

Analysis of the sound attack in context of computer evaluation of the singing voice quality

Edward Półrolniczak

West Pomeranian University of Technology
Faculty of Computer Science and Information Technology
ul. Żołnierska 52, 71-210 Szczecin, Poland
Email: epolrolniczak@wi.zut.edu.pl

Michał Kramarczyk

West Pomeranian University of Technology
Faculty of Computer Science and Information Technology
ul. Żołnierska 52, 71-210 Szczecin, Poland
Email: mkramarczyk@wi.zut.edu.pl

Abstract—Sound attack is meant as the initiation of the tone. Good attack involves the simultaneous start of exhalation along with the emission of sound. This situation is known as a soft attack. Much more common is a hard type of attack, which is normally used while speaking. It involves tightening of the vocal cords before the expiration. Unfortunately, this state is tiring for the ligaments and can cause damage to the voice. The idea was to propose a method to analyse this problem, which in the future would become a part of set of methods for computer analysis of the singing voice. This article presents a method of the extraction of attack parameters. The estimation of these parameters is carried out with the recorded audio samples. In the study sound sample of 'a' registered several times for each pitch will be used.

Index Terms—singing voice, sound attack, ADSR, quality of singing

I. INTRODUCTION

The motivation for taking up the research on the attack in the singing is the need of assessment of singing quality. The problem is to find some criteria of evaluation of the sound attack while singing. The goal is to find the criteria of automatic evaluation and involve computer methods to evaluate the attack in singing using these criteria. If properly defined, they allow for self-correction of selected voice parameters. It may be useful for training lessons of voice production. It can be useful to help singers make a progress. It can be very important for the choirs constantly working on the voice. Some criteria of evaluation, if properly defined, may allow for self-correction of selected voice parameters. The considered methods should evaluate these features in a similar way as human experts. To achieve that it is necessary to propose a computer method to analyse this feature and its parameters.

It should be noted here that it is difficult to find scientific publications which present the approach to computer evaluation of singing carried out in similar way as described in this article. The approach presented here reflects expert's evaluation in the computer aided assessment.

The singing voice is produced by the vocal instrument consisting of three basic components: the respiratory apparatus, the oscillating vocal folds and the vocal tract. Breathing has decisive influence on all activities related to the voice emission. Vocal folds open and close during breathing, but during speaking and singing, they start to oscillate with the

fundamental frequency of sound which comes out of the mouth. At the interface between these two phases (open and close) of the sound generation the musical phenomenon known as an attack on the sound occurs.

The ADSR envelope is a practical model that describes a single sound (note). It can be used for sound analysis [1], [2] and synthesis [3]. It is the essential description of the sound waveform in the MIDI standard [4]. In terms of sound synthesis attack is interpreted as a part of ADSR envelope. The attack is used for modification of a first phase of amplitude envelope [5] of generated sound in which sound gains the highest amplitude. In most basic models it's followed by decay section, during which amplitude is decreased. After that there is a stage of a sustain which is characterized by a stable amplitude. The ADSR envelope is finished then by a stage of release. During this part amplitude is completely decreasing. The ADSR sections are illustrated in fig. 1.

The word "attack" is probably too strong description of the release of air through glottal folds. Anyway it is the term commonly used in the literature. Some other, related terms used in literature are: initiation, onset [6] or transient. To be precise it is needed to define some terms related to each other: onset, attack, transient. The reason for making these distinctions clear is that different applications have different needs. The similarities and differences between these key concepts should be considered. Referring to the definitions found at [7] the attack is the time interval during which the amplitude envelope of the sound increases.

The transients are short intervals during which the signal evolves quickly (increases and decreases quickly) in some nontrivial, unpredictable way. It is often connected to the situation where the excitation of the sound is applied and then damped (leaving only a slow decay). So the transient would be a part of the amplitude envelope starting from the onset, including the whole attack and the most important part of decay (the strongest one). The onset [3], [8] is a single instant chosen to mark a transient. In most cases it is connected with the start of transient (also start of attack) or with the earliest time when the transient can be detected.

