

Computer analysis of the noise component in the singing voice for assessing the quality of singing

Abstract. The article focuses on the analysis of a noise in singing voices using samples obtained from choral singers. Analysis of the singing voice quality is a complex task. There are various points of view and methods of analysis used to solve that problem. Musicians consider the voice of singers from the point of view of voice emission. Doctors analyzing the voice of a singer trying to determine his medical parameters. To analyze the quality of singing voice the equipment and tools characteristic for clinical practice can be used. In the analysis of the signals mentioned in the article some methods for the voice analysis have been adopted. The quality parameters of the singing voice were calculated on the basis of recorded samples. This article presents the results of research at the frontier of computer signal analysis and medicine. To achieve the goal in the study a method for the analysis of noise Noise Parameter based on GNE (Glottal-to-Noise Excitation Ratio) was used.

Streszczenie. Niniejszy artykuł skupia się na analizie szumu w głosie śpiewaczym. Analiza jakości głosu śpiewaczego jest złożonym zadaniem. W praktyce istnieją różne podejścia do analizy zagadnienia szumu w głosie. Muzycy patrzą na głos śpiewaczy z perspektywy emisji głosu. Lekarze analizując głos śpiewaka próbują ustalić jego parametry medyczne. Stosowane są przy tym narzędzia wykorzystywane do analiz medycznych. W przedstawionej w tym artykule analizie parametry jakościowe głosu śpiewaczego obliczono dla próbek dźwiękowych pozyskanych od śpiewaków chóralnych. Przedstawiono tutaj wyniki badań z pogranicza komputerowej analizy sygnału i medycyny. Aby osiągnąć założony cel wykorzystano metody analizy parametrów szumu na podstawie współczynnika GNE (Glottal-to-Noise Excitation Ratio). **Analiza szumu w głosie śpiewaczym**

Keywords: signal analysis, singing voice, singing quality

Słowa kluczowe: analiza sygnału, głos śpiewaczy, jakość śpiewu

Introduction

The direct motivation to undertake the research in the field of quality of singing voice is the need to support by computer analysis the education of voice production according to the requirements of choral ensembles. One of the goals of the choir is continuous raising the artistic level. The presented research can support this expectation. The quality of the choir depends largely on the individual voice quality and the skills of individual choir singers. The developed criteria of computer quality voice assessment can be useful in work of voice coach and conductor. Properly described evaluation criteria may allow for correction of the selected parameters can support a self-education of voice production.

There are various characteristics of voice which can be assessed by experts (and also by computer algorithms). One of the most obvious and easy to assess by human experts is the intonation. It can give the answer for the question if the singer is singing the melody properly. Another easy to recognize parameter of singing voice is vibrato. The concept of vibrato is generally known as the waving of frequency and the volume occurring in the voices of singers. Trainers of voice emission define the vibrato as a periodic, pulsing, and amount of intensity and tone, which is experienced as color impression. In this perspective, the vibrato concept is wider than usual waving of frequency. The third is the timbre phenomenon. Although this factor is very subjective in perception, it can be used as an evaluation criterion of singing. Among the characteristics of the timbre, experts point to the bright, dark, sonorous, matte, clean, harsh, warm or cool timbre and others. Some of these characteristics, such as the matt or rough timbre, are perceived negatively in case of singing.

Among the various parameters (attributes) of singing voice there is also a phenomenon connected to a noise component. It can cause the impression of humming, hissing or rustling voice which we named here as the noisy voice as all of the listed characteristics are connected to noise component. From the medical point of view it is connected to a disorder of vibration of the vocal folds of the larynx [1]. It can be result of voice disease or just individual

voice characteristic of a singer. Sometimes it is connected to a loss of high frequency components in the signal.

The presented research may be significant not only for the quality assessment of singing but also for the health of the singer. Noise component in a singing voice can be a kind of warning signal determining that a singer must undergo a diagnosis.

