

Speech Spectral Transfer Function

Oleg N. Korsun¹, Evgeny.I. Mikhaylov^{2*} and Magomed Z. Nakhaev²

¹FSUE “State Research Institute of Aviation Systems”, Moscow, Russia

²Moscow Institute of Physics and Technologies (State University), Moscow, Russia

Abstract. The article discusses the problem of estimating the psychophysiological state of aircraft pilots by their speech. For this purpose, a new concept of an individual speaker’s transfer function is proposed. This definition is based on the classical results of automatic control theory. The article presents the algorithms for calculating transfer function and the examples of using this feature for medical purposes.

1 Statement of the problem

Today man-machine interfaces based on various physical principles are intensively developed in aviation [1]. For example, the main trends of audio interfaces’ development are the 3D-audio (surround sound effect) and automatic speech recognition [2], which are used for controlling onboard systems [3, 4].

It is known, that operator’s speech characteristics depend on outer conditions and his own psychophysical state. Estimation of these changes is the key task for improving man-machine interface. There is a number of articles, where this problem has been investigated. For example, the impact of aircraft overload on operator’s speech characteristics is presented in [3]. The connection between the degree of operator’s fatigue and parameters of his speech is described in [4]. This relation is based on the A. M. Lyapunov’s theory of stability [5,6]. The influence of the acoustic noises on the speech recognition performance score is analysed in papers [7,8]. The speech recognition algorithm, resistant to the noise, correlated with the voice signal, is proposed in [8]. The paper [9] discusses the speech characteristics of pilots with hearing loss diagnosis. This problem is particularly relevant to the helicopter aviation pilots. All the above mentioned researches are held in order to improve speech recognition algorithms. But on the other side, speech characteristics may be used to estimate operator’s psychophysical state and the influence of various factors on it.

Evidently, the principal changes in the speech signal occur in the frequency and time domains [2, 3, 4, 8]. The analysis of these changes in absolute values is coupled with the problem of representing the information because, for example, the power of speech signal in different frequency ranges varies by decades of decibels [2, 7]. So, the analysis of speech characteristics’ changes should be carried out not in absolute, but in relative terms. It means, that the speech signal should be compared with a relevant one, for example, with the

speech of another operator, considered as a reference, or standard.

To form a reference it is possible to use the mean value for the group of speakers. Another way is to choose a speaker, whose pronunciation is very close to the norms of the literature language. In this paper the concept of the speech transfer function is proposed in order to analyze the changes of operator’s speech characteristics.

2 The algorithm for calculation transfer function of the operator

Let us introduce the speech transfer function of the operator, that is similar to transfer function $W(p)$, well-known in the theory of automatic control [6]. This function is defined as the ratio of the Laplace transform of the input and output signals for zero initial conditions [6].

We arrive at a transfer function $W(f)$, which depends on the frequency f , Hz, applying to the Laplace variable p [6] the substitution

$$p = j2\pi f,$$

where j – imaginary unit, that is $j^2 = -1$.

For an arbitrary value of argument f function $W(f)$ is a complex number containing information about the signal change in amplitude and phase on the frequency f . As it is known, an operator perceives the amplitude of a speech signal, while the phase information is ignored. So, we consider only the absolute value $|W(f)|$. Let us call it *the speech transfer function of the operator in frequency domain*. According to the Wiener–Khinchin theorem [6], the following ratio holds for the absolute value of the speech transfer function

$$|W(f)| = \sqrt{s_y(f)/s_x(f)}, \quad (1)$$

where $s_y(f)$, $s_x(f)$ – spectral densities of the input and output signals.

* Corresponding author: mihevkonig@mail.ru

To form a practical algorithm for calculating estimates (1) let us use a parametrization algorithm, widely known in automatic speech recognition [2, 8]. According to this algorithm, the speech fragment, for example, a recorded word, is divided into N_t time intervals (frames). The duration of each frame is 20 ... 40 ms. Then the Hann spectral window and fast Fourier transform are applied to each interval [2].

