MECHANICS
Vol. 27 No. 4 2008

Marek PLUTA[*]

SIMULATOR OF INTONATION ERRORS: A TOOL FOR CONTROLLING PARAMETERS OF ACOUSTIC SIGNALS IN AUDITORY TESTS

SUMMARY

Intonation is one of the key elements of music performance and music perception, yet, due to a lack of suitable tools research on this subject was scarce. This paper presents a new and original tool developed particularly for a purpose of intonation perception research: the hardware-software simulator of intonation errors. The simulator is a controllable source of acoustic signals with properties similar to real musical signals. The paper discusses needs and possibilities of realistic simulation, as well as design and application of the simulator. Sample results of auditory tests that investigate impact of simulation's realism on intonation perception are presented.

Keywords: *simulation, control, acoustic signal, digital signal processing, sound synthesis, auditory tests*

SYMULATOR BŁĘDÓW INTONACYJNYCH JAKO NARZĘDZIE DO STEROWANIA PARAMETRAMI SYGNAŁÓW AKUSTYCZNYCH W TESTACH SŁUCHOWYCH

Intonacja jest jednym z kluczowych elementów wykonawstwa muzycznego oraz percepcji muzyki. Ze względu jednak na brak odpowiednich narzędzi, jej badaniom nie poświęcano dotychczas należytej uwagi. Niniejszy artykuł prezentuje nowe narzędzie zaprojektowane specjalnie w celu zastosowania w badaniach percepcji intonacji: sprzętowo-programowy symulator błędów intonacyjnych. Symulator jest kontrolowanym źródłem sygnałów akustycznych o właściwościach zbliżonych do rzeczywistych sygnałów muzycznych. Artykuł omawia potrzeby oraz możliwości realistycznej symulacji, a także projekt, wykonanie i zastosowanie symulatora. Prezentuje również przykładowe wyniki testów słuchowych, w których zbadano wpływ realizmu symulacji na percepcję intonacji.

Słowa kluczowe: *symulacja, sterowanie, sygnał akustyczny, cyfrowe przetwarzanie sygnałów, synteza dźwięku, testy słuchowe*

1. WHY SIMULATE WRONG INTONATION IN AN ORCHESTRA?

Music, particularly when it is performed by live musicians, is a source of highly complex acoustic signal. Not only is this complexity a result of diverse characteristics and variability of instruments' timbre, but it is also an effect of several levels of a higher order structure and form of a piece or composition. Therefore, it is frequently impossible to enclose such signal into an analytical form. Human music reception is partially based on the mechanism of auditory scene analysis (Bregman 1990), in which a sound signal is being separated into auditory streams. Those streams are perceptual units representing physical events or objects emitting sound. Through auditory scene analysis the listener acquires information regarding number of individual sound sources, their positions and characteristics, which creates the basis for auditory analysis of music.

Auditory tests employ simple, abstract signals, such as sinusoids, noises, amplitude or frequency modulated courses, and short impulses. Such signals differ from a far more complex real world sounds that sense of hearing is dealing with, for example in case of music. Therefore, tests based on simplified and abstract signals disregard the mechanism of auditory scene analysis, which relies more on recognition of patterns of sounds generated by physical events, and less on analysis of rudimentary parameters of signal components. That is why, in many cases, results of such tests may diverge from auditory system behaviour under natural conditions.

In order to test and study processes taking place in human auditory system while listening to music, it is vital to have a possibility to automatically generate acoustic signals in a form of realistic and repeatable fragments of music, and to be able to precisely set its chosen parameters. The need to manipulate signal parameters excludes employment of music recordings, as being poorly susceptible to interfere with parameters of their components, such as pitches of individual instruments in symphonic orchestra, due to irreversible process of sound mixing. Only tests applying simulations based on realistic sound synthesis, allowing generation and parametric description of controllable acoustic signals that possess near-natural characteristics, may deliver information regarding perception of sound under natural conditions, including effects of auditory scene analysis.

One particular aspect of sound perception, which is reception of intonation in a presence of other musical sounds, though highly important from the musical perspective, remains scarcely researched due to a lack of possibility to generate suitable test signals. Until now, tools for studying perception of intonation that could provide high sound realism, control over signal parameters, and repeatability, were non-existent. There are computer programs, used in ear training, such as GNU Solfege, Auralia, or Practica Musica, that offer some basic functionality, though their microtonal and intonation auditory exercises are overly simplified,

[*] Department of Mechanical Vibroacoustic, Faculty of Mechanical Engineering Robotics, AGH University of Science and Technology in Krakow

ignoring most musical context, and thus dealing with intonation in abstract situations. They present intonation errors as detuned music intervals, chords, or simple melodies, reproduced exclusively by a sound synthesiser. Moreover, as purely educational software, aforementioned programs do not provide necessary signal control. As a result, it is impossible to use them as a signal sources in a research on perception of intonation.

A new tool, hardware-software simulator of intonation errors, has been developed for a purpose of intonation perception research and was utilised in a series of auditory tests. The simulator, based on a high-fidelity sampling synthesis and digital recordings, provides a wide range of control over intonation of selected instruments or groups of instruments from the entire symphonic orchestra, leaving the rest of the orchestra instruments playing properly tuned. As elements of individual musical performance that can influence perception of intonation have been preserved, sound realism of fragments of orchestral pieces generated by the simulator is nearly natural.

2. SOUND PITCH AND INTONATION PHENOMENON

Pitch is a subjective auditory attribute of sound (Ozimek 2002), therefore it cannot be directly measured. However, it depends on physical parameters of sound: wave frequency, sound pressure, duration, envelope shape, and presence of other sounds. Humans perceive as having definite pitch sine waves, referred to as sine tones or pure tones, and harmonic complex tones, in which case pitch is defined by fundamental frequency, even if spectrum is highly degenerate. Some sounds with inharmonic spectra can also have definite pitch.