The sound attack in the singing is related to the breathing. During vocal development care should be taken to teach the attack on the breath [9]. Proper breath preceding the phonation

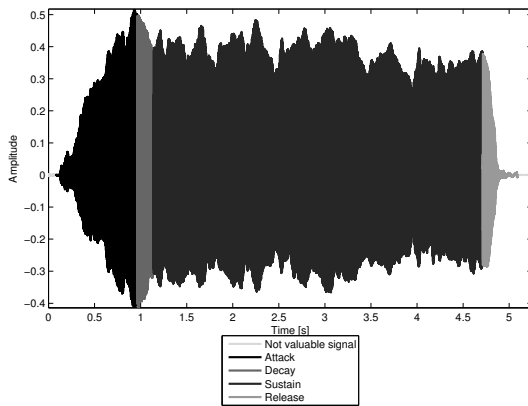


Figure 1. Example of the typical split of a sung phrase divided into sections ADSR (Attack-Decay-Sustain-Release)

phase will result in a good beginning of each phrase. It is especially important in phrases which begin with a vowel sound. Attack is more precise in the case of professional singers. Choral singers are less precise in that stage of voice production. It has to be noticed that practice can clear up that problem. Choir members are usually developing their voice in groups. Thus the vocal abilities are similar in a group of choir singers. It should be recalled here that the presented investigation concerns choir voices.

There are three known kinds of attack on sound depending on moment of tightening vocal cords in relation to the beginning of exhalation and on how strong the tightening of vocal cords is in the early stages of sound production:

- soft - when the vocal cords tightening moment coincides with the beginning of exhalation. This is the most favourable attack.
- hard - tightening of the vocal cords begins before the expiration. Unfortunately, this state is tiring for the ligaments and can cause damage to the voice.
- exhaling (aspirated) - short exhaust ahead of tightening of the vocal cords. Vocal cords do not close completely and the remaining gap influences the sound.

In this paper attack part of a signal is defined as first part of the sung sound. It ends by either stabilisation of amplitude (sustain) or total decrease of amplitude (decay).

In the article [10] an attempt to identify fast attacks has been presented. Fast attack transients are named there simply as attacks. The authors have defined the attacks as zones of short duration (a few ms) and fast variation of the signal short time spectrum with an abrupt increase in energy particularly noticeable in high frequencies since energy is usually concentrated in the low frequencies. The attack detection and modelling method developed there was based on the following requirements: the method should not use additive analysis results, in order to be usable for other purposes (segmentation, instrument recognition, etc.), it should succeed in every type of sound (particularly polyphonic sounds) with good time accuracy, it should be simple to use: the analysis parameters

should, as much as possible, be adjusted automatically, it should be tested on a data base of sounds including polyphonic mixtures of percussive and non-percussive sounds.

II. RESEARCH CONDITIONS

The database consisting of representative samples presenting the abilities of choir singers is important point of that study. The database used here is the extension of the database created under the research project of West Pomeranian University of Technology: "Computerized methods of supporting the process of training choir voices" [11]. The recorded singers are singing in the choir of the same university. The samples in the database are divided into categories reflecting different aspects of the singer's practices. The content of the database allows to extract the selected parameters of singing voice. The exercises were selected from a set usually used during the voice production training. It is possible to investigate for example intonation [12], vibrato feature, tremolo, sonority, noise [13] and other features. It is possible to perform some more general investigations over the database as, for example, singing voice quality assessment [14], [15]. For that study the exercise containing the vowel "a" sung at one pitch for a few seconds was chosen. In fact, the most interesting part here is situated at the beginning of voiced part of the sample.

The recordings used in this article were taken in the specially arranged room, with proper conditions for the recording session. All of the sound material was recorded with a 24 bit resolution, with the sampling frequency of 48 kHz. Higher recording parameters give some wider possibilities of editing of the recordings. The sessions were recorded as a whole for each singer. The division into elementary parts, called sequences or phrases (piano, singer), was done with use of a program with a tool for automatic segmentation. The process was performed under supervision of an expert to provide the possible best prepared samples.

