Scientific
Research
Publishing

The Impact of Selected Parameters of a Modified Sampling Synthesis on the Result of Its Auditory Assessment

R. J. Delekta[1], L. J. Spalek[2], M. Pluta[1]

[1]Academy of Music in Krakow, ul. Sw. Tomasza 43, 31-027 Krakow, Poland
[2]Institute of Physics, Cracow Institute of Technology, ul. Podchorazych 1, 30-084 Krakow, Poland
Email: nevermore@neostrada.pl

Abstract

One of the main shortcomings of standard sampling synthesis is the very limited number of sound parameters that are user-controllable. In the most general case, the user can choose a particular pitch, duration, and amplitude. If the sampler allows control over articulation, it simply switches from one sound sample to another. This makes fine-tuning of musical performances demanding and time-consuming if not an impossibility altogether. A synthesis system has been developed at the Academy of Music in Krakow, Poland. It uses a large collection of samples that contain short sequences of notes. The system implements a number of techniques to seamlessly connect recorded sequences, to control note durations as well as the tempo and the dynamics envelopes. Samples are automatically chosen, modified, and connected to keep the recorded, natural note transitions intact. The system uses performance rules to introduce variations into the regular playback akin to live performances by musicians. A user can either control the parameters manually or choose a desired expression and leave the particular decisions to the system. However, it is necessary to examine which parameters have the greatest impact on the listeners' impression and determine useful values. 15 expert listeners compared and evaluated variants of musical performances produced by the synthesis system with different sets of parameters. The paper discusses a selection of the examined parameters, the test methods employed and the results obtained.

1. Introduction

One of the aims of sound synthesis is the faithful recreation of sounds produced by acoustic instruments. Sampling is a popular method which yields satisfying results despite its simplicity. Its primary advantage rests upon a precise sound replication while requiring little computational complexity. High-quality samples allow a proper

reproduction of all the registers of a given instrument as the technical limitations associated with capacity and sample storage have been overcome. However, the few possibilities of influencing the sound being generated is a significant drawback. Using a sufficiently big sound sample set is the usual remedy. The set can encompass all of the pitches produced by the instrument (multisampling), in various performance techniques (dynamics and articulation), and in several interchangeable variants to avoid the impression of repeatability. Another difficult to eliminate drawback of sampling synthesis is the emulation of natural transitions between sounds [1]-[3].

The synthesis system for wind instruments in a symphonic orchestra being developed by the authors at the Academy of Music in Krakow [4] is a modified sampling synthesizer which utilizes specially prepared sound samples and signal processing methods. Analyzing and processing of a music score results in an acoustic signal being generated. From a music point of view, this non-real-time processing allows the modification of various performance aspects. This in turn, allows us to obtain the closest resemblance to that of performances of live musicians and not merely a verbatim reproduction of the music notation. The process takes place under defined performance rules, determining the variations in selected parameters dependent on musical context and style.

2. Current Implementation and the Resulting Aims

The current implementation (a Matlab/Octave code) consists of a score analysis module, specially recorded sound samples, and a method of connecting them. The joining is done through the use of a modified crossfade algorithm. The sound samples in the crossfade section have the same base frequency. As a result, phase problems may arise, hence a modification of the algorithm is necessary. A calculation of a cross-correlation with an associated micro time-shift aligns the phases and eliminates the problem [5]. A selection of performance rules has been chosen along with corresponding methods and algorithms needed to implement those rules [4]-[6].

The aim of this paper is to discuss the influence of a selection of parameters of a modified sampling synthesis via the test methods employed on the result of their auditory assessment. Specifically, the parameters used in the method for joining samples (crossfade length and the particular choice of connection position), the possibilities of altering note duration times, and shaping the amplitude are discussed.

3. Experiment Details

A listening test with the aim of comparing and evaluating variants of musical performances produced by our synthesis system with different sets of parameters was performed. The test involved a comparison of pairs of sound samples. A variant of the two-alternative forced choice method was utilized whereby the listener is presented with a pair of samples and listens to each one at least twice. Each pair was characterized by two different values of a selected parameter. Listeners were instructed to make a single selection based on their perception of the best resemblance to natural sound transitions. One should note that sound defects such as audible discontinuities, e.g. in pitch or amplitude, do not disqualify a sample.