Musical scales use intervals representing frequency ratios to measure differences between pitches. Pitch perception in music is periodic, therefore most scales divide octave (frequency ratio of 2:1) into smaller intervals, to obtain pitch classes, or chromas, repeated in every octave. Exact division is defined by particular tuning system. Intervals sound best when their frequency ratio is exactly equal to a ratio present in the harmonic series, i.e. when the interval can be found between partials in the harmonic multitone. Such intervals are commonly referred to as ‘natural’. However, it is arithmetically impossible to divide octave into twelve pitch categories, so that all intervals would remain natural. That is why all twelve-tone tuning systems have to use some kind of compromise: either slightly differentiate intervals based on different chromas keeping some intervals natural while detuning others, or divide octave equally, resulting in all intervals being finely detuned. In the latter system, named equal temperament, all semitones represent the same frequency ratio of $\sqrt[12]{2}:1$. For the purpose of finer pitch measurements a semitone can be divided further into 100 cents. One cent represents frequency ratio of $\sqrt[1200]{2}:1$. Instruments such as violin can, in practice, use more than one tuning system in a single performance, depending on melodic or harmonic context.

2.1. Intonation phenomenon

A musical performer has some freedom interpreting tempo, dynamics (in music representing softness or loudness of a sound or note), articulation (representing manner of performance of a particular note), or even rhythm. However, in case of pitch a performer should keep precisely to a music score and chosen tuning system. The precision of pitch is limited by auditory system properties and technical means of sound control on given instrument. Nevertheless, considering the aforementioned limitations, performed pitch should be as precise as possible, otherwise listeners will feel discomfort that, in extreme situations, can prevent any music reception at all. Musical practice refers to the precision of pitch as the intonation.

2.2. Complexity of intonation error phenomenon

Methods of acoustic signal digital analysis, based on either Fourier or wavelet analysis, are not up to the task of identification and estimation of real intonation errors of individual instruments in recordings of music ensembles, especially the bigger ones, like symphonic orchestra. The problem core is a difficulty to separate complex, mixed signal into signals of individual instruments. Even if that kind of separation was possible, there is a mathematical problem to describe real detuning phenomenon. Musical instruments generate sounds of complex and rich spectral structure, in which some partials are harmonic. Balance of those partials is dynamic, and is also related to:

- duration of sound,
- musical articulation and dynamics,
- changes in acoustic environment, including change of instrument position.

Even so, simplifying, intonation error b can be treated as a shift of fundamental frequency f_0 of generated sound in relation to a proper frequency f_p:

$$b = \frac{f_0}{f_p} \tag{1}$$

the real situation is more complex:

- fundamental frequency can be absent and located only basing on existing partials:

$$f_0 = f_{k+1} - f_k \tag{2}$$

- sounds of other instruments have different pitches, so that point of reference for intonation is not the same harmonic series in a different instrument, but rather many different harmonic series which fundamental tones f_{0i} can produce complex chord,
- intonation error is dynamic, since spectrum of each musical sound changes over time independently from other sounds, so that combination and balance of partials is time-dependant

$$y(t) = \sum_i A_i(t) \sum_k b_{ik}(t) \cos(2k\pi f_{0i} t) \tag{3}$$

– pitches change over time along with melody of each instrument, and usually those changes are not synchronised between instruments, causing variations in a chord structure, and altering both, point of reference for intonation, as well as pitch of detuned instrument, providing additional dynamic parameter:

$$y(t) = \sum_i A_i(t) \sum_k b_{ik}(t) \cos\left(2k\pi f_{0i}(t)t\right) \tag{4}$$

Therefore, even though time-variance of f_{0i} is specified in a musical score and as such is deterministic, the changes in sounds of individual instruments depend on the instrument, performer, and acoustic conditions, in big part remaining indeterministic.

3. SOUND REALISM

Musical composition passes several stages on its way from a composer to a listener. In order to create a realistic simulation, those stages should be analysed.

1. The first is idea and conception stage, based on a composer's vision, musical imagination and inner hearing.
2. The second is musical score, in which the composition has undergone significant reduction, because of limitations of parametric musical notation. In a way, score is a parametrisation of a composer's idea and remains invariable form of a musical piece.
3. A performer begins with analysis of composition in order to synthesise his own conception. Based on a performer's knowledge, musical imagination and inner hearing, it can significantly diverge from a conception of a composer or other performers.
4. During a music performance a performer's conception influences a process of instrument play transforming the composition into a physical phenomenon of sound. Each performance will produce different sound form not only in case of different performers possessing different skills, but also in case of the same performer, yet slightly different acoustic or other physical conditions.
5. In the last stage physical parameters of the sound phenomenon are converted into an auditory sensation by a listener, depending on his hearing, musical knowledge, and psychophysical condition. The listener performs an auditory analysis and synthesises the final musical image.

Being the basis of all performances, the score, as a parametric form, does not contain complete information that is finally perceived by a listener, and only some of its parameters are defined strictly and absolutely, like pitch. Some, such as rhythm, are precisely noted, but permit some degree of deviation, and others, such as dynamics, articulation or changes of tempo are noted relatively and inaccurately, permitting free interpretation. Therefore, some elements present in the final auditory sensation are missing from a musical score.

1. Score does not contain performer's conception manifested in his interpretation of adjustable parameters, such as dynamics, articulation, tempo, breathing, and transitions between their constant values, that give music live and individual feeling.
2. Deviations of some parameters can be accidental, since no performer can play two identical notes. That is why there will be random fluctuations, including performer's mistakes, in every performance. All of this can be treated as a kind of "musical information noise", corresponding to thermal noise in physical systems or round-off errors in digital systems. Level of this noise will depend on experience and class of a performer, as well as on type of musical instrument played.
3. Acoustic properties of a particular instrument, room acoustics (Adamczyk *et al*. 2003), as well as positions of performer and listener, have also some impact on final auditory sensation. However, this influence is less time-variable and in case of listening to a recording can even be treated as a constant.

Since more elements take part in an auditory perception of music than are included in musical score, therefore a score alone is not a sufficient source of information for a realistic simulation. In order for a simulation to be fully realistic it should include:

– a high fidelity sound synthesiser for chosen instruments, that could reproduce different kinds of articulation,
– a set of rules describing transitions between notes in different situations,
– a set of rules for proper musical interpretation of dynamics, articulation, tempo or breathing,
– a method to insert random deviations with a distribution resembling live performance,
– effects to simulate proper room acoustics (Gołaś 2000).