Eight males, representing baritone voice, were subjects for the experiment. They were chosen as representative voices from the group of singers. Each person was in good vocal and health condition. The samples were recorded twice to have the possibility to compare the results. The sound pitches selected for the study were chosen to be comfortable to sing for all of the singers. The pitches covered the range from pitch number 10 in octave 2 (sound A) to pitch number 5 in octave 4 (sound e2). There were 2 sets of the samples for each singer. The sets consisted of 5 samples, so together there were 80 samples available to analyse (8 singers, 2 sets, 5 samples for each set).

III. RESEARCH IDEA

The idea of the research was to develop a descriptors of attack on sound. Those descriptors should map experts' evaluation of analysed signal. For this reason, the evaluations of attack on sound were carried out parallelly by three experts. Evaluation marks given by the experts will be presented in the following part of this publication. Finally both results, obtained by computer methods and given by human experts,

will be compared to decide whether the similar conclusion can be drawn.

A. Experts' Assessment

Three experts were asked to assess the sound attack in the samples of singing. They were instructed about the definition of the attack on sound defined in this study. They had a possibility to listen to the voice examples before they started the evaluation. It was done to teach the experts the rules of the assessment and to make sure the samples would be evaluated in the same way by all of the experts. The experts had to assess the attack feature using 1-3 scale, where 1 meant so called soft attack, 2 meant medium attack and 3 meant hard attack. The experts were also evaluating the quality of the attack on the sound. To assess the quality they used a rating scale 1-5 where 1 was bad quality and 5 meant very good quality. It should be reminded here that every expert had to evaluate 2 sets of samples for each singer. Every expert was assessing the set of the samples according to the procedure:

- at first they were able to listen to 6 different, randomly selected, samples,
- 1st set of samples was played as first and then it was evaluated by the expert,
- 2nd set of samples was played in the random order.

The experts did not see the ratings of each other. During the assessment process three experts have assessed 8 male singers. There were 2 sets for each singer. Each set consisted of 5 samples. Concluding, samples were assessed 240 times. Taking into account there were 2 marks for each sample (the type of the attack and the quality of the attack) it gave 480 marks.

B. Computer Aided Evaluation

During the preliminary studies, it has been established that the attack on the sound will be analysed according to the definition described in the first section of the article. At the beginning of the study it was assumed that the factors will be applied to three areas of the voice signal: time, amplitude and pitch [2]. Parameters of time have been given in seconds. The generalized Hilbert envelope was used to describe the characteristics of the amplitude. To illustrate each envelope the following parameters, which are also described in fig. 2, were selected:

- angle of attack on volume - directional factor of linear function of approximation for the amplitude envelope (fig. 2b),
- mean square error between the linear approximation and the actual envelope (fig. 2b),
- ratio of the energy of attack in relation to the energy of the entire signal represented by equation 1,
- values of derived envelope calculated by eq. 2 - the total sum of values and the number of values above and below zero (fig. 2c).

According to the theory, if both the first and the second parameter has a low value, we should be dealing with perfectly

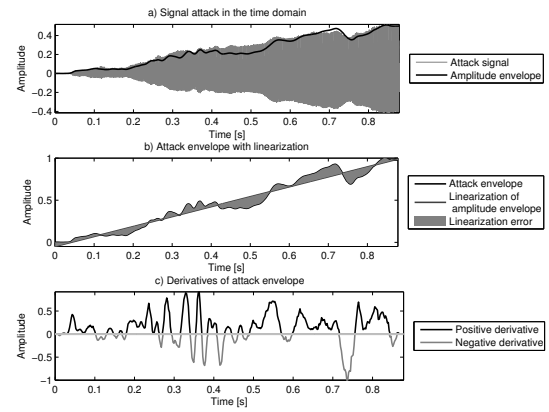


Figure 2. Presentation of some attack parameters based on amplitude values

soft attack. Other parameters were selected as complementary ones, in order to test the possibility of their use.

$$E = \frac{\sum(Attack^2)}{\sum(Signal^2)},$$

where E -is energy parameter,

$Attack$ -fragment of signal during the attack phase,

$Signal$ -whole signal.

(1)

$$f'(x) = \begin{cases} 0 & : x = 1 \\ f(x) - f(x-1) & : x > 1 \end{cases}, x \in \mathbb{N}_+,$$

where $f'(x)$ -is derivative in point x ,

$f(x)$ -is base signal

x -is index of element in vector.

