DOI: 10.15439/2019F316
ISSN 2300-5963 ACSIS, Vol. 18

Comparison of singing voice quality from the beginning of the phonation and in the stable phase in the case of choral voices

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*—In the process of acoustic voice analysis, in this case of singing, it is important that the sound samples contain a stable phase of phonation. Sometimes, however, it is not possible. This study was prepared to determine how big are the differences between the values of the acoustic parameters obtained for the initial phase of phonation and for the stable phase of phonation. The values of acoustic parameters, such as, among others shimmer, jitter, RAP, PPQ, APQ, HNR or SPR were estimated for registered singing samples in the initial phase of phonation and in the middle phase. The analysis were performed over the samples of singing of the vowel 'a' recorded many times for different pitches. In the process of analyzing of the obtained results, it was found that the impact of the selection phase of phonation for analysis is crucial in assessing the singing voice quality.**

I. Introduction

THE motivation for taking up the research on the singing voice acoustic parameters analysis was the need of assessment of singing quality. It may be useful for training lessons of voice production. It can be useful to help singers make a progress and it may allow for self-correction of selected voice parameters. It can be also very important for the choirs constantly working on the voice.

To analyse singing voice a must is to determine intonation or vibrato [1]. Some authors try to analyse the singing voice based on mel-cepstral features [2] or voice and speech features like Singing Power Ratio [3]. Anyway, there are many available acoustic parameters which may be investigated in the singing voice quality assessment.

One of the problems in the case of singing voice analysis is to obtain stable values of the parameters from the samples of singing. Due to the character of the singing signal envelope, it seems that determining the values based on the initial fragments of the singing recording may have an impact on the analytical process. So if it turns out that the values of the parameters from the beginning and the middle of the phrase differ significantly, it means that short signals for which a significant part is the attack and decay phase should not be applied to analysis.

During the creation of the database, the authors of this work paid attention to the quality of the samples. The recordings were carried out in appropriate conditions, the samples were subjected to precise segmentation. The samples obtained by us last for 3-4 seconds so they ensure that the middle part is the most valuable. In order to determine the analysed parameters, the samples were cut at an additional 5% from the beginning and the end. Regardless of these treatments, the authors were not sure if the samples throughout the entire run have similar quality or maybe the initial fragments are out of quality from the middle ones. This doubt was behind the undertaking of the described research and observations.

The voice, in general, is produced by a vocal instrument consisting of three elements: a breathing apparatus, oscillating vocal folds and a vocal tract. Breathing has a decisive impact on all activities related to voice emission. The entire phonation process can be represented using the ADSR (Attack-Decay-Sustain-Release) model that describes production of a single sound. It can be used for sound analysis [4], [5] and synthesis [6]. It is also the description of the sound waveform in the MIDI standard [7].

The sound attack is first part of the ADSR envelope. The attack ('on the sound') is used to modify the first phase of the amplitude envelope [8] of the generated sound, in which the sound gains the highest amplitude. This is followed by decay section during which the amplitude is reduced. As the next a sustain stage characterized by a stable pitch amplitude is visible. The ADSR envelope is completed at the release stage. The ADSR sections are illustrated in fig. 1.

The quality of singing is related to breathing. The proper breath before the phonation results in a good beginning of each phrase [9]. This is especially important in singing phrases that start with a vowel. The attack is more precise in the case of professional singers. Choral singers are less precise at this stage of voice production. The practice can solve this problem, but choir members usually develop their voice in groups, making they vocal abilities are similar in the group. It should be kept in mind that the presented investigation concerns the choral voices analysis.

