Modification of notes recognition algorithm with a sharp change of the pitch frequency

A Yu Yakimuk¹, E S Kataeva¹, M N Golovchiner²

¹Tomsk State University of Control Systems and Radioelectronics, Tomsk, Lenin str., 40, 634050, Russia
²Tomsk State University, Tomsk, Lenin str., 36, 634050, Russia

E-mail: yay@keva.tusur.ru, kes@keva.tusur.ru, golovchiner@mail.ru

Abstract. The paper presents the results of signal processing using the simplest methods of time series analysis to localize the main note in the presence of vibrato or ascending / descending glissando. The idea of using smoothing to find the vibrato effect is to evaluate segments of the speech signal, within which it was not possible to identify the note. Data on frequencies in this segment are averaged to localize the main note, relative to which the deviation in sound and the frequency of changes are determined in percentage terms. If the deviation from the main note is more than half a tone, the frequencies are perceived as noise.

1. Formulation of the problem

Application software in learning is common in many areas of knowledge. In the field of automatic speech processing, specialized programs can be used in the context of speech synthesis or recognition for signal dynamics studying or keywords determination. For example, in [1-2], the assessment of word intelligibility is carried out as part of solving the problems of speech rehabilitation after complex treatment of cancer. The use of specialized programs can help in teaching foreign languages or in performing exercises to develop vocal skills. However, in the field of vocal training, existing programs do not always accurately determine the note sung by the performer. The biofeedback effect is useful in the field of musicians training [3], which leads to the need to develop a music recognition system with high accuracy.

The developed software package includes an algorithm notes recognition [4]. The principle of operation of the algorithm for recognizing notes in vocal performance, used in the software package, is given below. The value of the pitch frequency is determined for each time point, so it is possible to determine which note "was sung" at the time countdown. Determination of the pitch frequency is performed by an algorithm based on a modification of the mathematical model of the human auditory system which is shown in [5]. Preliminary testing of software system has shown that improving the quality of the music identification is possible with the expansion of the range covered by the frequency and identifying situations that reduce the accuracy of the algorithm.

The tests of the algorithm showed that the software complex allows you to recognize more than 95% of notes sung by the singer. Testing the software package consisted in processing audio recordings that include various approaches to singing notes in the range from 70 to 800 Hz. It was determined that there is no effect on the algorithms quality in the software package [6] while singing arpeggios, crescendos and decrescendos. The absence of influence on the notes identification quality lies in the fact that the analysis takes into account only the fundamental tone frequency. It is important to note that one of the
stages of the musical notes identification algorithm is focused on pure performance. This stage is responsible for determining whether a vocalized area belongs to a note. At this stage, the obtained sets of fundamental frequencies are compared in duration with the minimum duration of possible notes. If the duration of the fragment is insignificant, this fragment will be eliminated.

It is quite difficult to play a note so that all frequencies at the time of singing are within the band assigned to it. For this reason, adjacent notes above and below the note are counted. In figures 1-3, the frequency spectrum is divided into sections corresponding to the boundaries of notes. Each section is highlighted by a horizontal strip on the background of the frequency graph. On the graph, the axes are indicated as follows: along the abscissa axis is time ($t$), along the ordinate axis are the fundamental tone frequencies ($F_0$).

As can be seen in Figure 1, the segment sung within the same note was recognized by the program. The number of time points that fell into neighboring notes was insignificant and did not affect the identification result. The predominant number of recognized fundamental tone frequencies was in the range of the sung note. Given the requirements for accuracy in the algorithm, the program was able to conclude a sung note. If the number of fragments in each of the 3 sections turned out to be approximately the same, the algorithm perceives the entire voiced segment as noise.

![Figure 1](image1.png)

**Figure 1.** A segment with a note sung without vibrato in the singer's voice.

It was determined that errors arise for situations with approaches that imply a change in the value of the sound frequency during the performance: vibrato (Fig. 2) and glissando (Fig. 3).

![Figure 2](image2.png)

**Figure 2.** A segment with a note sung with vibrato in the singer's voice.

When vibrato is singing, fluctuations in the frequency of sound relative to some main note occur within a semitone [7-8]. It should be noted that there are varieties of vibrato-like fluctuations. This
includes tremolation and "voice swing". Due to the fact that the algorithm is configured to determine pure notes, such segments are perceived as noise. As you can see, the oscillations during the execution of the main note (in this case, the note “E of the small octave”) immediately cover 4 notes. Taking into account the minimum duration of note hold leads to the fact that each of the 4 notes was singing within its boundaries during an insufficient amount of time. In this case, a return to this note occurs after switching to other notes which reduces the proportion of frequencies related to the main note.