Literature review

Measures used to determine the degree of noise in voice signal are important group of features used for assessment of voice quality. Due to the fact that they are largely limited to the assessment of the accuracy of the harmonic structure the methods are commonly named HNR (Harmonic-to-Noise Ratio). In the literature a number of indicators can be found, but the most important are:

1. HNR_{Yumoto} (Yumoto, 1982) – is estimated in the time domain: the average glottal period is calculated after the length of each period is normalized, then each segment is compared with the average (segment). The variance of deviations is the measure of noise [4]. n the modified version HNR_{Qi} (Qi, 1992) segments are normalized in the time domain using the dynamic time warping [5].
2. NNE – Normalized Noise Energy (Kasuya, 1986) – is the ratio of non-harmonic signal energy to the total energy of signal. It is estimated by location of the minima between the harmonics in the range 1-5 kHz, which carry information about the level of noise [6].
3. HNR, SPI, TNI (MDVP, 1993) parameters – are based on the same transformation: spectral comb filtering, separating the harmonic and non-harmonic energy of the signal. Each measure is the ratio of energy value in specific frequency ranges [7].
4. CHNR – Cepstral HNR (de Krom, 1993) – it involves the cepstrum to determine the baseline of spectrum. After that the ratio of total energy to the base energy of spectrum is determined [8]. Further modifications rely on better fitting the baseline for the spectrum [9]. Often the measure is calculated according to the algorithm in [3]. The method uses discrimination between harmonic and noise energy in the magnitude spectrum by means of a comb-filtering operation in the cepstrum domain. The authors claim the

method is a valid technique for determining the amount of spectral noise and may be a useful parameter in the analysis of voice quality. The HNR measure is calculated here by using a variable window length equal to 5 pitch periods, liftering the pitch component of the cepstrum and comparing the energy of the harmonics with the noise floor. A sub-measures can be found in a literature. Those are for example: HNR5 which measures the HNR between 0-500Hz, HNR15 which measures the HNR between 0-1500Hz and HNR25 which measures the parameter in the range 0-2500Hz.

Reliability of all these measures depends on the effectiveness of the F0 detection algorithm. HNRQ_i and HNR_{Yumoto} measures are particularly sensitive due to the need of precise detection of glottal segments in the time domain. Their disadvantage is also the fact they do not allow to estimate a HNR parameter for the required frequency range. Other measures require the mean F0 correctly estimated in the test signal frame.

Significant problem while determining the NNE is detection of harmonics and de-termination of their range. While calculating the CHNR the problem can be precise estimation of the baseline of the spectrum which is used to separate non-harmonic and harmonic energies.

As a competitive ratio of noise a parameter GNE (Glottal-to-Noise Excitation Ratio) have been proposed in literature. The authors indicate the advantage of their method over CHNH and NNE [10]. The algorithm does not require the information estimated on F0 trajectory so it should result in reliability of the measure. The disadvantage of the method is high complexity.

In that study, to measure the singing voice quality of choral singers in the context of noise in the singing GNE parameter is used. Comparing GNE to a competitive CHNR parameter it should be noted that CHNR is reported to be F0 dependent so the results can be compared when calculated for the same pitches. Although the whole text is written using Arial font, for symbols please use the Times New Roman italic – for example *J* and not *J* (i.e. the same symbols as in the equations).

Research material

At the study a database consisting of representative samples presenting the abilities of choir singers were used. Detailed information about the database used in the study, conditions of recordings and about singers are presented in subsections.

The database

In the study the database created in the frame of a research project of West Pomeranian University of Technology: "Computerized methods of supporting the process of training choir voices" [11] was used. It was decided to use for the project own sound material. The reason for this decision was the necessity of having a database which would represent the abilities of choir voices as a specific class of voices. The Assumption was that there are certain parameters of singing which may be measured irrespectively of what is being sang and that these parameters are influenced by what is being sang. Therefore, the approach to the creation of such a database is also of significant importance.

The database consists from five vocal exercises which are used as material for assessing the quality of singing voice. The exercises were selected from a set usually used during vocal trainings. The most important features for particular exercises are: the text, sound scheme in reference to the key of C-major and the approximate tem-po of the performance of the exercise. Exercises have been

chosen to allow determining different problems: intonation, timbre, vibrato, noise and others.

The database is still developed. Most of the recorded singers are singing in the choir of the same university. The new, recently developed part of the database, is the set of recordings of amateurs (non-singers).