The sound samples were based on two instrumental symphonic music fragments for the flute and bassoon by A. Dvořák and W. A. Mozart, respectively (please see **Figure 1** for details). The piece by Dvořák was chosen so as to determine the influence of parameter values on closely combined sounds in a melodious phrase. Parameter changes should be easily observable, due to connections of various sort, *i.e.* on different rhythmic values, and

Figure 1. The two selected instrumental symphonic music fragments for the flute and bassoon used in the listening tests. A piece from Symphony No. 5 in F major, Op. 76, III Movement-Trio, bars 285 - 292 by A. Dvořák (top) and Symphony No. 35 in D major, K. 385 I Movement, bars 1 - 5 by W. A. Mozart (bottom).

intervals. This piece is of greater scientific significance among the tested couple. The listeners need to subjectively judge the influence of a particular parameter on the piece as a whole. In the piece by Mozart the sounds are mostly separated. Thus, the effects of rare (in occurrence) and short changes in the parameters may be examined in the surroundings of sounds uninfluenced by the said changes. In summary, the piece by Dvořák contains a larger number of legato connections, and may be characterized as melodious. The piece by Mozart contains separate notes, staccato, with connections restricted to the end part, and may be characterized as distinct.

There were 42 different sound samples used in the tests (please see **Table 1** for details). No crossfade can be done in the areas of note beginning and endings. The crossfade is located as needed between them in the sample connecting procedure. Samples were processed with an 88,200 Hz sampling frequency but resampled to 44,100 Hz 1-channel (mono) for playback. Their duration was approximately 7 s (the Mozart piece) and 10 s (the Dvořák piece). Reference variant parameters were established on the basis of preliminary tests, as described in [5].

Due to a practical, fatigue-related time limit for a listening test, not all of parameter combinations could have been compared. In general, there was only one parameter changed at a time, while the others were kept at their reference values. Each listener rated 80 pairs of sound samples, *i.e.* 40 each by Mozart and Dvořák with corresponding parameters. Out of these there were 6 variants in the changes of tempo. Each appeared once in the test, due to an easily identifiable difference. The other pairs appeared twice with a change in the order of samples, *i.e.* AB, BA, so there were: 7 crossfade variants (14 pairs), 4 beginning length variants (8 pairs), 3 ending length variants (6 pairs), and 3 phrase arch variants (6 pairs). The order in which the pairs were presented was the same for all of the listeners. All of the parameters were grouped according to the particular parameter being tested.

The test group was composed of expert listeners: 10 experienced university-level ear training teachers, and 5 orchestra conductors (PhD students). With 15 listeners, the sample size equaled 15 for all the pairs in the tempo comparison part, and 30 for the other comparisons (two occurrences of each pair-with different order-treated as independent events). The test procedure was automatic, software controlled. Closed studio headphones (Beyerdynamic DT 770 Pro) were used for diotic signal presentation (1-channel/mono samples, presented simultaneously to each ear). The test lasted in the range of 50 - 120 minutes with an average duration of approximately 70 minutes. Listeners were allowed to take a break at any moment during the test.

Table 1. A characterization of the samples used in the tests.

Excerpt	No	T	B	E	X	D	A	
Dvořák	1	100	68	23	113	1	1	Reference variant
	2 - 4	60, 80, 117						Tempo variants
	5 - 11				6, 11, 28, 57, 85, 170, 283			Crossfade region length variants
	12 - 15		23, 45, 91, 113					Note beginning length variants
	16 - 18			11, 45, 68				Note ending length variants
	19 - 21					0, 0, 1	0, 1, 0	Phrase arch presence variants
Mozart	22	150	68	23	113	0	0	Reference variant
	23 - 25	125, 175, 200						Tempo variants
	26 - 32				6, 11, 28, 57, 85, 170, 283			Crossfade region length variants
	33 - 36		23, 45, 91, 113					Note beginning length variants
	37 - 39			11, 45, 68				Note ending length variants
	40 - 42					0, 1, 1	1, 0, 1	Phrase arch presence variants

Table key: T-Tempo [BPM] (beats per minute); B-Note beginning (part not used in crossfade) [ms]; E-Note end (part not used in crossfade) [ms]; X-Crossfade length [ms]; D-the presence of a dynamics arch; A-the presence of a tempo arch (agogics). Values not listed are the same as in the reference variant.