![Figure 3. A segment sung with ascending glissando in singing.](image)

As is known from the theory of music [9], when glissando is sung, there is a smooth glide from one note to another. Because of such singing, the transition is too fast to be able to identify each individual note at the moment of sliding, since less than 0.1 seconds are allocated to each segment of the covered notes. Glissando can be either upward (as shown in Figure 3) or downward. As you can see in the figure, the algorithm is able to determine the starting and ending notes between which the slide is performed. However, the coverage of 12 notes in between is perceived as noise.

Thus, there is a need to identify areas perceived by the program as noise. We assume that a directed transition from one note to another is glissando, and fluctuations with respect to one note within a single segment are tremolation. At this stage, the determination of the type of tremolation will not be considered.

2. Analysis of the performance of notes with a sharp change in the fundamental tone frequencies
Since the identification algorithm gives us the time series of the fundamental tone frequencies, within which it was not possible to identify a clean note, it was decided to conduct a preliminary data analysis. Among the preliminary analysis procedures, there are: search for anomalies, time series smoothing, checking for a trend and process dynamics calculating.

From the point of view of the subject area, abnormal observations can perceive bursts of frequencies determined outside the sound of the main melody. This includes any noise screened by algorithms at the stage of note segmentation. Accordingly, we can consider this procedure as completed.

As part of the time series smoothing, the true levels of the series are replaced with calculated values that have smoother dynamics than the original data. It was decided to focus on the methods of mechanical smoothing, since in the problem under study it is necessary to evaluate each individual time series values taking into account the actual adjacent values. The weighted moving average method is not applicable for the problem under study, since it cannot contain a quadratic or cubic trend. There is also no need to use the exponential smoothing, because it is used in the tasks of predicting the
development of the process after the study area. In this regard, the simple moving average method will be used. As the number of observations \( P \) included in the smoothing interval, it was decided to use the value of the minimum note duration used in the segmentation and note identification algorithm. The choice is due to the fact that when singing vibrato and glissando, the duration of the singing within such sections is significantly higher than this parameter, but the identification of individual clear notes was not carried out by the basic algorithm. A disadvantage of this method is that the first and the last \( P \) observations remain unsmoothed but we can neglect that fact because of the estimated total duration of portion.

Checking for the presence of a trend is a hypothesis test that the average value of the time series does not change. For verification, the series will be divided into \( n \) parts, and each of these parts is considered as a separate time series. The resulting averages for each part will be compared. If the average values increase or decrease, we assume that the studied segment may contain glissando (ascending or descending, respectively).

2.1. Analysis of notes sung with vibrato in the voice

Note that all calculations are performed in Microsoft Office Excel. The built-in functional contains the “Data Analysis” function, in which a moving average \([10]\) can be built for the selected series. In the built-in method, only the previous values are analyzed without taking into account the subsequent values, which allows you to get an envelope that is not informative as part of research.

In order to solve this problem, the classical formula for sliding smoothing was used taking into account the measurements before and after smoothing, \( 0.5*P \) values preceding the smoothing and the same number of subsequent values were included in the smoothing interval (Fig. 4). As a result, in the interval from \( 0 \) to \( 0.5*P \), the values of the series will coincide, but we can see the averaging from the next value. The solid line indicates the values of the analyzed fundamental tone frequencies, and the dashed oscillating - the smoothed values.

![Figure 4. Moving average for vibrato-like singing chart.](image)

In the presented figure, for the note “E of the small octave”, parallel dashed lines indicate the lower (160.121 Hz) and upper (169.643 Hz) sound boundaries determined using the average logarithmic. Notably that judging by the schedule in the studied segment there are 7 oscillations in the frequency of the fundamental tone frequency. In a further study of the type of vibrations in singing this aspect can be applied as a criterion for the difference between vibrato and tremolation.

As mentioned earlier, the presence of a trend is determined by comparing the obtained average values for each of the sets of the studied series. Also, using Excel tools on a graph of the fundamental tone
frequencies, you can specify the construction of a linear trend (Fig. 5). At the stage of analyzing the method applicability to the studied problem we will use the built-in function in Excel, and at the implementation stage in the software package we will compare the average values for the segments of the studied range.

![Figure 5. Linear trend for vibrato-like singing chart.](image)

As can be seen from the figure above, the linear trend allows us to determine that the average frequencies refer to one note in most of the studied area. However, according to a linear trend, it is impossible to assess the presence of oscillations in the fundamental tone frequencies. In this regard, it is not possible to use only a linear trend to assess areas with tremolation. The trend allows you to assess the number of notes covered in the study area, but not the type of performance. In turn, the moving average simultaneously interprets the presence of vibrations and the covered notes, which makes this assessment method more preferable for vibrato-like sections.