Many of the aforementioned elements are in range of contemporary technologies, yet their implementation is difficult. Software samplers (Roads *et al*. 1996) can realistically simulate sounds of many, but not all musical instruments. Convolution methods can help to include room acoustics into a simulation. There are also some efforts to codify rules of transitions between notes (Strawn 1985). Still, one important problem remains unsolved: there is no method to simulate a process of live interpretation of a musical piece. That is why contemporary simulator must take slightly different approach.

4. STRUCTURE, IMPLEMENTATION AND OPERATION OF THE SIMULATOR

4.1. Simulation guidelines

A simulation of intonation errors is a process of preparing and reproducing music samples containing errors purposely introduced into intonation of selected instruments of symphonic orchestra. Music samples are based on fragments of recordings lasting 10–20 s, constituting a musical entirety, that are supplemented with a given intonation error, while the rest of the orchestra plays correctly tuned.

In ensemble performance errors occur in particular instruments rather than in a whole ensemble simultaneously. If error occurred in a group of instruments, it would most likely be different in every instrument of the group, and its correction would require individual adjustments in every detuned instrument. For an error in one instrument, the rest of the ensemble constitute point of reference for intonation, while detuning more instruments creates deformed chord structures and causes determining of point of reference complicated. The simulator can recreate both situations: error in only one instrument, or different errors in a group of three instruments.

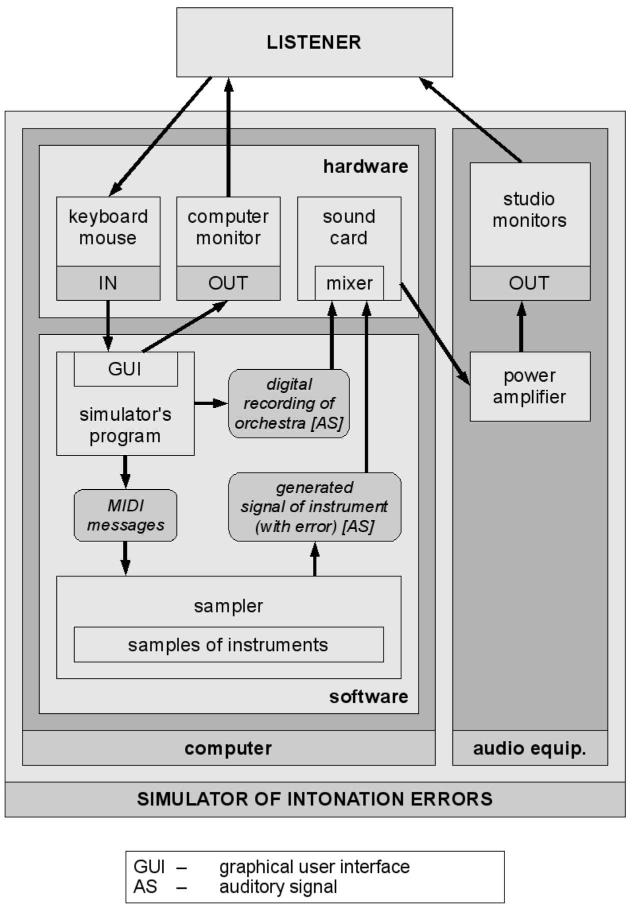

Fig. 1. Elements of the simulator

Even though all orchestra instruments are present in music samples, only wind instruments are being detuned. Most percussion instruments emit sounds of indefinite pitch, and in case of a few that have a definite pitch, usually performer has no way to control intonation. Therefore, intonation errors in percussion would be unrealistic. Exclusion of intonation errors in string instruments has several reasons. Almost every melody line in strings is performed by a group of instruments. Since their timbres are similar, unless an error is significant, a mean pitch is perceived. Even in case of significant errors, every performer's error could have different magnitude and direction, so that proper error estimation would be impossible. In a symphonic orchestra most cases of intonation errors originate in wind instruments. While playing a wind instrument it is difficult to compare its pitch to the rest of the orchestra, due to a sound being propagated not only through air, but also through performer's cranial and facial bones. Those instruments play individual melody lines, so there is no effect of blurring, errors are distinct, and can be precisely estimated.

In order to precisely set or estimate intonation errors, their magnitudes have been quantised to the following values: –50, –25, 0, 25, and 50 cents. Negative values mean lower than proper pitch, and positive values – higher. Quantisation is based on equal temperament properties and differences between tuning systems. Finer quantisation would be too close to possible pitch ambiguities resulting from coexistence of different tuning systems in a single performance. The simulator can generate not only constant, but also time-variable intonation errors, defined by error values in specified moments in time and one of predefined shapes of transition between those moments.

4.2. Structure and implementation of the simulator

In order to control intonation of individual instruments in a symphonic orchestra, the simulator combines fragments of digital recordings with one or three tracks generated by the most realistic method of sound synthesis: sampling synthesis in a form of software sampler. Synthetic tracks can be easily controlled, while digital recording of a real symphonic orchestra provides elements of live performance.

The simulator consists of hardware and original software elements, as shown in Figure 1. It is applied in two test stands: in Group of Ear Training laboratories in Academy of Music in Krakow and in Laboratory of Sound Engineering in Department of Mechanics and Vibroacoustics, AGH University of Science and Technology. Its main hardware element is a PC equipped with a sound card connected to an audio amplifier and speaker columns or to an active studio monitors. Hardware equipment is described in Table 1.

A dedicated program controls the process of simulation, i.e. generation and reproduction of music samples, as well as interacts with user through a graphical interface (GUI).

Table 1. Simulator hardware

Test stand	Academy of Music in Krakow	AGH University of Science and Technology
Computer hardware	Based on Intel Pentium 4 processor	Based on AMD Athlon processor
Sound card	E-mu APS	Creative Sound Blaster Audigy 2 Platinum
Audio hardware	Power amplifier: Technics SU-V500, passive monitors: Tannoy PBM 6.5 II	Active studio monitors: Genelec 1032A

It communicates with software sampler Gigasampler by means of MIDI (Musical Instrument Digital Interface) protocol. Gigasampler utilises a set of sound samples stored on computer hard drive containing recordings of wind instruments. During initial tests some of the samples turned out to be detuned and needed correction of intonation. In case of many sound samples frequency evaluation algorithm in a program Cool Edit failed, therefore a more robust had to be developed (Pluta 2006). Only application of this new method allowed correction of intonation in all sound samples.