The study group and used samples

The study was carried out on the group of twenty choir singers. Each person was in good health. They were representing different levels of advancement to which were assigned by experts (beginner, intermediate and advanced).

In the male group three singers named in base as s05m, s14m and s16m were assigned to the beginners group and the group of intermediate singers consisted of another three singers: s02m, s04m and s05m. The last two male singers (s07m and s10m) were assigned to the highest group of good singers (advanced).

In the female group three singers named as s01f, s15f and s20f were assigned to the group of beginners. The group of intermediate singers consists of seven singers. One female singer (s13f) was assigned to the highest group of good singers (advanced).

Table 1. Description of the audio database records

Symbol of singer (v2.0 od DB)	Gender	Voice	Level	Years in the choir
s01f	Female	Soprano	Beginner	5
s02m	Male	Bas	Intermediate	16
s03f	Female	Soprano	Intermediate	15
s04m	Male	Bas	Intermediate	3
s05m	Male	Bas	Beginner	1
s06f	Female	Soprano	Intermediate	5
s07m	Male	Bas	Advanced	18
s08f	Female	Alto	Intermediate	7
s09f	Female	Alto	Intermediate	3
s10m	Male	Bas	Advanced	6
s11f	Female	Soprano	Intermediate	3
s12m	Male	Bas	Intermediate	3
s13f	Female	Soprano	Advanced	15
s14m	Male	Bas	Beginner	2
s15f	Female	Alto	Beginner	2
s16m	Male	Bas	Beginner	2
s17f	Female	Soprano	Intermediate	4
s18m	Male	Tenor	Intermediate	6
s19f	Female	Alto	Intermediate	2
s20f	Female	Soprano	Beginner	1

The group was singing the sequence of vowels 'a-e-i-o-u'. The exercise, named 'e01', used for the tests is close to the real singing conditions.

Conditions of recordings

The samples were recorded in a specially arranged place to provide proper conditions for recordings. The acoustic of the room was specially modified to avoid sound reflections and reverberations. To achieve that a special sound-absorbing foam panels with an appropriate geometric profiles was used. For the purpose of recordings the capacitor microphone AT 4050 connected to professional audio interface was used.

The recorded sound was processed at the computer equipped with appropriate software. The singers were singing with their faces directed at the microphones. The distance face-microphone was about 35-40 cm. Also at the back of the singer the sound absorbing foam was situated. The recordings were carried out by 2 technical persons –

one was the person giving instructions and initial sounds to the singer (singer should sing his phrase on the specific pitch), second was operator of the recording equipment.

The study of noise in the singing voice

For the purpose of this article a noise estimation have been performed. To achieve the goal of the study NP (Noise Parameter) have been used.

NP parameter used in that study is based on GNE (Glottal-to-Noise Excitation Ratio) parameter which has been developed as an acoustic measure to noise determination in the signal generated by vocal folds [2]. The authors claim that the measure is almost independent of frequency modulation noise (jitter) and amplitude modulation noise (shimmer). It is generally F0-independent method.

According to [2] NP (Noise Parameter) ratio of the voice signal based on the GNE is also calculated and is defined as following:

$$(1) \quad NP = 1.5 + \frac{0.695 - (1 - 10^{GNE})}{0.242}$$

The basis of GNE coefficient is the assumption that it the vocal folds, during the closing arouse synchronously frequencies in different bands. Irregular air flow is manifested uncorrelated frequencies in different bands. In that study the GNE parameter is estimated for the band 3000Hz and is estimated as following:

1. the signal is down-sampled to 10 kHz,
2. the inverse filtering of the signal is performed,
3. the Hilbert envelopes are calculated (the absolute value of the complex analytic signal) of different frequency bands with fixed bandwidth and different center frequencies,
4. every pair of envelopes for which the difference of their center frequencies is equal or greater than half the bandwidth is considered and the cross correlation function between such envelopes is calculated,
5. the maximum of each correlation function is saved in a maxima vector,
6. the maximum from the maxima vector obtained in the previous stage is the resulting value.

