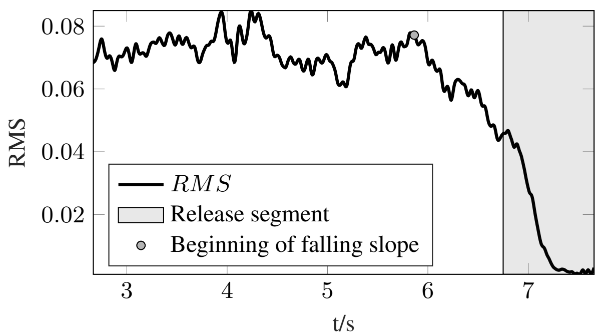


Figure 7: RMS trajectory of a note end with release segment (gray) and beginning of the falling slope (SampLib_19)

of the tone. As shown in Fig. 7, there is often a significant decrease in signal energy before the actual release starts. For items with low dynamics, the release is also covering the very last part of the excitation.

The *ease of annotation* varies between dynamic levels, as well as between the fundamental frequency of the items. Notes played at fortissimo show clear attack and decay segments with a steady sustain part, whereas pianissimo tones have less prominent boundary segments and a parabolic amplitude envelope. The higher SNR in fortissimo notes allows a better annotation of the transients. Tones with a high fundamental frequency have less prominent partials, whereas the bow noise is emphasized. They are thus more difficult to label, since attack transient are less clear in the spectrogram. The segmentation of high pitched notes at low velocities is hence most complicated.

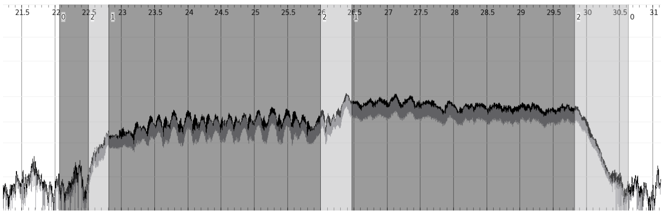
4.2. Two-Note Sequences

The two-note sequences contain the the segments note, rest and transition with the labels listed in Table 6. Stationary sustain parts are labeled as notes, whereas the transition class includes attack and release segments, as well as note transitions, such as glissando.

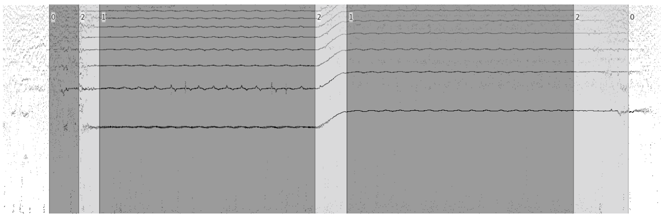
All two-note sequences follow the same sequence of segments (0-2-1-2-1-2). Figure 8 shows a labeling project in Sonic Visualiser for a two-note item with glissando. The transition segment is placed according to the slope of the glissando transition.

Table 6: Segments in the two-note labeling scheme

Label	**Segment**
0	rest
2	transition
1	note



(a) Energy trajectory



(b) Peak frequency spectrogram

Figure 8: Sonic Visualiser setup for annotation of two-note item 22

4.3. Solo

Solo items have been annotated using the guidelines proposed by von Coler et al. [5]. Due to the choice of the compositions, only few parts violated the restriction to pure monophony. Solo item 10, for example, starts with a chord, which is labeled as a single transitional segment.

5. STATISTICS

This section reports selected descriptive statistical properties of the sample library which are potentially useful when considering the use of the data.

5.1. Single Sounds

Fig. 9 shows the RMS for all single sounds, in box plots for each dynamic level. The median for the dynamic levels is logarithmically spaced.

Table 7: Segment length statistics for the single-sounds

	$\bar{l}/s$	μ/s
Attack	0.247	0.206
Sustain	5.296	1.118
Release	0.705	0.802

Statistics for the segment lengths of the single sounds are presented in Table 7 and Figure 10, respectively. With a mean of 5.296 s, the sustain segments are the longest, followed by release segments with a mean of 0.705 s. Attack segments have a mean

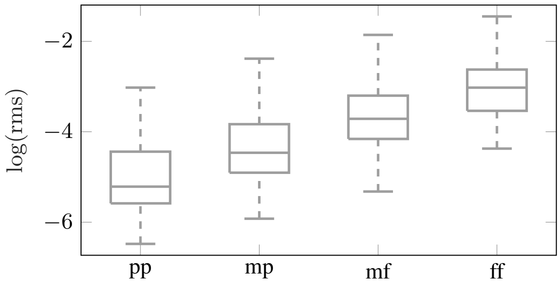


Figure 9: Boxplot of RMS for the sustain from the BuK microphone

length of 0.247 s. Extreme outliers in the mean attack length are caused by high pitched notes with low dynamics.

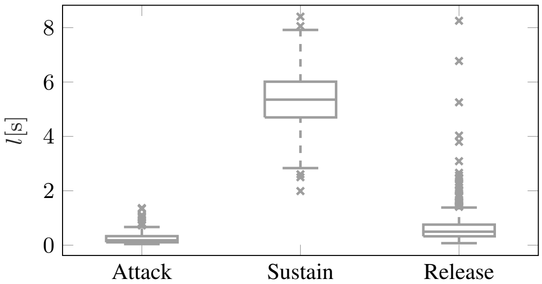


Figure 10: Box plots of segment lengths for all single sounds

5.2. Two-Note

The two-note sequences allow a comparison of different articulation styles. Figure 11 shows the lengths for detached, legato and glissando transitions in a box plot. With a median duration of 0.72 s, glissando transitions tend to be longer than legato (0.38 s) and detached (0.37 s) transitions.