4.3. Structure of music samples

Music samples are created by synchronous reproduction of two layers: digital recording of a real orchestra, and a synthetic layer containing detuned instruments. Digital recording layer is reproduced directly, without modifications, by the simulator and includes information constituting the sensation of sound reality that could not be simulated from the ground using sound synthesis. This layer is also a point of reference for intonation. While reproducing digital recording the simulator controls a software sampler using MIDI messages sent in specified moments. The sampler generates the synthetic layer, i.e. tracks of one or three instruments containing given intonation errors (Fig. 2). Timing, timbre, articulation, and dynamics of the synthetic layer have to be adjusted to match the recording.

For every fragment of digital recording the simulator stores a matching set of data, referred to as MIDI track, defining notes generated by the sampler as a synthetic layer. Every note has four parameters: start time t_i, timbre b_i, MIDI *velocity* a_i, and pitch w_i. Timbre is a number of sound sample and defines both, instrument and articulation. MIDI *velocity* controls amplitude envelope (ADSR – attack, decay, sustain, release) and represents musical dynamics. Even though parameters of MIDI track are defined and stored, intonation errors are generated basing on parameters provided in real time. That is how the simulator permits intonation control.

4.4. Synchronisation of layers

Preserving natural tempo fluctuations present in the digital recording is vital for achieving realistic simulation. Timing of the synthetic layer has to match timing of music in the recording. It is not enough to calculate appropriate lengths of notes, since both layers are reproduced by independent devices (audio and MIDI device) and can be easily desynchronised. The simulator has to synchronise start of every note, so it uses position pointer in reproduced waveform as a clock for the synthesiser (Fig. 3). This method requires a specific format of MIDI tracks: waveform sample numbers indicating note beginnings in digital recording are used in a place of note durations. During playback the simulator sends a *note on* message to the sampler if the waveform position pointer has surpassed a value for a particular note. This method allows MIDI tracks to follow natural rhythm and tempo nuances present in digital recording.

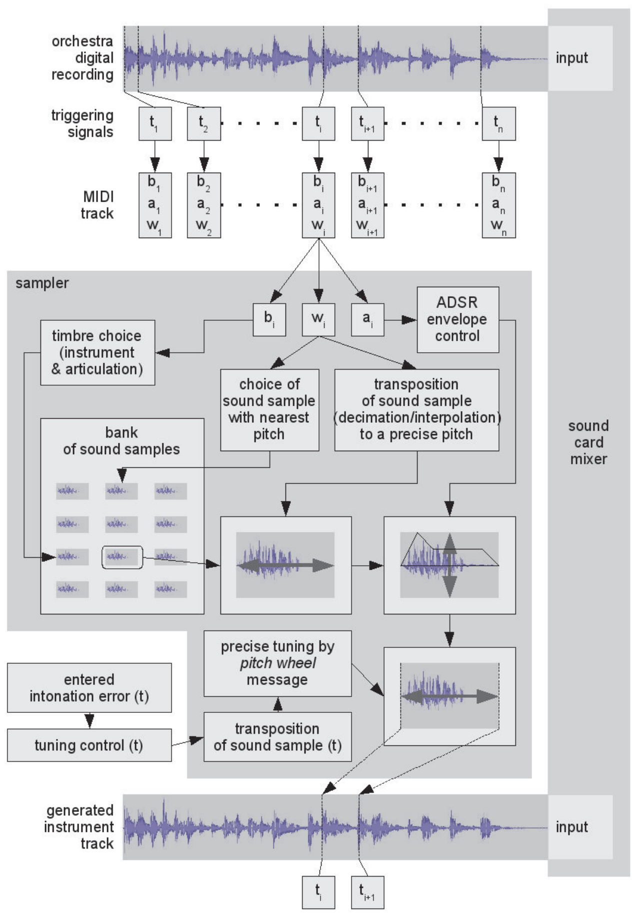

Fig. 2. Operation diagram of the simulator

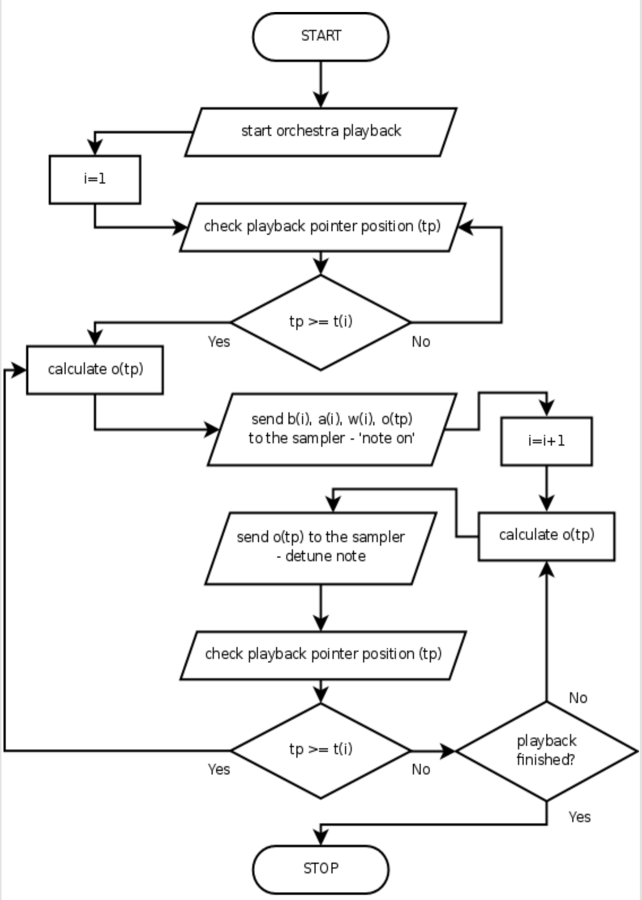

Fig. 3. Synchronisation of layers

4.5. Intonation control

A MIDI message *pitch wheel* that allows pitch shift in both directions is used to control intonation of instruments in the synthetic layer. It has theoretical resolution of 12 bits per semitone, but in practice depends on synthesiser implementation. *Pitch wheel* message can be sent not only during note start, but also while note is sounding. In order to use those properties for smooth and continuous change of intonation during music sample playback, the simulator sends appropriate messages every 5 ms (Fig. 2).