Due to the correlational nature the maximum of GNE equals 1. It happens when two different envelopes (for different bands) are exactly the same. GNE decreases in the case of increasing contribution of noise in voice and, at the same time, reducing the frequency coefficients. This factor, in an extreme case, may be associated with the so-called humming voice.

The experiment and the results

The goal of the performed experiment was to find a new concept of comparing singers by Noise Parameter, calculated as in formula (1), for choir singers singing sounds at successive pitches. Singers were singing the sequence 'a-e-i-o-u' starting from the highest pitch they were able to achieve. Parameter was estimated over 50 second windows. To obtain more observations the mean, minimum and maximum values of this parameter were calculated for each sequence. Figure 1 presents example plot of those values for singer s04m. The numbers on the x axis represents consecutive sequences of 'aeiou' sung by singer. The y axis corresponds to values of NP.

The x axis represents the numbers of successive singing samples starting from the highest pitch in descending order. The y axis represents NP values. The graph shows that not only the mean value but also the range (difference of the maximum and minimum values) can give interesting, usable information about singers.

While interpreting the above graph it can be seen that the singer has some problems with voice stability in the context of noise. It can be said that the singer has less vocal problems in the lower part of his vocal scale. It can be also visible that there is a pitch limit beyond which singer has serious vocal problems. He shouldn't sing at the lowest pitches. We expected the singer will achieve the best results – the lowest NPmean value, the lowest NPrange in the middle of his scale but it happened differently. Another example shows a person which seems to be generally good singer. The NPrange values are really low, excluding the first sung notes (Fig. 2).

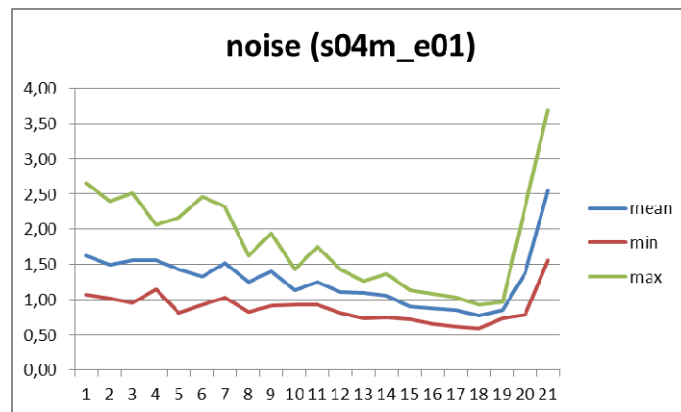


Fig. 1. NPmean, NPmin and NPmax estimated for s04m singer for the set of the samples 'aeiou'

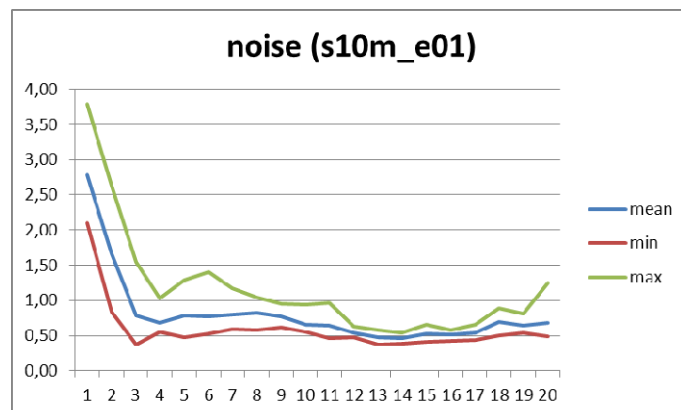


Fig. 2. NPmean, NPmin and NPmax estimated for s10m singer for the set of the samples 'aeiou'

Visual inspection of the calculated graphs allowed to draw conclusions about the possibility of assessing the quality of a singer using Noise Parameter. Here we present the main conclusions:

to compare quality of singers in the context of Noise Parameter (NP) the parameters describing preserving quality of the voice should be introduced,

to reliably estimate the parameters samples from the ends of the ranges shouldn't be taken into account (we decided to exclude 3 samples from both sides).