4. Results and Discussion

The results of different test pair variants are presented below with accompanying commentary. In order for the result to be deemed significant (with a level of significance = 0.05 of a two-sided exact binomial test), that is, for the listeners to prefer one variant over another in a pair, the difference must be at least 60 percentage points for a tempo variant comparison and at least 40 percentage points in the other cases.

4.1. A Comparison of Crossfade Lengths

Short length crossfade values (*i.e*., up to 28 ms) were rejected by the listeners in the Dvořák pieces. There was no clear preference in all the other cases (please see **Figure 2** for details).

4.2. A Comparison of Note Beginning Lengths

Listeners showed a clear preference towards note beginning lengths with the initial test value of 68 ms, but only in the case of the Dvořák piece. In the Mozart piece no significant preferences were observed (**Figure 3**).

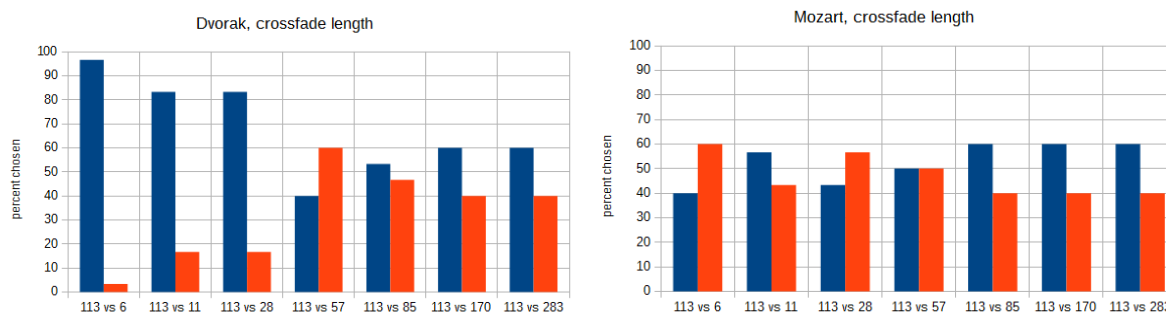### 4.3. A Comparison of Note Ending Lengths

A comparison of listener preferences with regards to note ending lengths for the Dvořák piece yielded the same results as for the case of note beginning lengths. When listening to the Mozart piece, a preference towards longer time values was apparent, as compared with the starting values, *i.e*. 68 ms vs. 23 ms, respectively (**Figure 4**).

### 4.4. A Comparison of Phrase Arch Variants

A comparison of the different variants of the phrase arch yields no clear listener preference in the Dvořák playback. The Mozart piece samples with an agogic arch (DA and A) were met with a listener rejection. The dynamical arch was somewhat more tolerated but it was not a clear preference (**Figure 5**).

4.5. A Comparison of Tempo Variants

In the legato piece (*i.e*., the Dvořák fragment) listeners rejected samples with large tempo values, *i.e* with 117