(2)

The third set of analysed parameters concerned the pitch. This set meant to show how the singer sets the pitch during the attack on the sound. The parameters are displayed in fig. 3, and include following sets of data:

- angle of attack on pitch - directional factor of linear function of approximation for the pitch track (fig. 3b),
- mean square error between the linear approximation and the actual pitch track (fig. 3b),
- maximum pitch value registered with the upper pitch limit given by the value of expected tone plus 200 Hz,
- values of derived pitch track calculated with eq. 2 - the total sum of values and the number of values above and below zero (fig. 3c).

If the first parameter is less than zero that means the singer is attacking the sound from a higher frequency and trying to reduce the frequency to match the sound with the referenced tone. In the opposite case the singer increases the frequency in order to match the tone. Other parameters were selected as complementary ones, to test the possibility of their use.

IV. THE RESULTS

As it was mentioned before, the investigation consisted of two parts. In the first part experts were asked to assess

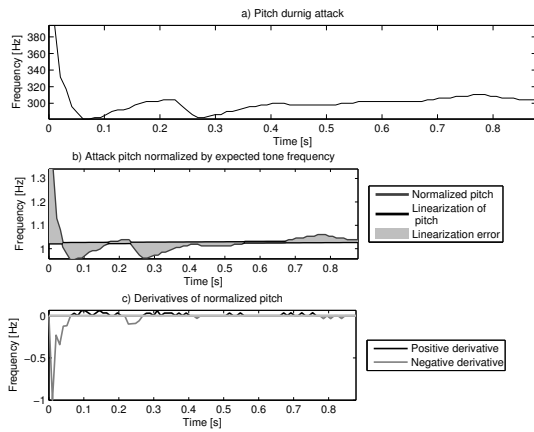


Figure 3. Presentation of some attack parameters based on pitch values

the samples from the database. In the second part a set of calculations over the samples was performed. After that, correlations between the results were searched.

A. Computer Aided Evaluation Results

As it was planned before, the special calculations over the chosen samples were performed to obtain information about the attack on the sound. One of the most natural parameters of the attack calculated on the sound sample was the angle of attack calculated over the volume envelope. The parameter is interpreted here as a directional factor of a linear function which is an approximation of the amplitude of the envelope (fig. 2b). The values obtained for 5 samples for each singer are presented together in fig. 4. It can be seen in the graph that the angle of the attack is higher in case of the boundary samples. It is because these pitches are too high or too low to be enough comfortable for singing. It can be observed that in the tested group there are singers having problems at every sample and the singers having some discomfort only at boundaries. It is apparent from the graph that the best are singers s21 and s27, as the angle is generally lower for those singers.

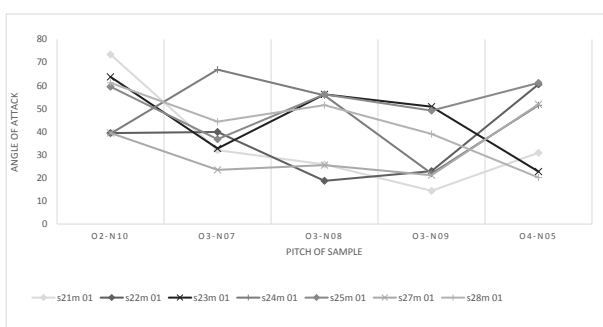


Figure 4. Angle of amplitude attack parameter (version 1 of set of samples)

The results obtained for mean square error between the linear approximation of the sound attack and the actual envelope (fig. 2b) are presented in fig. 5. Analysing that graph, and comparing to the previous one it should be noted that the

parameter for singers s21 and s27 differentiate the singers. The singer s21 has higher values of these parameters than s27 and for this reason it can be stated that s27 is better, more stable, than s21 (although, considering the second attempt, s21 became more stable). Combining those two parameters together it is possible to obtain a decision rule better describing the abilities of the singers.