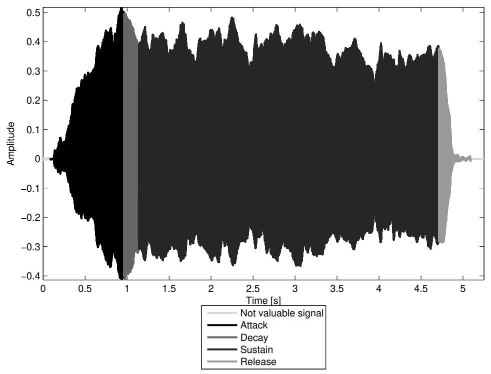

Figure 1. Example of a sung phrase divided into ADSR sections

The first phase of voice production perceived from the point of view of the ADSR envelope looks highly variable as its values increase from 0 up to the highest envelope value. However, if to take into account the physiological aspects of the human body, it will turn out that the production of sound with varying loudness and stable fundamental frequency requires a lot of effort and experience. This leads to the suggestion that greater differences in the qualitative parameters between the initial phase of voice emission and the phase with stable amplitude may be more visible in people with less vocal experience - which should also be noted when analysing the obtained values.

II. Research Conditions

The database used here was created as part of the research project of the West Pomeranian University of Technology: 'Computer methods to support the process of choral voice training' quote [10] and expanded at a later time. The content of the database allows to estimate selected parameters of the sung voice. It is possible to examine, for example, the intonation [11], the function vibrato [1], tremolo, sonority, noise [12] and other variables. It is possible to carry out more general database research, such as for example the voice quality evaluation [13], [14].

For this study, recordings containing the vowel /a/ sung on one pitch for a few seconds, were selected. In the further part of the analysis, the subjection will be the initial part of the signal and the middle part (sustain) of the samples.

The recordings used in this article were made in a specially arranged environment, with appropriate conditions for the recording session. All recordings were done with a resolution of 24 bits and with a sampling rate of 48 kHz. All singers where provided with referential signal at the beginning of recording of each sample. The process was carried out under the supervision of an expert to ensure the best quality of the samples.

The analysed group of singers consists of 16 men and 7 women. All these people have so much vocal experience that they are able to sing the sound at a given frequency. The examined persons are characterized by a varied work experience in the team (1-20 years). The pitches range recorded for each person reflects their vocal abilities - recorded sound represents the person's voice scale.

III. Research Idea

The aim of the research was to confirm or reject the hypothesis that that values of the quality parameters estimated for the first part of the signal present worse quality of the singing comparing to the middle part of the signal.

To reach the goal of the study a number of acoustic (vocal) parameters can be used. Some of them are: SPR, LTAS. Another popular are: jitter and shimmer measures, harmonic-to-noise-ratio (HNR), formants (including singer's formant (SF)), Spectra Centroid, energy ratio (ER), percentual variability (PV) [15] and others.

Many of those mentioned above are used in this study. Analysing acoustic parameters we were observing a differences in values estimated for the Entry of phonation and the middle.

The chosen acoustic parameters have been estimated and analysed based on the recorded vowel /a/. The set of the estimated parameters consisted of the most recognized by the scientists in the field of voice analysis:

- Jitter,
- Shimmer,
- HNR35,
- SPR.

Above were implemented and calculated using Praat [16] (via Parselmouth [17]) and Matlab using dedicated libraries (VoiceSauce [18], YIN [19]) or our own implementations.

Jitter and shimmer are the two common perturbation measures in acoustic analysis. Jitter is a measure of frequency instability, while shimmer is a measure of amplitude instability. In Praat we have access to multiple kinds of jitters (local and local absolute, RAP, PPQ5, Jitter DDP) and shimmers (local and local in dB, APQ3, APQ5 and APQ11, Shimmer DDP)

Harmonics to Noise Ratio (estimated here in Matlab via VoiceSauce [18]) indicates ratio of harmonics values comparing to noise level. In our case HNR35 is ratio measured between 0-3500Hz.

SPR (Singing Power Ratio), for the needs of this publication, was calculated in Matlab on the basis of [20].

For each sample we omit 5% of signal from both endings of file just to be sure it contains valid signal. Remaining signal was divided equally in to five parts and for each of them we calculated total set of parameters. They were named as SET 1, SET 2 and so on. For the analysis we've chosen SET 1 which we believe is the most unstable and SET 3, which we believe that it is the most stable part of the signal in the context of parameters fluctuations.