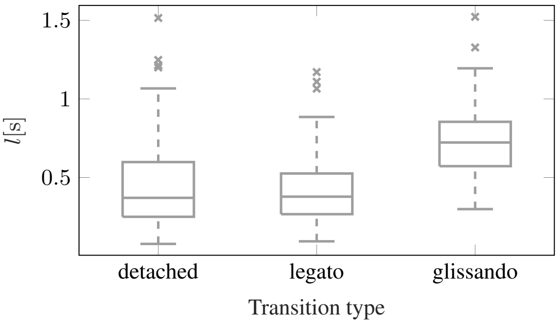


Figure 11: Box plot of transition lengths for all two-note sequences

5.3. Solo

Table 8: Note statistics for items in the solo category

Solo item	Number of notes	$\bar{l}/s$	μ/s
1	8	0.698	0.745
2	8	0.721	0.768
3	8	0.728	0.776
4	8	0.707	0.753
5	8	0.724	0.771
6	8	0.774	0.848
7	104	0.695	0.661
8	75	1.074	0.899
9	89	0.911	0.923
10	63	0.735	0.690
11	76	0.689	0.707
12	56	0.615	0.740

For the solo category, the basic statistics on the note occurrences and lengths are listed in Table 8. All scales (items 1 - 6) contain 8 notes, compositions (items 7-12) have a mean of 77 notes per item. With a mean note length of 0.614 906 s, item 12 has the shortest, and with 1.074 361 s, item 8 has the longest notes.

6. CONCLUSION

The presented sample library is already in application within sinusoidal modeling projects and for the analysis of expressive musical content. Overall recording quality proves to be well suited for most tasks in sound analysis. Since the segmentation ground truth follows strict rules and has undergone repeated reviews, it may be considered consistent.

7. ACKNOWLEDGMENTS

The author would like to thank the violin player, Michiko Feuerlein, and the sound engineer, Philipp Pawlowski, for their work during the recordings, as well as the SIM Berlin for the support. Further acknowledgment is addressed to Moritz Götz, Jonas Margraf, Paul Schuladen and Benjamin Wiemann for the contributions to the annotation.

8. REFERENCES

[1] Masataka Goto et al. "Development of the RWC music database". In: *Proceedings of the 18th International Congress on Acoustics (ICA 2004)*. Vol. 1. 2004, pp. 553–556.

[2] Tuomas Eerola and Rafael Ferrer. "Instrument library (MUMS) revised". In: *Music Perception: An Interdisciplinary Journal* 25.3 (2008), pp. 253–255.

[3] Gregory J Sandell. "A Library of Orchestral Instrument Spectra". In: *Proceedings of the International Computer Music Conference*. 1991, pp. 98–98.

[4] J.P. Bello et al. "A Tutorial on Onset Detection in Music Signals". In: *IEEE Transactions on Speech and Audio Processing* 13.5 (2005), pp. 1035–1047.

[5] Henrik von Coler and Alexander Lerch. "CMMSD: A Data Set for Note-Level Segmentation of Monophonic Music". In: *Proceedings of the AES 53rd International Conference on Semantic Audio*. London, England, 2014.

[6] Henrik von Coler, Moritz Götz, and Steffen Lepa. "Parametric Synthesis of Glissando Note Transitions - A user Study in a Real-Time Application". In: *Proc. of the 21st Int. Conference on Digital Audio Effects (DAFx-18)*. Aveiro, Portugal, 2018.

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

[8] Jürgen Meyer. "Musikalische Akustik". In: *Handbuch der Audiotechnik*. Ed. by Stefan Weinzierl. VDI-Buch. Springer Berlin Heidelberg, 2008, pp. 123–180.

[9] Han-Wen Nienhuys and Jan Nieuwenhuizen. "LilyPond, a system for automated music engraving". In: *Proceedings of the XIV Colloquium on Musical Informatics (XIV CIM 2003)*. Vol. 1. 2003, pp. 167–171.

[10] E. Gómez et al. "Melodic Characterization of Monophonic Recordings for Expressive Tempo Transformations". In: *Proceedings of the Stockholm Music and Acoustics Conference*. 2003.

[11] Norman H. Adams, Mark A. Bartsch, and Gregory H. Wakefield. "Note Segmentation and Quantization for Music Information Retrieval". In: *IEEE Transactions on Speech and Audio Processing* 14.1 (2006), pp. 131–141.

[12] Chris Cannam, Christian Landone, and Mark Sandler. "Sonic visualiser: An open source application for viewing, analysing, and annotating music audio files". In: *Proceedings of the 18th ACM international conference on Multimedia*. ACM. 2010, pp. 1467–1468.

[13] Xavier Rodet and Florent Jaillet. "Detection and Modeling of Fast Attack Transients". In: *Proceedings of the International Computer Music Conference*. 2001, pp. 30–33.