2.2. Analysis of notes sung with glissando in voice

To determine the glissando in the study area you must perform another task. The basic algorithm for segmentation and note identification determines 2 notes in similar situations: the one for beginning of the slide, and the one for the end. Accordingly, we consider a comparison of notes at the boundaries of the investigated area as a first task to be solved for determining the presence of a glissando. If the notes turn out to be identical, then the area between them cannot correspond to the glissando and should be considered for attribution to vibrato. In other cases, we will observe a trend line passing through several notes. The greater the difference between the notes and the faster the transition between them was made, the greater the angle of the obtained trend line relative to the time axis.

![Figure 6. Moving average for chart with ascending glissando in singing.](image)

As for the plot with tremolation, we consider the moving average for the studied range. As you can see in Figure 6, in the transition from the note “G of a small octave” to the note “E of the 1 octave” there
are no obvious fluctuations in the fundamental tone frequency. The solid line on the graph indicates the fundamental tone frequencies, and the dashed line indicates the moving average for the studied range. Parallel dotted lines correspond to the boundaries of notes. The resulting moving average line goes through 9 notes. Thus, this result indicates an ascending glissando. Below (Fig. 7) is a graph for a linear trend for the studied range. In fact, the linear trend for the studied range is close to the result obtained with constructing the moving average.

![Graph showing glissando](image)

**Figure 7.** Linear trend for a chart with upward glissando in singing.

3. **Analysis of the results**

Analysis of the graphs of estimates for time series with tremolation in singing showed the following results. As you can see in the figures with the estimation of the moving average, the areas with tremolation are characterized by an oscillatory appearance for this graph. You can also notice that the linear trend does not fluctuate at the indicated moments. For glissando, on the other hand, the moving average and linear trend behave in a similar way. This allows us to apply both methods comprehensively to determine the current situation.

The proposed idea is to use both estimates for the studied segment. The obtained data sets will be compared for the similarity between their estimates for a certain parameter N. Determination of the value of this parameter must be studied separately. In this experiment, to reduce the number of errors of the 1st and 2nd kind when determining the type of the investigated ranges of the fundamental tone frequency we consider mixed situations in which time series behave ambiguously.

If there was an effect of glissando in the singing, the results of application of the moving average and linear trend for this time series will be similar. For this situation, it will be necessary to check for the number of notes through which the transition was made, and the direction of the transition (ascending or descending glissando). As a result, it will be possible to filter out areas with noise between adjacent segments corresponding to one note, and areas with a burst of noise that is outside the range of sound between notes.

For situations in which the difference in the estimates of the time series will exceed the value of the criterion N, we assume that there was singing with effect of vibrato. As previously noted, the moving average has a wavy structure in the case of oscillations in the studied area. Counting the number of vibrations per unit time can be used in the classification problem of tremolation type. This parameter will be useful in teaching vocal performance in case of too frequent or rare fluctuations in the performance of a note. In addition, in the case of determining vibrato in singing, it will also be necessary to control the number of notes through which the trend passed.
As you can see in Figure 8, the vibrations for the sung note pass with a slide down. In fact, in this example, glissando and vibrato are mixed. However, this section cannot be attributed to any of the types for two reasons: for the glissando the presence of vibrations in the voice is not characteristic, and for the vibrato there should be no sliding on another note.

Acknowledgments
This research was funded by the Ministry of Science and Higher Education of Russia, Government Order for 2020–2022, project no. FEWM-2020-0037 (TUSUR).

References
[1] Bondarenko V P, Choinzonov E T, Balatskaya L N et al. 2007 Biomedical Engineering 41 pp 154–156
[2] Kostuchenko E, Novokhrestova D, Pekarskikh S, Shelupanov A et al. 2019 SPECOM 2019: Speech and Computer 11658 pp 359-369
[3] Leydon C, Bauer J J, Larson C R 2003 Acoustical Society of America 114(3) pp 1575-1581
[4] Konev A, Kostyuchenko E, Yakimuk A 2017 Journal of Physics: Conference Series 803(1) 012077
[5] Konev A A, Meshcheryakov R V and Kostyuchenko E Y 2018 Optoelectronics, Instrumentation and Data Processing 54 pp 361–366
[6] Yakimuk A Yu, Konev A A, Andreeva Yu V, Nemirovich-Danchenko M M 2019 IOP Conference Series: Materials Science and Engineering 597(1) 012072.
[7] Reddy A, Subramanian 2015 Journal of Voice 29(5) pp 603–610
[8] Michel C, Ruiz M 2017 Physics Education 52(4) pp 1-6
[9] Deutsch D and Hamaoui K 2005 Acoustical Society of America 117 2476
[10] Canela M Á, Alegre I, Ibarra A 2019 Quantitative Methods for Management pp 111-119