IV. The Results

As mentioned earlier, the study consisted in determining advanced voice parameters for acquired voice samples in the initial and middle part, and then on observing general trends.

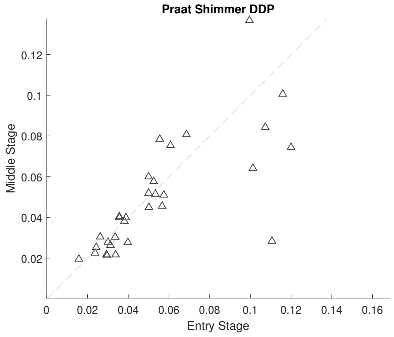

Figure 2. SET1 to SET3 Shimmer DDP relation - singer s34f

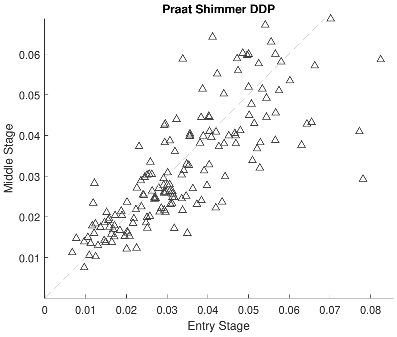

Figure 3. SET1 to SET3 Shimmer DDP relation – all females

Table I
THE TABLE OF 'IMPROVEMENT' FOR ALL RECORDED SINGERS

	Better	Worse	Unchanged
Praat jitter (local)	666	369	16
Praat jitter (local absolute)	669	372	10
Praat jitter (rap)	579	452	20
Praat jitter (ppq5)	630	404	15
Praat Jitter DDP	579	452	20
Praat shimmer (local)	642	372	37
Praat shimmer (localdB)	654	357	40
Praat shimmer (apq3)	621	407	23
Praat shimmer (apq5)	637	371	40
Praat shimmer (apq11)	643	356	46
Praat Shimmer DDP	621	407	23
HNR35-mean	437	602	12
HNR35-std	588	439	24
SPR	627	389	26

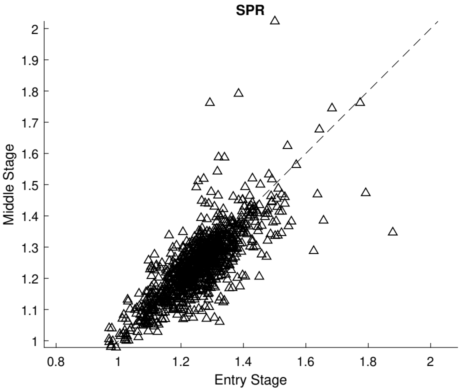

Figure 4. SPR for all singers

One of the estimated parameters was the average absolute difference between consecutive differences between the amplitudes of consecutive periods (Shimmer DDP).

The shimmer DDP, similarly to other measures of this class, indicates the magnitude of changes in amplitudes of appropriate periods and can be identified with voice tremors based on small but high frequency changes in volume. This phenomenon adversely affects the quality of the voice produced. In order to check whether the SET3 presents the improvement of the DDP parameters in relation to the first set, the values from both sets were compared. In the presented figure 2, prepared for the example singer s34f, most of the Shimmer DDP values, shown in relation SET1 to SET3, are below the line determining the lack of differences/changes what can it mean, in that particular case, that the voice quality may be improved.

What is the situation for all female voices in the context of this particular parameter?

Also in the case of the Figure 3, there is a tendency to decrease values of irregularity parameters. Next, a simple statistic is presented, which consists of determining the number of 'improved' values for the analysed parameters. Table I contains the values obtained for all individuals taking part in the study.

It's shows that an improvement in quality may be observed. Particular attention should be paid to the SPR parameter, which is the ratio indicating ratio of two different formants in the Long Term Average Spectrum in the analysed signal and is understood as an important quantitative measurement for evaluating singing voice quality for all voice types. It can be seen that this ratio improved in most of the cases. At the same time. Most of the jitter and shimmer values have been decreased, and what is more, values of the standard deviation for HNR35 have decreased. This should be considered as a simple indicator which shows that a stable part of the signal represents a signal which may have better vocal quality.