Intonation changes during playback are defined by a function which parameters are set when a decision about an error is made. Pitch shift values are calculated in real time, as a function of waveform position pointer. Depending on chosen function and its parameters, the simulator can create music samples with constant or time-variable intonation errors. Tuning of each instrument is controlled independently, so in case of detuning a group of three instruments each instrument can have a different error. The function that controls detuning can either be constant, or gradually change between specified moments in time ('nodal points': the beginning, the end, and optionally 2 more between them). Intonation shift in constant function, as well as in nodal points of time-dependent function takes one of five values from a range of –50 to +50 cents. Between nodal points intonation changes in one of four possible ways (Fig. 4). If b_p and b_k from a set of numbers {–2, –1, 0, 1, 2}corresponding to pitch shifts of {–50, –25, 0, 25, 50}cents are intonation errors of the beginning t_p and the end t_k of section, the intonation error in the moment of time t between them is described by one of the following formulas:

$$b_1(t) = b_p + (b_k - b_p)\left(\frac{t - t_p}{t_k - t_p}\right) \tag{5}$$

$$b_2(t) = b_p + (b_k - b_p)\sin\left(\pi\frac{t - t_p}{t_k - t_p}\right) \tag{6}$$

$$b_3(t) = b_p + (b_k - b_p)\left(1 + \mathrm{tg}h\left(s\frac{t - t_k}{t_k - t_p}\right)\right) \tag{7}$$

$$b_4(t) = b_p + (b_k - b_p)\left(\mathrm{tg}h\left(s\frac{t - t_p}{t_k - t_p}\right)\right) \tag{8}$$

Parameter s in two last formulas determines slope of the curve in its beginning or end. Curve b_3 simulates an 'end of breath' situation, when constant velocity of air stream cannot be sustained. Curve b_4 simulates performer's attempt to correct intonation just after starting to play. The attempt can be unsuccessful, as in Figure 4. The effect of 'pitch searching' is simulated by curve b_2. In case of 4 nodal points only curve b_1 is used, and function is represented by a broken line that can simulate freely adjustable intonation error.

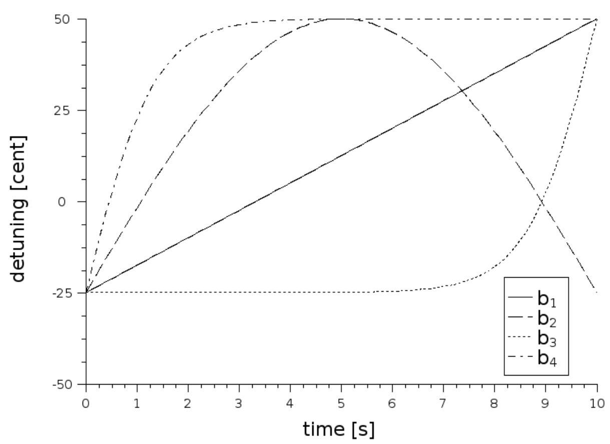

Fig. 4. Detuning curves for a sample set of parameters: $t_p = 0$, $t_k = 10$, $b_p = -1$, $b_k = 2$

4.6. Simulation of articulation and dynamics

Software sampler allows using sound samples that include many different kinds of articulation to introduce articulation changes into synthetic layer of simulation in order to match it with articulation in recording layer. In real musical performance articulation changes frequently. It is quite often that those changes are too subtle to be precisely recreated by a limited set of sound samples that contain 3 to 9 kinds of articulation, depending on instrument. However, approximate match can be accomplished. Such a process has two stages. First, articulation is initially set according to score of the musical piece. In the next stage both layers are listened to multiple times and articulation of synthetic notes distinguishing from the recording is corrected.

Musical dynamics of synthetic tracks has been matched with the course of dynamics in the recording using original method (Fig. 5) based on the fact, that unlike difficult to parametrise articulation, musical dynamics can be parametrised with RMS of signal x consisting of n samples as:

$$x_{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} \tag{9}$$

Measured for appropriate periods of time, this parameter can help reproduce musical dynamics course in recording. Ratio of signal RMS of a real instrument to signal RMS of a real orchestra constituted the norm that had to be obtained for every note separately.

The simulator controls dynamics of synthetic notes using 7-bit MIDI controller *velocity* that reflects velocity of pressing a key on the instrument, and changes amplitude envelope (ADSR) of sound sample accordingly. Even though it is not defined how exactly does it affect RMS of synthesised sound, it could be measured if the sound was recorded. First stage of the musical dynamics match is reproduction and simultaneous recording of the real orchestra's track. It allows to take into account parameters of sound card mixer. Recorded signal is then split into sections containing separate notes. In the beginning all notes have their *velocity* parameter set to middle of range (64), and their *D* parameter representing change of *velocity* set to half of this value (32).