To compare singers in the context of Noise Parameters we propose to estimate the following values:

- mean value of NPmean values in the evaluated set of samples – meanNPmean,
- range value of NPmean values in the evaluated set of samples – rangeNPmean
- difference between maxNP and minNP values calculated in the maxNPmean location (named here rangemaxNPmean, and finally named rmaxNPmean),

- maxNP-minNP values calculated in the maxNPmean location (named here rangeminNPmean, and finally named rminNPmean).

The meanNPmean is the most general parameter describing voice quality here. It gives general information about the level of noise in the voice over the evaluated voice scale.

The rangeNPmean gives a little bit detailed information about behavior of a voice during singing. It provides information about stability of NP over the evaluated set of samples.

The rmaxNPmean and rminNPmean are giving additional information about voice behavior. It gives information about the scale of a problem with a voice.

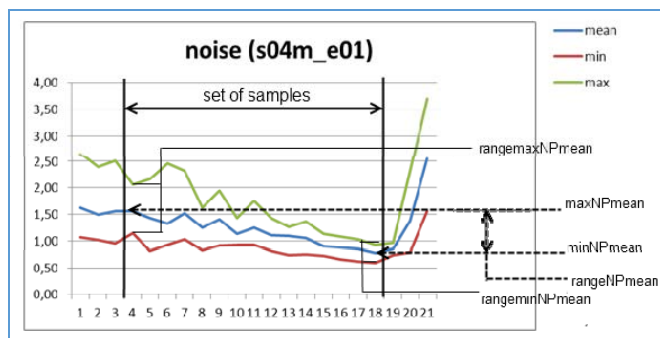


Fig. 3. Explanation of introduced quality parameters for NP

The calculations were performed for men and women as well but because it is not possible to compare those groups directly the results presented in the table 2 represent male singers as an example. In the second part of that section the results for women will be presented.

Table 2. Introduced quality parameter calculated over a set of samples for each singer

Singer	s02m	s04m	s05m	s07m	s10m	s12m	s14m	s16m	s18m
meanNP _{mean}	1,36	1,17	0,91	0,85	0,64	1,08	1,49	0,91	1,16
rangeNP _{mean}	0,75	0,79	1,29	0,44	0,36	0,39	1,46	0,83	0,76
rmaxNP _{mean}	1,76	0,92	0,92	1,86	0,46	0,93	1,12	1,00	0,67
rminNP _{mean}	0,59	0,33	0,41	0,60	0,16	0,86	0,97	0,48	0,41

Analyzing the table it can be noticed that there are significant differences in the results between the singers. Parameter meanNPmean, the most general parameter describing voice quality here, gives general information about level of noise in the voice over the evaluated voice scale. It shows that the singer s10m achieved the lowest mean NP over all his samples. It means that his singing is almost free of noise. It is interesting that also other calculated parameters are at the low level. The rangeNPmean value is low. It means that the noise is stable among all the samples. The rmaxNPmean and rminNPmean are giving additional information about the voice behavior – here the values are also low. Low rmaxNPmean means that for the least comfortable sound, where the NPmean is the highest, the voice keeps his parameters. The next singer worth mentioning is s07m. Almost all values are on low level. The only one problem here is the high value of rmaxNPmean which shows highly unstable noise parameter for the less comfortable pitch. Both singers belongs were assigned to advanced group by experts.

A surprise are the singers s05m and s16m. They were classified by experts to the beginner group but it should be noted that level of advancement is not the same as level of noise in the voice. Experts determine a levels of advancement taking into account also such parameters as vibrato, timbre and sonority.

Other singers are generally close to each other in the context of NP. Taking into account the range of acceptable NP values mentioned in [2] it must be noted that the level of noise in the examined group of singers doesn't indicate any pathologies.

The same calculations were performed for female singers. The results are presented in the table 3.