As the next, we present an 'improvement graph' for SPR values. Looking at the figure 4, it can be seen that in this case above 60% of the values has been improved. The above figures and tables show uncritical statistics and comparisons, estimated for all samples in the database. However, samples that were sung at the limit of a singers abilities should be excluded and statistics should be set again. It will also be interesting to identify changes in quality parameters only for the most comfortable samples (sung at the very centre of the vocal range).

A. *Analysis for the comfort vocal range*

An attempt was made to look at the data from which singing samples were excluded on the border of the possibilities of particular people. In the scope of sung samples, 3 samples were removed from the bottom of the range and 3 from the top of the range to analyse the range of comfortable sounds for

Table II
IMPROVEMENT OF THE QUALITY PARAMETERS FOR THE COMFORT VOCAL RANGE

	Better	Worse	Unchanged
Praat jitter (local)	555	303	26
Praat jitter (local absolute)	537	276	71
Praat jitter (rap)	486	394	4
Praat jitter (ppq5)	521	335	28
Praat Jitter DDP	486	394	4
Praat shimmer (local)	541	311	32
Praat shimmer (localdB)	553	303	28
Praat shimmer (apq3)	521	350	13
Praat shimmer (apq5)	539	309	36
Praat shimmer (apq11)	544	296	43
Praat Shimmer DDP	521	350	13
HNR35-mean	379	499	6
HNR35-std	496	370	18
SPR	525	331	22

Table III
IMPROVEMENT OF THE QUALITY PARAMETERS FOR THE MIDDLE VOCAL RANGE

	Better	Worse	Unchanged
Praat jitter (local)	240	113	17
Praat jitter (local absolute)	234	106	30
Praat jitter (rap)	206	158	6
Praat jitter (ppq5)	222	141	7
Praat Jitter DDP	206	158	6
Praat shimmer (local)	235	113	22
Praat shimmer (localdB)	238	113	19
Praat shimmer (apq3)	232	131	7
Praat shimmer (apq5)	228	121	21
Praat shimmer (apq11)	233	123	14
Praat Shimmer DDP	232	131	7
HNR35-mean	162	204	4
HNR35-std	221	140	9
SPR	216	143	10

singers. The answer was to find out if the presence of samples in the recordings that are not comfortable for singers will have a significant effect on the overall picture of the results obtained.

The results presented in the table II show that the overall picture of the whole has not changed - all ratios have improved mostly. In order to determine the influence of border samples on the result correction, we determined the improvement factors for both situations as the ratio of corrected values to all samples (for all parameters) and compared with the results after discarding border samples. There has been a slight improvement here, which does not mean that it does not matter. In both cases a factor of over 60% was obtained, which indicates that the extreme samples do not affect the results. This may be due to the fact that the recordings paid attention to the comfort of singing and interrupted the session at the moment when singing a certain pitch of sound made it difficult.

B. Analysis for the middle of vocal range

In the next scenario, samples from the middle of the vocal range were selected for the study and shown in table III. In this scenario, the results increased on average by a few percentage points. Some parameters showed improved values in 70% of cases. This shows that while recording of the samples is worth to determine the most comfortable sounds to sing and choose for testing those from the centre of the vocal ranges.

C. Analysis for the corresponding frequencies

In the next scenario, samples from the middle of the vocal range were selected for the study and shown in table IV. The quality coefficients aggregated for selected frequencies finally confirm the hypothesis that the central part of the recording is more valuable for the analyses.

Among many qualitative parameters estimated for the recorded samples and aggregated for particular frequencies sung by the surveyed persons, for the presentation jitter (local, absolute) was chosen. As it was mentioned before, jitter is a measure of frequency instability. Differences of the values of that parameter gives us information about differences of the quality of the signal. In the presented example, the jitter parameter (mean value) decreased in the case of the middle segment in the case for most of the investigated sung frequencies (additionally standard deviation of the parameter also decreased) so it should be considered as the final confirmation of the hypothesis that the central part of the sample, associated with the sustain phase, presents a signal with greater stability and thus better quality.