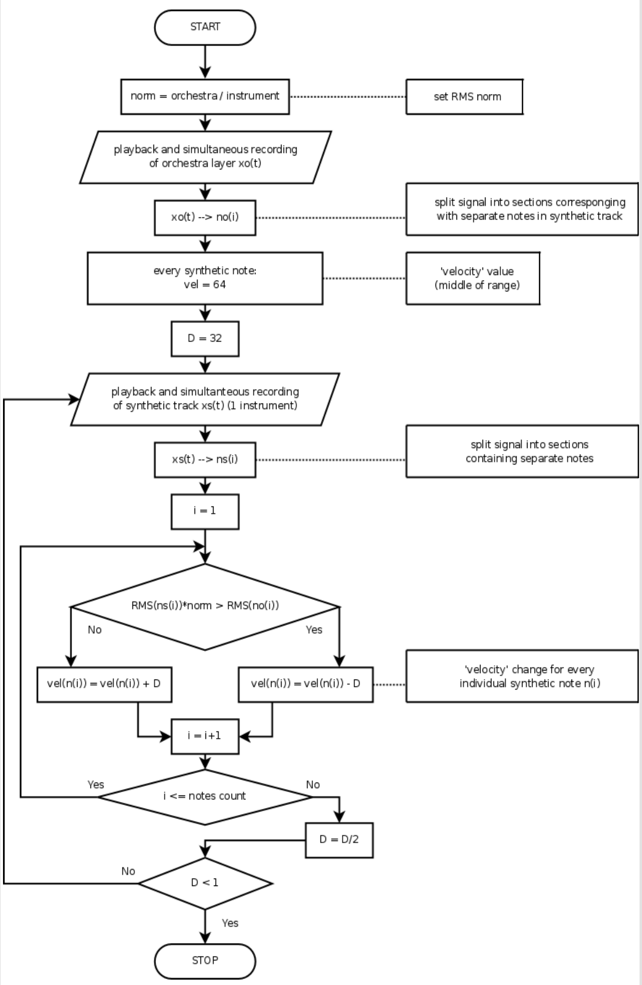

Fig. 5. Dynamics match algorithm

The dynamics match process runs in a loop that starts by reproducing the synthetic track with current *velocity* values. Just as in case of signal of the orchestra, synthetic track is recorded during playback, and then split into sections containing separate notes. Through recording both simulation layers, RMS is measured in signals generated under the same conditions, including parameters of sound card mixer that could influence each layer differently. RMS of every synthetic note is calculated, scaled accordingly to the established norm, and compared to RMS of corresponding orchestra's note. Depending on comparison result *velocity* parameter for the particular note is increased or decreased by *D*. After this *D* is halved. The loop continues until *D* reaches zero and can be halved no more.

Thanks to described method the musical dynamics course in synthetic tracks imitates natural changes inherent in recording of the orchestra, thus realising one more element of real performers' musical interpretation. During initial tests a group of conducting students noticed an audible improvement in sound realism after application of described method in comparison to an older method, similar to the one used for articulation.

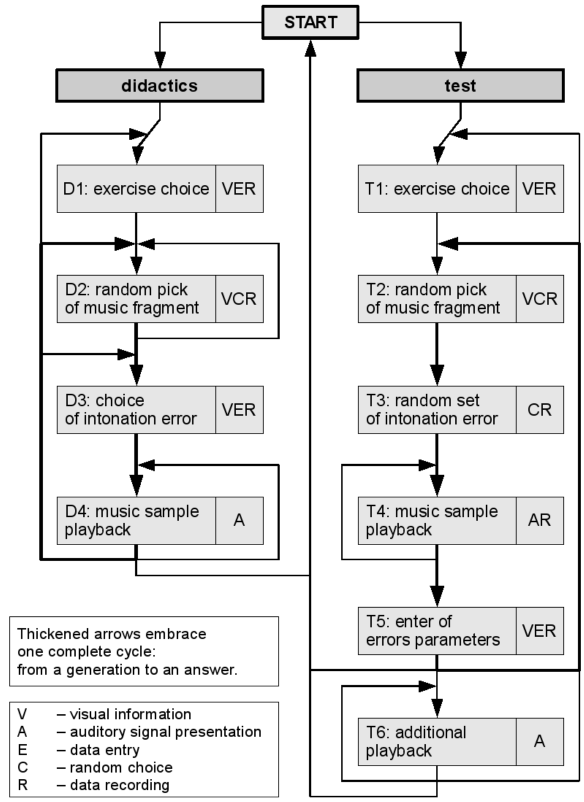

Fig. 6. Stages of the simulator operation

4.7. Operation of the simulator

There are two modes of the simulator's operation (Fig. 6). In the 'didactic' mode the user sets error parameters that are introduced into the music sample and reproduced. In the 'test' mode it is the simulator that automatically generates music sample by introducing random intonation error into a randomly chosen fragment of orchestra recording. In this mode the user (or the listener) listens to the music sample and enters values of error's parameters basing solely on auditory analysis. The simulator evaluates correctness of this analysis.

During operation the simulator automatically records the following user information for every music sample:

- date and starting time of the first reproduction,
- designation of the recording fragment,
- kind of auditory exercise,
- duration of generated music sample,
- total time of the auditory analysis (until parameters are entered),
- number and durations of the particular music sample reproductions,
- generated intonation error parameters,
- error parameters entered by the user as a result of the auditory analysis.

Twelve kinds of auditory exercises can be generated – each one corresponding with a specified category of intonation errors. Categories are divided according to a number of simultaneously detuned instruments, time-variability of errors, and required precision of auditory analysis, as shown in Table 2. Sample program window (for exercise 9) is shown in Figure 7.

Table 2. List of simulator's auditory exercises

		Error in 1 instrument	Error in 3 instruments
Time-invariant error	Detection of error	Exercise 1	Exercise 4
	Detection of error and its direction	Exercise 2	Exercise 5
	Detection of error's magnitude and direction	Exercise 3	Exercise 6
Time-variant error	2 nodal points, curve b_1	Exercise 7	Exercise 10
	2 nodal points, all curves	Exercise 8	Exercise 11
	4 nodal points, curve b_1	Exercise 9	Exercise 12

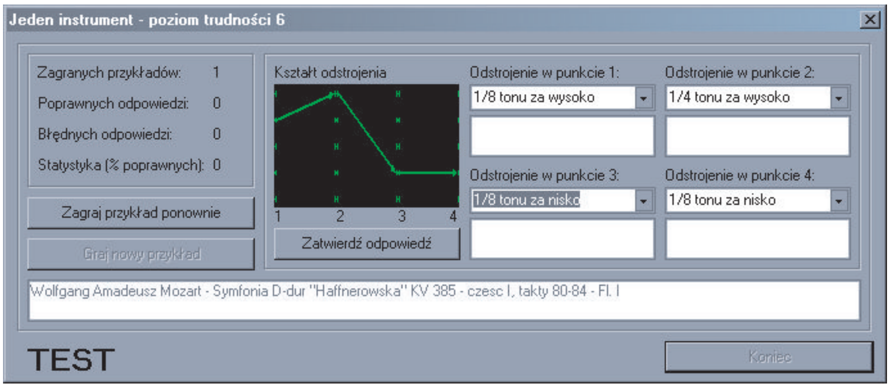

Fig. 7. GUI of simulator – exercise 9

5. USE OF THE SIMULATOR IN AUDITORY TESTS

Auditory tests of an intonation perception process were a target application of the simulator. Without a tool to generate musically realistic, repeatable and controllable intonation errors a research on such a process was impossible. The simulator allowed to test the process under auditory conditions similar to reception of a real music, so that perception mechanisms such as auditory scene analysis would work naturally. Detailed test results will be presented in work (Pluta 2008).