After that an absolute value of the Fourier transform is calculated for each frame. Then the whole frequency range, limited by Nyquist frequency [6], is divided into a predetermined number of bands $N_f = 20 \dots 40$, and the average absolute value (the square root of the spectral density's estimation) is calculated for each band, as it is necessary for the formula (1). As a result, we obtain a matrix of the word's parametric portrait:

$$\mathbf{X} = \{x_{ij}\}; i = 1, \dots, N_f, j = 1, \dots, N_t, \quad (2)$$

which dimension is $N_f \times N_t$; each j -column describes the spectral content of the speech signal for the j - frame; each i -row describes a change in time of the mean absolute value of the signal components belonging to the i -frequency band.

It is known, that the error of the estimates, calculated by the formula (1), is considerable [2,6]. To improve the accuracy of estimation we will increase the amount of data. We take $M = 10 \dots 30$ realizations for each word and determine the matrix of the mean parametric portrait

$$\bar{\mathbf{X}} = \frac{1}{M} \sum_{k=1}^M \{x_{ij}\}_k, \quad (3)$$

where $\bar{\mathbf{X}}$ – matrix of dimension $N_f \times N_t$ of the operator's mean parametric portrait.

The above mentioned parametrization algorithm is a standard procedure of the speech recognition theory [2, 8]. When analyzing the general change in the spectral properties of speech, the time domain quantification is not necessary, because we do not need to take into account the features of sounds or syllables. So let us calculate the mean over all time intervals, i.e., over elements of the matrix's $\bar{\mathbf{X}}$ rows, and obtain the vector \mathbf{a} of the mean amplitudes of the frequency components, belonging to the i -frequency band

$$\mathbf{a} = \frac{1}{N_t} \sum_{j=1}^{N_t} \{\bar{x}_{ij}\}, i = 1, \dots, N_f, \quad (4)$$

where \bar{x}_{ij} – the elements of the mean parametric portrait matrix (3).

In order to find estimates of the transfer function between two speakers it is necessary to calculate matrices $\bar{\mathbf{X}}_1$ and $\bar{\mathbf{X}}_2$, vectors \mathbf{a}_1 and \mathbf{a}_2 for each frequency band and for each speaker's speech data. Then the formula (1) is to be applied.

$$|W_{1,2}(f_i)| = \frac{a_{2i}}{a_{1i}} = \frac{1}{N_t} \sum_{j=1}^{N_t} \{\bar{x}_{2ij}\} / \sum_{j=1}^{N_t} \{\bar{x}_{1ij}\}, i = 1, \dots, N_f, \quad (5)$$

where f_i – frequency corresponding to the middle of the i -frequency band; a_{1i} , a_{2i} – elements of vectors \mathbf{a}_1 and

\mathbf{a}_2 , corresponding to the i -frequency band; \bar{x}_{1ij} , \bar{x}_{2ij} – elements of matrices $\bar{\mathbf{X}}_1$ and $\bar{\mathbf{X}}_2$ of the mean parametric portraits.

In (5) indices 1 and 2 denote respectively the first and second speakers.

The transfer function between the different states of the same speaker is determined by the same way.

The values of the transfer function (5) are expressed in dB as follows [6]:

$$|W_{1,2}(f_i)| = 20 \lg \left(|W_{1,2}(f_i)| \right) = 20 \lg \left(\frac{a_{2i}(i)}{a_{1i}(i)} \right), \\ i = 1, \dots, N_f. \quad (6)$$

The results of estimation will be improved, if the word is preliminarily divided into few parts in accordance with the algorithm [10].

3 Speakers in noisy conditions

Let us consider the applications of speech transfer functions in order to study the effects of noise on the speaker's speech. For this purpose the following experiment has been carried out. The noise recorded in the cockpit during the flight was fed only into the speaker's headphones, so as not to interfere with the record of the speaker's words. The registration frequency was 22 kHz. During the experiment audio data was recorded in three conditions: without noise; with noise in headphones 80 and 90 dB. Figure 1 shows plots of the speech transfer function for three speakers in noisy conditions (80 and 90 dB in the headphones). Transfer functions were calculated for the same speaker and the record without noise was taken as a reference.