V. CONCLUSION

The article in general concerns the subject of signal analysis and is focused on the analysis of the quality of the signal representing the voice, in particular the voice of the singers. In the process of analysing of the voice quality, as with any signal, it is important to have samples whose content faithfully reflects the examined features to the maximum. In the case of singing samples analysis, the specifics of generating this signal should be taken into account. The singing signal characteristics can be described in an approximate way using the ADSR model. It indicates that in the initial phase of voice production and in the final phase, physiological phenomena occurs (reflected in ADSR by the attack and decay phases), which may affect the analysed features. In the analysis the middle part of the stable phase should be taken into account. The problem starts when the samples are too short. Very often it happens that the people being recorded try to shorten the sung phrase. When the analysed recording is too short, the impact of the attack phase becomes noticeable, as documented in this article. All values of the analysed signals, for the most of the samples, indicated higher signal quality in the sustain phase. This is best seen in the case of the quality coefficients aggregated for frequencies. Additionally it was confirmed that samples from the middle of the vocal range are those the best reflecting the voice of the singer.

The results concerning voice quality analysis presented here may be useful for constructing a singing quality assessment system. A large number of the results obtained for this study

Table IV
QUALITY COEFFICIENTS AGGREGATED FOR FREQUENCIES - PART OF THE RESULTS - JITTER (LOCAL, ABSOLUTE)

Sung Frequency [Hz]	Stage	Mean value	Standard Deviation	Percentile 25%	Median	Percentile 75%
146.8324	Entry	3.2081e-05	1.1301e-05	2.2417e-05	3.0432e-05	4.2687e-05
	Middle	2.4737e-05	8.3209e-06	1.7707e-05	2.2877e-05	2.8792e-05
155.5635	Entry	2.2129e-05	7.1437e-06	1.6498e-05	2.2769e-05	2.4551e-05
	Middle	1.8161e-05	7.3445e-06	1.2832e-05	1.6569e-05	2.3387e-05
164.8138	Entry	1.8351e-05	7.3706e-06	1.2684e-05	1.8483e-05	2.1523e-05
	Middle	1.5216e-05	7.1313e-06	1.0105e-05	1.2216e-05	1.8996e-05
174.6141	Entry	1.7062e-05	7.4216e-06	1.1709e-05	1.4635e-05	2.1104e-05
	Middle	1.5071e-05	7.8123e-06	9.859e-06	1.2466e-05	1.6579e-05
184.9972	Entry	1.3369e-05	6.2646e-06	9.6237e-06	1.1677e-05	1.4684e-05
	Middle	1.1412e-05	4.0202e-06	8.406e-06	9.9462e-06	1.4236e-05
195.9977	Entry	1.1873e-05	4.703e-06	8.4799e-06	1.3474e-05	1.5916e-05
	Middle	1.0632e-05	6.1637e-06	6.4303e-06	8.584e-06	1.2656e-05
207.6523	Entry	7.6085e-06	3.5426e-06	5.4417e-06	6.8487e-06	8.2179e-06
	Middle	6.6576e-06	4.2819e-06	3.6837e-06	5.8339e-06	7.4392e-06
220	Entry	5.8739e-06	1.5366e-06	4.6282e-06	6.1168e-06	7.1197e-06
	Middle	4.9587e-06	1.6141e-06	3.6099e-06	5.1528e-06	6.3075e-06
233.0819	Entry	6.653e-06	4.0259e-06	4.6118e-06	5.1807e-06	6.7364e-06
	Middle	5.5737e-06	3.6728e-06	3.6742e-06	4.5198e-06	5.4754e-06
246.9417	Entry	6.9293e-06	4.3558e-06	4.9796e-06	5.2084e-06	6.693e-06
	Middle	7.3143e-06	5.343e-06	3.0391e-06	6.4562e-06	8.0467e-06
261.6256	Entry	7.9658e-06	3.076e-06	6.03e-06	8.0923e-06	9.9593e-06
	Middle	6.3619e-06	4.8693e-06	3.9123e-06	4.3487e-06	6.9787e-06
277.1826	Entry	5.3634e-06	1.7809e-06	4.233e-06	5.465e-06	6.4937e-06
	Middle	6.2812e-06	3.2574e-06	3.6387e-06	6.5349e-06	8.9237e-06
293.6648	Entry	4.6625e-06	1.4611e-06	3.6059e-06	4.4312e-06	5.7768e-06
	Middle	7.5257e-06	5.2186e-06	3.7047e-06	6.9339e-06	1.1495e-05