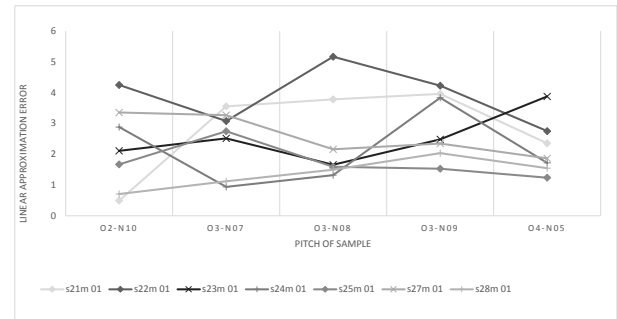


Figure 5. Mean square error parameter for the actual attack envelope (version 1 of set of samples)

Another example of the calculated parameter is ratio of the energy of attack in relation to the energy of the entire signal (represented by equation 1). The results visible in fig. 6 again highlight the singers s21 and s27. Those singers have greater energy values for the attack than the others. Especially if the energy values for each of the samples are summed up the differences between those singers and the others become really clear. The other results will not be described here in details, as the ones presented above sufficiently illustrate the general idea.

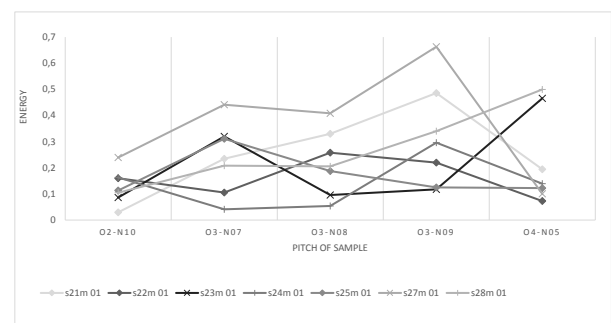


Figure 6. Energy parameter calculated over attack envelopes (version 1 of set of samples)

B. Experts Results

Experts were asked to assess the feature of sound attack of the sung samples. They were instructed about the rules adopted in this study. They had a possibility to listen to the voice examples before evaluation of the process. The procedure was set to guarantee the assessment of the samples in the same way. The experts evaluated, among others, whether the sound attack in the recorded sample is soft, medium or

hard. Those marks have been counted for each singer. That gave an entry information about the approach of the singer to the sound initiation. To illustrate that the figures 7 and 8 characterising the results of a good singer are presented. He was, generally, evaluated in the same way by all the experts. Some problems were observed on both ends of the analysed range. Additionally, from figure 8 one can see that the second trial (every singer was recorded twice) was better than the first one. Each expert has generally evaluated the initiation of the sound as soft, in this case. It should be noticed that all the samples lying in the middle range of the pitch were evaluated as soft. This is a result consistent with expectations of the experts.

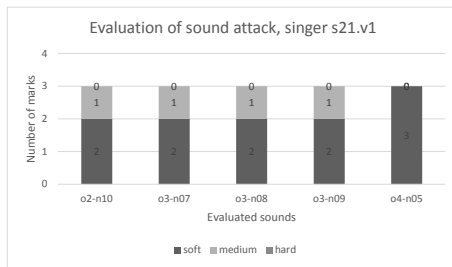


Figure 7. Evaluation of the singer s21 (sample version 1)

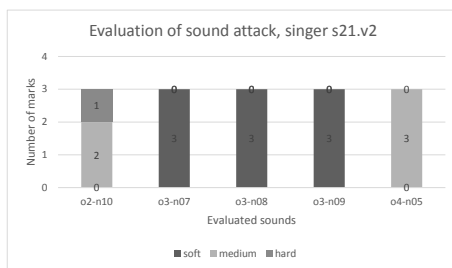


Figure 8. Evaluation of the singer s21 (sample version 2)

Another situation is presented in figure 9. The singer's results presented in that figure are not stable. This may be connected with some evaluation problems or, on the other hand, with low abilities of the singer.

All evaluation results were collected in one table and compared with the results obtained in an automatic way.

C. Correlation of The Results

The results, both given by the experts and obtained from computer calculations were compared to find an answer to the question - whether the computer methods proposed here are able to evaluate singing samples in the similar way as human experts do. To this end the relation between the vector containing experts' decisions and the vector containing values of each previously calculated parameters for each singer was estimated. The example provided here shows comparison of the expert marks given to each singer with the calculated angle parameters.