Table 3. Introduced quality parameter calculated over a set of samples for female singers

Singer	s01f	s03f	s06f	s08f	s09f	s11f	s13f	s15f	s17f	s19f	s20f
meanNP _{mean}	1,04	1,33	1,24	1,10	1,43	1,68	1,60	0,83	1,40	1,04	2,44
rangeNP _{mean}	1,71	1,45	1,64	0,89	1,65	1,97	1,96	0,84	2,02	1,60	1,80
rmaxNP _{mean}	2,44	2,83	2,30	2,62	2,15	2,39	0,81	1,00	2,61	1,87	2,12
rminNP _{mean}	0,44	0,35	0,29	0,30	0,92	0,91	0,47	0,42	0,59	0,46	1,30

The figure 4 is a visualization of the table 3.

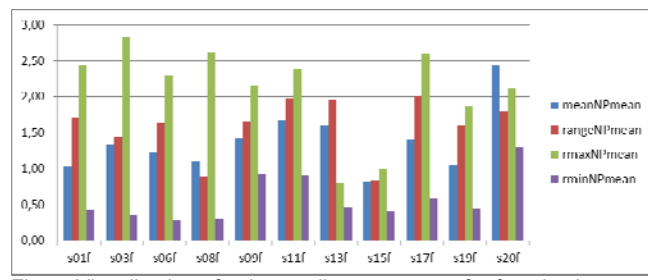


Fig. 4. Visualization of noise quality parameters for female singers

It can be seen that also in the female group there are leading voices in the context of noise. Singer s15f presents the low values of all the parameters. Mean noise is on the lowest level among all the singers. Next two singers (s01f and s19f) have achieved quite low level of mean noise value but other parameters show a differences between them and the best singer. The rangeNPmean shows that voice quality is changing significantly during singing the subsequent sounds (in descending order). Singer s08f shows generally good quality of voice but it had significant emission problems while singing the least comfortable sounds (probably the pitch was too high for the singer).

Observations of the results show that the rmaxNPmean and rminNPmean parameters are not giving general information about the voice quality in the context of noise. They provide rather local information as (for example) at which pitch the singer is having a voice production problem. It is often that the biggest noise problem can be observed at the highest pitches. The smallest problems are usually observed at the end of the sung sequence. It is either not clear whether it is not the F0-dependency effect.

The meanNPmean and rangeNPmean parameters should give similar possibilities of assessment to a real experts. To justify this statement the Pearson Correlation Coefficient for vectors of NPmean values and vectors of points describing amount of noise given by experts in the consecutive samples was estimated. For example, the correlation coefficient in the case of male singers between those singers reached the value of 0,54 when taking into account the first expert and 0,76 for the expert number 3 (the expert number 2 was excluded from the investigation after Rater Agreement Coefficient calculated for experts showed a low agreement value between this expert and two others). It proves that the estimated values of Noise Parameter was changing in the same way as ratings of the experts.

After taking into account:

- the fact that the value of Noise Parameter for the normal voice shouldn't exceed the value of 2,0 [12],
- the fact that lower NP value means lower noise level,

- the fact that the meanNP_{mean} is more significant than the rangeNP_{mean} parameter, and the earlier observations a decision rule is proposed here to be used in automatic voice quality assessment. The proposed Decision Parameter (DP) is a sum of weighted meanNP_{mean} and rangeNP_{mean} parameters (2).

$$(2) \quad DP = 0,75 * \text{meanNP}_{\text{mean}} + 0,25 * \text{rangeNP}_{\text{mean}}$$

The final decision about voice quality should be taken according to the below rule:

if(DP>=2,0) then D = 'pathological (P)', else
 if(DP>=1,6 and DP<2,0) D = 'unsatisfactory (U)', else
 if(DP>=1,2 and DP<1,6) D = 'weak quality (WQ)', else
 if(DP>=0,8 and DP<1,2) D = 'medium quality (MQ)', else
 if(DP>=0,4 and DP<0,8) D = 'high quality (HQ)';
 if(DP>=0,0 and DP<0,4) D = 'very high quality (VHQ)';

The presented strategy of the assessment of singing voice can be implemented into an automatic system which will indicate the weaknesses of a singer. The table 4 presents the obtained Decision (D) values for men singers.