Figure 2. A comparison of different variants of crossfade lengths in milliseconds.
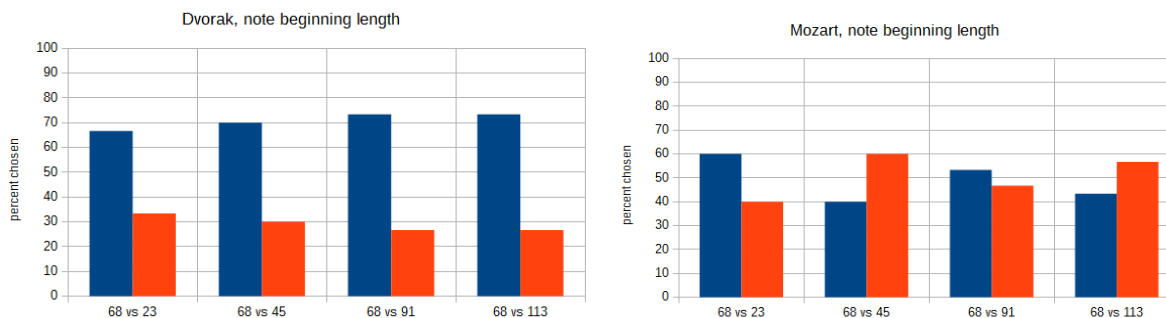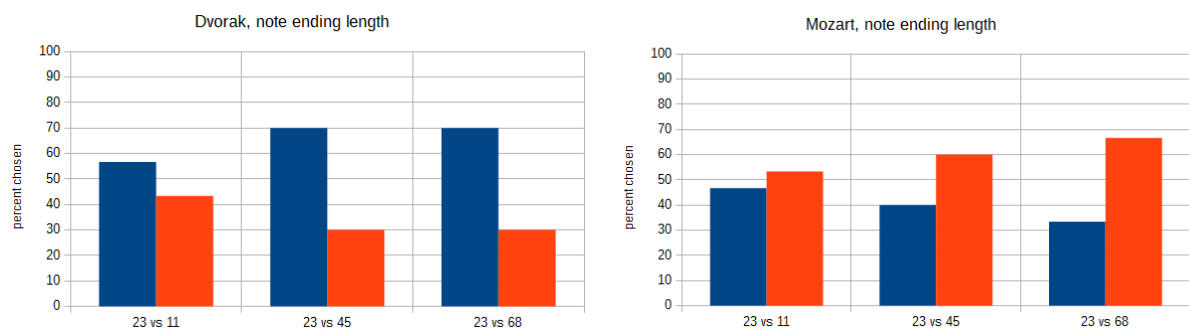

Figure 3. A comparison of different variants of note beginning lengths in milliseconds.

BPM. There was no significant preference towards moderate values when choosing between slow and moderate tempos (60 vs. 100 BPM). Listeners clearly rejected the highest tempo (200 BPM) in the Mozart piece (**Figure 6**).

4.6. Listeners' Feedback

Listeners noted that longer notes appeared idle, in contrast to how they are normally performed. This issue requires further consideration and the introduction of some dynamical change, depending on the context. As certain phrases transition from one instrument to another when they are performed, one needs to establish an appropriate rule and a suitable realization of an overlap between two instruments in the score. There have been a considerable number of remarks with regards to articulation and accentuation. However, these were unfortunately mutually contradictory. The release segments of notes require attention. In current implementation, a sample is shortened to a desired length with a fade out. The first issue to consider is the shape of the fade envelope. Secondly, and more importantly, the silence of the recording room should be audible in the track

Figure 4. A comparison of different variants of note ending lengths in milliseconds.
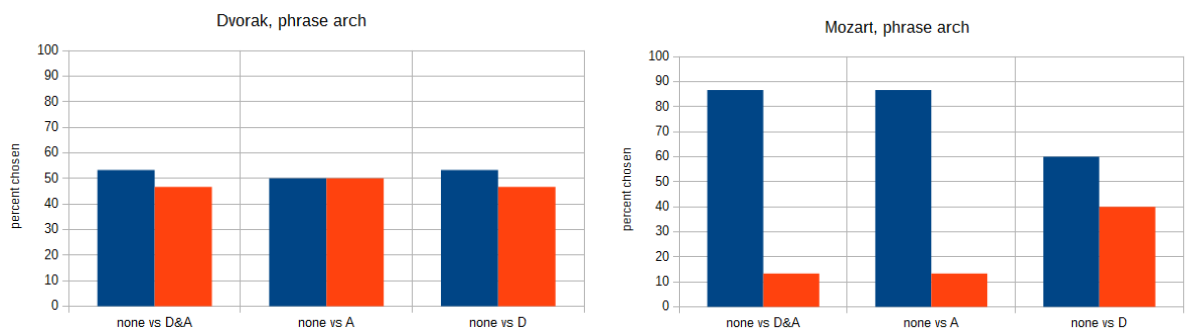
Figure 5. A comparison of the presence of different variants of phrase arches. D&A denotes a dynamic and agogic arch, A denotes an agogic arch, and D denotes a dynamic arch.
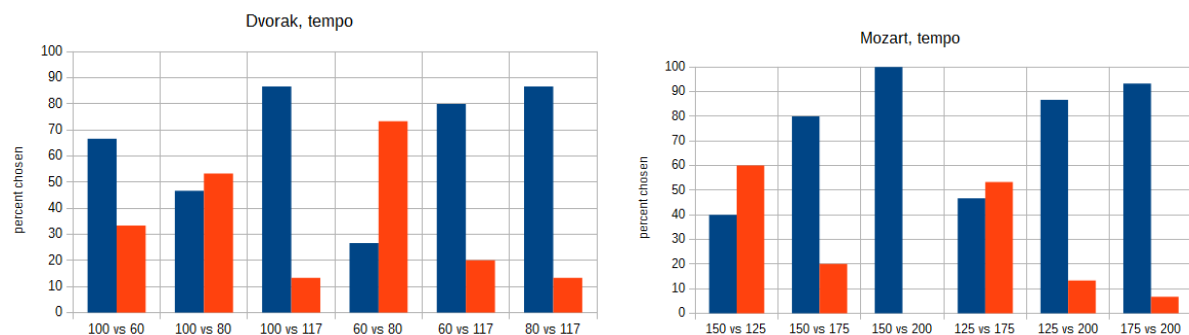
Figure 6. A comparison of different tempo variants in beats per minute (BPM).