The analysis of results shows that the impact of 80 dB noise causes a significant increase in the amplitude over the entire frequency range, with the largest rise in the range 1 ... 4 kHz, followed by decline in the range of high frequencies 4 ... 11 kHz. The maximum increase in amplitude is: for the first speaker 9 ... 10 dB with a decline to 3 ... 4 dB at high frequencies (figure 1a); second speaker 5.0 ... 6.5 dB with a decline to 3.5 ... 4.5 dB (figure 1b); third 4.5 ... 5.5 dB with a decline to 2.5 ... 3.5 dB (figure 1c). Noise augmentation from 80 to 90 dB increases the volume of speech by 1 ... 2 dB for all speakers. Thus, the proposed function allows to identify common and individual changes in the speech of speakers in noisy environment.

4 Speech transfer function in medical applications

Let us consider the use of the proposed function for investigating the speech of helicopter pilots with hearing loss diagnosis. A speaker without diseases of hearing and speech was chosen as a reference. Figure 2 shows plots of the speech transfer function for three speakers with a diagnosis of hearing loss with respect to the speaker without diseases of hearing and speech.

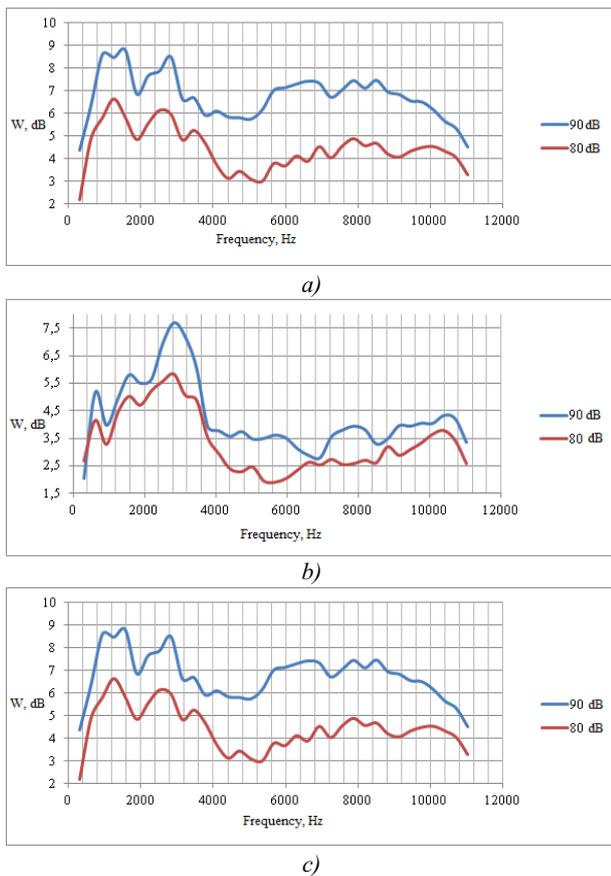


Fig. 1. Plots of speech transfer function for three speakers (a–c) with the noise in the headphone 80 (1) and 90 (2) dB.