Table 4. Decision Parameter value and Decision for male singers

Singer	s02m	s04m	s05m	s07m	s10m	s12m	s14m	s16m	s18m
meanNP _{mean}	1,36	1,17	0,91	0,85	0,64	1,08	1,49	0,91	1,16
rangeNP _{mean}	0,75	0,79	1,29	0,44	0,36	0,39	1,46	0,83	0,76
DP	1,21	1,08	1,01	0,75	0,57	0,91	1,48	0,89	1,06
D	WQ	MQ	MQ	HQ	HQ	MQ	WQ	MQ	MQ

It can be seen in the table 4 that the automatic decision values overlap with the opinions of the experts. A similar situation occurred in the case of women's ratings however it should be mentioned that the quality of one of the women singers was assessed as pathological.

Conclusion

The article was focused on the problem of computer assessment of presence of noise in singing voice. This characteristic is well recognizable by perception and it can be used as an evaluation criterion of singing voice quality. For the purpose of computer evaluation it was assumed that the noise parameter in a singer's voice can be represented by objective parameters describing a quality in the context of noise. The goal of the study was to answer the question whether the calculated parameters can be used to assess quality of voice from perspective of noise component. To achieve the goal a method functioning in conditions similar to the real singing was necessary. In the study the investigation on Noise Parameter was performed and strategies of assessment were proposed. The new

parameters were calculated for each singer. From the set of four parameters two were chosen as possible to be applied in a computer method of quality assessment in the field of noise. Finally a decision parameter and decision rule was proposed to be used in an automatic voice quality evaluation.

REFERENCES

- [1] Hammarberg B., Fritzell B., Gauffin J., Sundberg J., and Wedin L.: Perceptual and acoustic correlates of abnormal voice qualities. *Acta Otolaryngol*, No. 90, pp.441–451 (1980)
- [2] Michaelis D., Gramss T., and Strube H. W., Glottal to noise excitation ratio - a new measure for describing pathological voices, *Acustica / acta acustica*, No. 83, pp.700–706 (1997)
- [3] Guus de Krom: A Cepstrum-Based Technique for Determining a Harmonics-to-Noise Ratio in Speech Signals, *Journal of Speech, Language, and Hearing Research*. Vol. 36, 254-266 (1993)
- [4] Yumoto, E., Gould, W.J., Baer, T. Harmonics-to-noise ratio as an index of the degree of hoarseness. *Journal of the Acoustical Society of America*. Vol. 71, 6 (1982)
- [5] Qi, Y.: Time normalization in voice analysis. *Journal of the Acoustical Society of America*. Vol. 92 (1992)
- [6] Kasuya, H., and others. Normalized noise energy as an acoustic measure to evaluate pathologic voice. *The Journal of the Acoustical Society of America*. Vol. 80, 5 (1986)
- [7] Delyski, D. D.: Acoustic Model and Evaluation of Pathological Voice Production. *Proceeding of EUROSPEECH'93*. Berlin (1993)
- [8] de Krom.: A cepstrum based technique for determining an harmonics-to-noise ratio in speech signals. *Journal of Speech and Hearing Research*. Vol. 36 (1993)
- [9] Qi, Y., Hillman, R.: Temporal and spectral estimations of harmonics-to-noise ratio human voice signals. *The Journal of the Acoustical Society of America*. Vol. 102 (1997)
- [10] Fröhlich, M., Michaelis, D., Strube, H. W.: Acoustic "breathiness measures" in the description of pathologic voices. *In Proceedings ICASSP'98* (1998)
- [11] Łazoryszczak, M., Pórolniczak, E.: Audio database for the assessment of singing voice quality of choir members. *Elektronika*. No. 3, pp. 92-96 (2013)
- [12] Michaelis, D., Fröhlich, M., Strube, HW.: J Acoust Soc Am. Selection and combination of acoustic features for the description of pathologic voices. Vol. 103(3), pp. 1628-39 (1998)

Authors: dr inż. Edward Pórolniczak, West Pomeranian University of Technology, Szczecin, Al. Piastów 17, 70-310 Szczecin E-mail: epolrolniczak@wi.zut.edu.pl; mgr. inż. Michał Kramarczyk, West Pomeranian University of Technology, Szczecin, Al. Piastów 17, 70-310 Szczecin, E-mail: mkrmarczyk@wi.zut.edu.pl.