225

background during the whole duration of the playback as complete silence during rests is perceived as unnatural. An appropriate remedy should also mask the unnatural fade out connected with the note ending.

5. Conclusions

The listeners clearly preferred the initially chosen crossfade length of 113 ms. Shorter time lengths were met with an especially negative reception. The discrepancies in the case of note ending lengths suggest that it is best to refrain from using averaged values. Instead, crossfade fragments should be determined for each sample separately in the process of preparing the samples. Attempts to use average values should only be made for note beginning lengths. Larger tempo changes are badly received by the listeners. This suggests that multi-note samples should contain as many tempos as possible. These would be only slightly altered in the process of sample joining. Most probably with a preference towards slowing them down rather than accelerating them. The automatic introduction of phrase arches should be abandoned. Especially with regards to the agogic arch which met with a bad listener reception. This function should be user controlled. Dynamical arches may be introduced with a larger degree of freedom. However, in this case the listeners did not show any clear preference between cases containing this arch as well as without it. In the Mozart piece, where close connections and multi-note samples seldom appear, particular choices of crossfade and note beginning lengths have little significance. The lack of preference with regards to the phrase arch in the cantabile melodic line (*i.e.*, in the Dvořák piece) is quite puzzling. Perhaps the influence of distortions introduced by the additional signal processing counteracts the potential improvement in the perception of naturalness.

Acknowledgements

References

[1] Strawn, J.M. (1985) Modeling Musical Transitions. CCRMA, Department of Music, Stanford University, Stanford, California.

[2] Almedia, A., Chow, R., Smith, J. and Wolfe, J. (2009) The Kinetics and Acoustics of Fingering and Note Transitions on the Flute. *Journal of the Acoustical Society of America*, **126**, 3. http://dx.doi.org/10.1121/1.3179674

[3] Roads, C. and Strawn, J. (1996) The Computer Music Tutorial. The MIT Press, Cambridge, Massachusetts.

[4] Delekta, R. and Pluta, M. (2014) Synthesis System for Wind Instruments Parts of the Symphony Orchestra. In: 7*th Forum Acusticum*, Kraków, 7-12 September 2014.

[5] Pluta, M. and Delekta, R.J. (2015) Technique to Seamlessly Connect Sound Samples in Sampling Synthesis. *Conference proceedings OSA*, Wrocław-Świeradów.

[6] Delekta, R.J. and Pluta, M. (2015) Implementacja Reguł Wykonawczych w Syntezie Dźwięku Instrumentów Dętych Zmodyfikowaną Metodą Samplingową. *Proceedings of the* 16*th International Symposium on Sound Engineering and Tonmeistering*, Warszawa, 8-10 October 2015, 1.