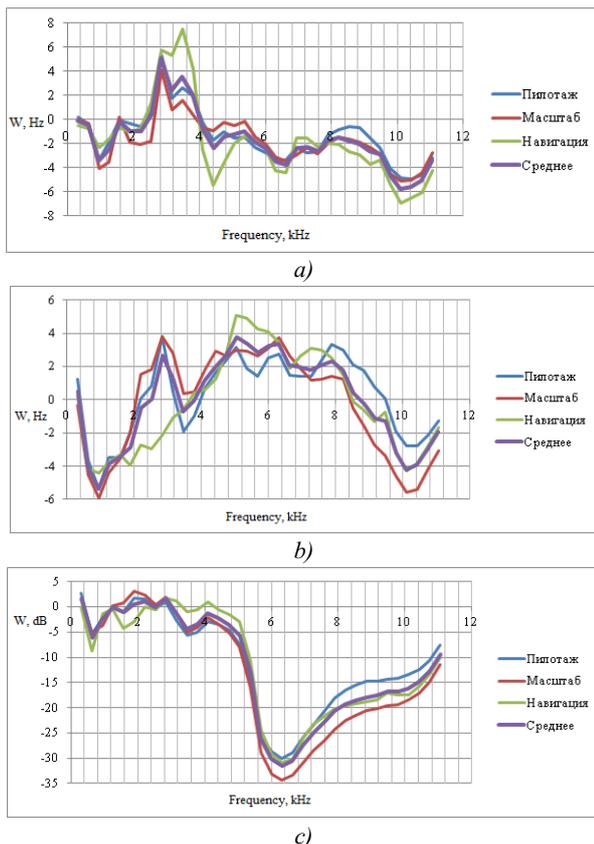


Fig. 2. Plots of speech transfer function for three speakers (a–c) with a diagnosis of hearing loss with respect to the speaker without diagnosed diseases of hearing and speech.

The plots show transfer function estimates calculated from records of the russian words "пилотаж", "масштаб", "навигация", corresponding to such english words as "pilotage", "scale", "navigation", and the mean ("среднее") between them.

Plots have obvious individual characteristics, but the common feature is the wide variation of values, which makes approximately ± 6 dB (figure 2a, b), and ± 20 dB (figure 2c).

The final experiment, which is to be discussed in this paper, is related to means for correcting teeth. Speech data of the speaker before installing corrective means was chosen as a reference. Then speech transfer function was calculated before and after installation of these means. Obtained results are presented in figure 3. This figure shows the significant changes in speech for the frequencies above 6 kHz.

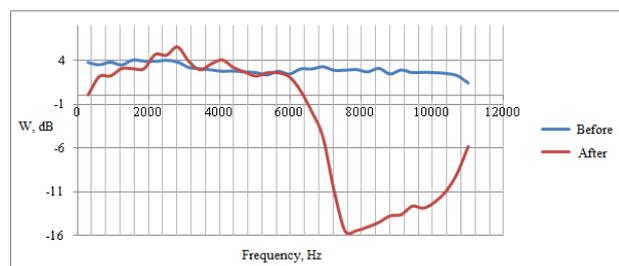


Fig. 3. Plot of speech transfer function for the same speaker before and after the installation of means for correcting the teeth.

5 Conclusion

In this paper we introduced the concept of speaker's speech transfer function in frequency domain in order to analyze the changes in speech characteristics and the influence of the outer conditions and psychophysical state of the speaker. The paper proposes the algorithm for calculating the estimates of speech transfer function, using experimental data. Some applications of the proposed function are also presented.

References

1. V.N. Evdokimenko., R.V. Kim, M.N. Krasil'shchikov, G.G. Sebyakov, *Journal of Computer and Systems Sciences International*, **54**(4), 609-620 (2015)
2. L. Rabiner, B.-H. Juang, *Fundamentals of Speech Recognition* (Prentice-Hall International, Inc., 1993)
3. A.I. Ivanov et al., *Vestnik komp'uternykh i informatsionnykh tekhnologii*, **6**, pp. 3-7, (2012)
4. Iu.G. Bondaros, A.I. Ivanov, A.A. Tishchenko, *Vestnik komp'uternykh i informatsionnykh tekhnologii*, **6**, 22-30, (2010)
5. K. Galkowski et al., *Journal of Computer and Systems Sciences International*, **55**(4), 503-514 (2016)

6. A.A. Erofeev, *Automatic control theory: textbook for universities*, (Moscow: Politekhnik, 2008)
7. A.S. Kolokolov, I.A. Liubinskii, *Automation and Remote Control*, **10**, 144-151, (2015)
8. O.N. Korsun, *Vestnik komp'iuternykh i informatsionnykh tekhnologii*, **4**, pp. 3-7, (2014).
9. A.I. Ivanov, *Meditcina truda i promyshlennaia ekologiya*, **11**, 40-45, (2014).
10. O.N. Korsun, A.V. Poliev, *Journal of Computer and Systems Sciences International*, **55**(4), 609-618 (2016)