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 from the West Pomeranian University of Technology who have devoted their time to recording voice samples. Thanks are also due to the experts in the field of voice emission, especially for Mr. Adam Kuliś from Study of Culture, for substantive consultations.

REFERENCES

[1] E. Półrolniczak and M. Kramarczyk, "Analysis of the signal of singing using the vibrato parameter in the context of choir singers," *Journal of Electronic Science and Technology*, vol. 11, no. 4, pp. 417–423, December 2013.

[2] J. Godino-Llorente, P. Gomez-Vilda, and M. Blanco-Velasco, "Dimensionality reduction of a pathological voice quality assessment system based on gaussian mixture models and short-term cepstral parameters," *Biomedical Engineering, IEEE Transactions on*, vol. 53, no. 10, pp. 1943–1953, Oct 2006. doi: 10.1109/TBME.2006.871883

[3] K. Omori, A. Kacker, L. M. Carroll, W. D. Riley, and S. M. Blaugrund, "Singing power ratio: Quantitative evaluation of singing voice quality," *Journal of Voice*, vol. 10, no. 3, pp. 228 – 235, 1996. doi: http://dx.doi.org/10.1016/S0892-1997(96)80003-8. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0892199796800038

[4] Y. Meron and K. Hirose, "Separation of singing and piano sounds." in *ICSLP*, 1998.

[5] 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.

[6] 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.

[7] L. Mazurowski, "Computer models for algorithmic music composition," in *Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on*. IEEE, 2012, pp. 733–737.

[8] K. Jensen, "Envelope model of isolated musical sounds," in *Proceedings of the 2nd COST G-6 Workshop on Digital Audio Effects (DAFx99)*, 1999.

[9] R. M. Alderson, *Complete handbook of voice training*. Parker Publishing Company, 1979.

[10] M. Łazoryszczak and E. Półrolniczak, "Audio database for the as sessment of singing voice quality of choir members," *Elektronika: konstrukcje, technologie, zastosowania*, vol. 54, no. 3, pp. 92–96, 2013.

[11] E. Półrolniczak 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.

[12] E. Półrolniczak 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.

[13] 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. ISSN 2326-0262 pp. 200–204.

[14] 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.

[15] E. H. Buder, "Acoustic analysis of voice quality: A tabulation of algorithms 1902–1990," *Voice quality measurement*, pp. 119–244, 2000.

[16] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [Computer program]," Version 6.0.43, retrieved 8 September 2018 http://www.praat.org/, 2018.

[17] Y. Jadoul, B. Thompson, and B. de Boer, "Introducing Parselmouth: A Python interface to Praat," *Journal of Phonetics*, vol. 71, pp. 1–15, 2018. doi: https://doi.org/10.1016/j.wocn.2018.07.001

[18] Y.-L. Shue, P. Keating, C. Vicenik, and K. Yu, "Voicesauce," *p. Program available online at http://www. seas. ucla. edu/spapl/voicesauce/. UCLA*, 2009.

[19] A. De Cheveigné and H. Kawahara, "Yin, a fundamental frequency estimator for speech and music," *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1917–1930, 2002.

[20] E. Yumoto, W. J. Gould, and T. Baer, "Harmonics-to-noise ratio as an index of the degree of hoarseness," *The journal of the Acoustical Society of America*, vol. 71, no. 6, pp. 1544–1550, 1982.