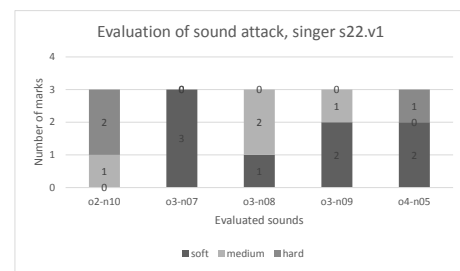


Figure 9. Evaluation of the singer s22 (sample version 1)

The upper part of the table I ("Summary of weighted expert's evaluation marks") summarize the marks given by the experts. The marks given by three experts were summed up and weighed to obtain general mark for each singer. For each mark (obtained in the previous step) it has been assigned an adjective describing the type of the attack according to the rule: mark 1-3 meant hard attack, 4-6 meant medium, 7-9 meant soft attack. The numerical marks were also averaged to give the average adjective mark. Similar operations were carried out on the second part of the table consisting of the values calculated using previously described procedure. The bottom part of the table contains the comparison of the marks, both, calculated and given by the experts. The row "Consistent values" shows how many values are consistent in both sets (expert's marks and calculated values). Taking into account all 5 samples (for each singer) it gives the mean value of consistency at the level of about 60 percent. Taking into account only the three middle samples the average consistency level between experts and computer evaluation also gives the value of about 60 percent. It leads to the conclusion that to achieve better results other parameters should be also involved.

The approach presented here is based on the expert's way of assessing the singing. The parameters proposed and calculated here are directly connected to the singing signal and can be easily observed in the signal. In the present article an example application of one of the parameters has been described. Therefore, further improvement of efficiency of computer assessment using the remaining, previously mentioned, parameters seems to be possible.

V. CONCLUSION

The article focuses in general on the problem of the sound attack in singing. Sound attack is understood here as an initiation of the sound. This characteristic is very subjective in perception, but it can be used as a criterion of singing quality evaluation. The attack on sound can be divided into several types. One of the types is a soft attack. This is identified with the beginning of the sound in a soft, fine, smooth manner. Much more common is a hard type of attack, which is normally used while speaking. Unfortunately, this state is tiring for a singer and can cause damage of the voice. That is why the hard attack is not advised while singing for a long time. The goal was to involve computer methods to evaluate

Table I
THE ANGLE OF THE ATTACK

Summary of weighed experts evaluation marks - 1-3: hard, 4-6: medium, 7-9: soft													
s22m 01	mark	s23m 01	mark	s24m 01	mark	s25m 01	mark	s27m 01	mark	s28m 01	mark	s21m 01	mark
4	medium	5	medium	6	medium	4	medium	5	medium	5	medium	8	soft
9	soft	9	soft	7	soft	3	hard	9	soft	5	medium	8	soft
7	soft	7	soft	8	soft	3	hard	9	soft	3	hard	8	soft
8	soft	7	soft	9	soft	4	medium	8	soft	3	hard	8	soft
7	soft	7	soft	8	soft	6	soft	8	soft	6	medium	9	soft
Mean value	Mean mark	Mean value	Mean mark	Mean value	Mean mark	Mean value	Mean mark	Mean value	Mean mark	Mean value	Mean mark	Mean value	Mean mark
7	soft	7	soft	7,6	soft	4	medium	7,8	soft	4,4	medium	8,2	soft
Calculated values of the angle parameter													
6	medium	4	medium	6	medium	4	medium	6	medium	4	medium	3	hard
6	medium	7	soft	3	hard	6	medium	8	soft	6	medium	7	soft
8	soft	4	medium	4	hard	4	medium	7	soft	5	medium	7	soft
8	soft	5	medium	8	soft	5	medium	8	soft	6	medium	9	soft
4	medium	8	soft	5	medium	4	medium	5	medium	8	soft	7	soft
Mean value	Mean mark	Mean value	Mean mark	Mean value	Mean mark	Mean value	Mean mark	Mean value	Mean mark	Mean value	Mean mark	Mean value	Mean mark
6,4	soft	5,6	medium	5,2	medium	4,6	medium	6,8	soft	5,8	medium	6,6	medium
Consistent values (%)													
60		60		40		40		80		40		80	
Consistent values excluding boundary pitches (%)													
66		33		33		33		100		33		100	