Auditory tests were carried out in two test stands: in Group of Ear Training laboratories in Academy of Music in Krakow and in Laboratory of Sound Engineering in Department of Mechanics and Vibroacoustics, AGH University of Science and Technology. Test equipment consisted of computer and audio hardware described in Table 1 and original, dedicated simulator software. 16 test subjects, between 20 and 29 years old, participated in a research, including:

– 6 conducting students and conductors,
– 3 music theory students,
– 5 sound engineering students,
– 2 subjects having primary or secondary education in music.

Participants' task was to listen to a randomly generated music samples, carry out its auditory analysis in order to determine parameters of listened intonation error, and enter those parameters into the simulator's program, which evaluated correctness of the auditory analysis. Tests were divided into sessions. Listeners participated in 10 individual sessions lasting 45–60 minutes each, carried out every day or every few days. Exact duration of each session was chosen by participants according to their capabilities. Test subjects operated the simulator individually and could adjust difficulty according to their own level making decisions regarding a kind of exercises carried out, a number of cycles (Fig. 6) in every exercise, time spent on particular exercises, and total session time. Participants were obliged to carry out a certain 'minimal program' consisting of 10 cycles in each one of 12 exercises.

5.1. Impact of the realism of simulation on intonation perception process

Primary tests of the intonation perception process have been carried out using the most realistic version of the si-mulator. However, ear training tools are based on very simple methods of sound synthesis. Similarly, most psychoacoustics research draws conclusions regarding complex perceptual phenomenons basing on test signals which structure is significantly simplified comparing to real-world or musical sounds. It is possible that the intonation perception process, since it is based on recognition of complex sound structures. could be affected by the realism of the test signal.

In order to test whether such an impact occurs, a series of tests was carried out. Participants used two versions of simulator's software, while hardware remained unchanged:

– the original 'high-fidelity' version utilised software sampler Gigasampler and not looped samples of wind instruments that included different kinds of articulation,
– 'low-fidelity' version represented inferior sound realism, reduced by utilisation of sound card hardware sampler and *General MIDI* (GM) compliant set of instruments containing 8 MB of looped sound samples.

The second version, as GM compliant (GM defines a list of 128 timbres to use), did not contain samples of different kinds of articulation, thus decreasing sound realism.

Additionally, looping of samples removed natural amplitude envelope of musical instruments, leaving only artificial ADSR envelope. On one hand, such a modification caused synthetic instruments to be easier to distinguish when played simultaneously with a recording of a real orchestra, on the other hand, differences between synthetic instruments were reduced. Both simulator's versions were switched during tests every few cycles. Each version was used in the same number of cycles.

Time needed to correctly identify intonation error, measured from the start of a music sample presentation to the end of auditory analysis indicated by confirmation of answer, was analysed. A distribution of identification times for all listeners contains four local maxima, therefore in the presented graphs answers are grouped into four time-intervals.

For both versions of the simulator the following were tested:

– distribution of correctly identified detuning cases in time-intervals (Figs 8 and 9),
– efficiency of detuning identification (as percentage of correct identifications) in each time-interval (Figs 10 and 11),
– profile of detuning identification mistakes (as percentage of a particular type of mistake in all cases of in-correct identifications) in each time-interval (Figs 12 and 13).

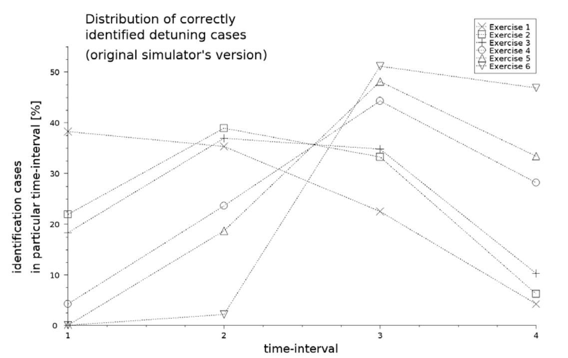

Fig. 8. A distribution of correctly identified detuning cases for the original version of the simulator

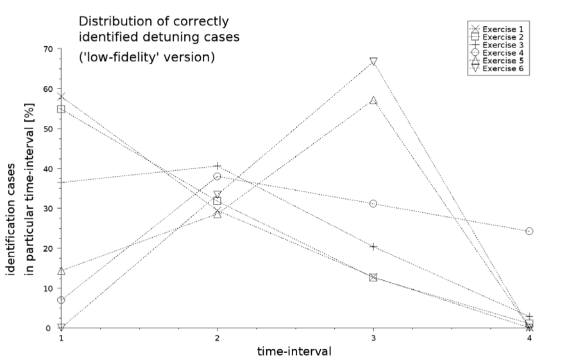

Fig. 9. A distribution of correctly identified detuning cases for the simplified version of the simulator

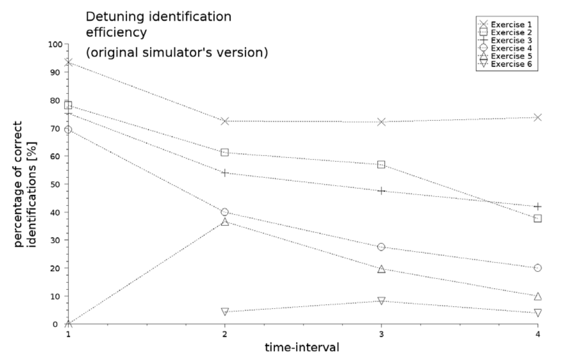

Fig. 10. Detuning identification efficiency for the original version of the simulator