the attack in a singing to help the singer make progress. The methods should evaluate that feature in a similar way as human experts. To achieve that it was necessary to propose a method to analyse this problem. The results of the above automatic computer evaluation were compared to the marks given by the human experts. For the purpose of automatic evaluation it was assumed that the attack on the sound can be represented by objective parameters of signal describing subjective human impression. There was a need to answer the question whether among the calculated parameters were such that reflect experts' evaluation and thus can be useful to construct computer aided assessment system. The evaluation of the sound attack would be a part of that system. The investigation consisted of two parallel parts. In one of them the experts were assessing recordings of singing. In the other part a set of signal parameters was calculated for each singer in the context of sound attack. Among others the following parameters were calculated: the angle of attack calculated over the volume envelope, mean square error between the linear approximation of the sound attack and the actual envelope, the energy of attack in relation to the energy of the entire signal. During the study it has been found that among estimated features angle parameter, mean square error parameter and energy parameter can construct feature vector applied in a computer method of a sound attack quality assessment. The results concerning sound attack presented here may be useful for constructing a singing quality assessment system.

A large number of the results obtained for this study requires further, deeper analysis and may lead to subsequent applications.

ACKNOWLEDGMENT

The authors would like to thank the members of The Jan Szyrocki Memorial Choir of West Pomeranian University of Technology for their cooperation in building database of singing voices. The authors would also like to thank the experts who spent a lot of time listening to and evaluating the singing samples.

REFERENCES

- [1] Y. Meron and K. Hirose, "Separation of singing and piano sounds." in *ICSLP*, 1998.
- [2] M. Muller, D. P. Ellis, A. Klapuri, and G. Richard, "Signal processing for music analysis," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 6, pp. 1088–1110, 2011.
- [3] A. Holzapfel, Y. Stylianou, A. C. Gedik, and B. Bozkurt, "Three dimensions of pitched instrument onset detection," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, no. 6, pp. 1517–1527, 2010.
- [4] L. Mazurowski, "Computer models for algorithmic music composition," in *Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on*. IEEE, 2012, pp. 733–737.
- [5] K. Jensen, "Envelope model of isolated musical sounds," in *Proceedings of the 2nd COST G-6 Workshop on Digital Audio Effects (DAFx99)*, 1999.
- [6] J. Davids and S. LaTour, *Vocal Technique: A Guide for Conductors, Teachers, and Singers*. Waveland Press, 2012.
- [7] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler, "A tutorial on onset detection in music signals," *Speech and Audio Processing, IEEE Transactions on*, vol. 13, no. 5, pp. 1035–1047, 2005.
- [8] C.-C. Toh, B. Zhang, and Y. Wang, "Multiple-feature fusion based onset detection for solo singing voice." in *ISMIR*, 2008, pp. 515–520.
- [9] R. M. Alderson, *Complete handbook of voice training*. Parker Publishing Company, 1979.
- [10] X. Rodet and F. Jaillet, "Detection and modeling of fast attack transients," in *Proceedings of the International Computer Music Conference*, 2001, pp. 30–33.
- [11] M. Łazoryszczak and E. Pórolniczak, "Audio database for the assessment of singing voice quality of choir members," *Elektronika: konstrukcje, technologie, zastosowania*, vol. 54, no. 3, pp. 92–96, 2013.
- [12] E. Pórolniczak and M. Łazoryszczak, "Quality assessment of intonation of choir singers using f0 and trend lines for singing sequence," *Metody Informatyki Stosowanej*, pp. 259–268, 2011.
- [13] E. PÓROLNICZAK and M. KRAMARCZYK, "Computer analysis of the noise component in the singing voice for assessing the quality of singing," *Przegląd Elektrotechniczny*, vol. 91, pp. 79–83, 2015.
- [14] E. Polrolniczak and M. Kramarczyk, "Formant analysis in assessment of the quality of choral singers," in *Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), 2013*, Sept 2013, pp. 200–204.
- [15] P. Zwan and B. Kostek, "System for automatic singing voice recognition," *Journal of the Audio Engineering Society*, vol. 56, no. 9, pp. 710–723, 2008.