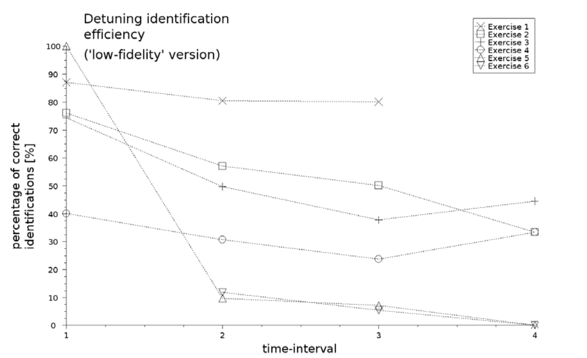

Fig. 11. Detuning identification efficiency for the simplified version of the simulator

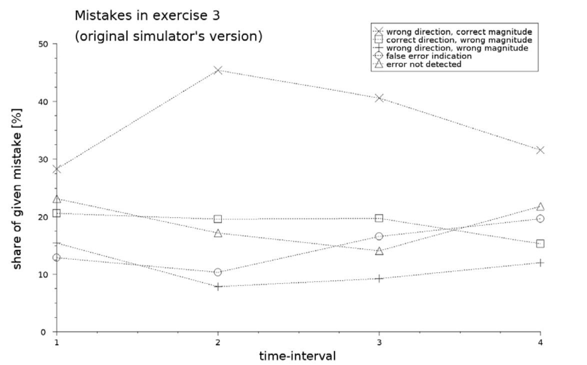

Fig. 12. A profile of identification mistakes for the original version of the simulator

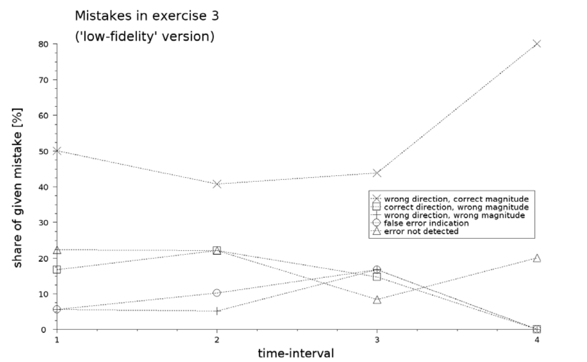

Fig. 13. A profile of identification mistakes for the simplified version of the simulator

Distributions of correctly identified detuning cases differ between both versions of the simulator. If the realism is reduced (Fig. 9) the last time-interval contains only few cases of identification, while the first one contains significantly more cases than in the original version of the simulator (Fig. 8). A difference in efficiency between both versions is interesting. In exercises 1–3, where only one instrument was detuned, impact of sound realism is negligible. However, in exercises 4–6, where three instruments were simultaneously and independently detuned, decreased sound realism (Figs 10 and 11) caused a decrease in detuning identification efficiency, most notably in exercise 5. It can be explained by difference in quality of the samples and removal of natural amplitude envelope from low-fidelity instrument samples. Since such samples are easier to distinguish from an orchestral recording, but more difficult to distinguish from other samples, it causes problems with a separation of auditory streams and effects in decreased efficiency of detuning identification.

A difference between both versions can also be observed in the profile of detuning identification mistakes, and involves the most common case of detuning direction mistake while detuning magnitude is identified correctly. In case of the original high-fidelity version (Fig. 12) this kind of mistake was only slightly more frequent then other kinds in first time-interval, but much more frequent in second and third time-intervals. It could indicate a different mechanism used by listeners when the identification was fast (below 8 seconds) compared to normal, slower identification mechanism. In case of the reduced reality version (Fig. 13) the situation has changed: the same mistake was much more frequent then others not only in second and third (respectively, 41% and 44% of all mistakes) but also in first time-interval (50%). Decreased realism could have somehow handicapped the mechanism used in realistic version.

6. CONCLUSIONS

Intonation is one of the key elements of music performance and music perception. However, due to a lack of suitable tools research on this subject was scarce. This paper presented a new and original tool developed particularly for a purpose of intonation perception research: the hardware-software simulator of intonation errors. The simulator serves as a controllable source of acoustic test signals which properties are similar to real musical signals: fragments of music pieces performed by the symphonic orchestra. Additionally, the simulator aids a data collecting process in auditory tests through automatic recording of test subjects' detailed results.

Results of auditory tests with two versions of the simulator: original 'high-fidelity' version, and less realistic 'low-fidelity' version indicated, that a change in level of realism of the simulation has an impact on intonation perception process, and decrease in sound realism handicaps particular mechanisms of intonation perception. Therefore not only further research on intonation perception, but also intonation-related music teaching should be based on tools providing realistic signals.

References

Adamczyk J., Stryczniewicz L., Łopacz H. 2003, *Parametry akustyczne sali operowej w Krakowie*. Prace Naukowe Instytutu Telekomunikacji i Akustyki Politechniki Wrocławskiej, Seria: Konferencje, 2003, vol. 84, No. 28, pp. 59–62.

Bregman A.S. 1990, *Auditory Scene Analysis: The Perceptual Organisation of Sound*. MIT Press, Cambridge Massachusetts.

Gołaś A. (Ed.) 2000, *Podstawy sterowania dźwiękiem w pomieszczeniach*. Uczelniane Wydawnictwa Naukowo-Dydaktyczne AGH, Kraków.

Ozimek E. 2002, *Dźwięk i jego percepcja – aspekty fizyczne i psychoakustyczne*. Wydawnictwo Naukowe PWN, Warszawa-Poznań.

Pluta M. 2006, *Musical instruments samples' tuning evaluation algorithm*. Proc. of the LIII Otw. Sem. Akust. OSA, Kraków-Zakopane, Poland, September 2006, pp. 199–204.

Pluta M. 2008, *Badanie interakcji słuchacza ze sprzętowo-programowym systemem symulowania błędów intonatyjnych*. Ph. D. diss. (in preparation), AGH University of Science and Technology, Department of Mechanics and Vibroacoustics, Kraków.

Roads C., Strawn J., Abbot C., Gordon J., Greenspun P. 1996, *The Computer Music Tutorial*. MIT Press, Cambridge Massachusetts.

Strawn J. 1985, *Modelling musical transitions*. Ph. D. diss., Stanford University Department of Music, Stanford.