Environment. Technology. Resources. Rezekne, Latvia
Proceedings of the 15th International Scientific and Practical Conference. Volume III, 300-305

Technical Obstacles to Speech Intelligibility

Georgi Stanchev
Department of Electronics, navigation and communication in the aviation
"Georgi Benkovski" Bulgarian Air Force Academy
Dolna Mitropoliya, Bulgaria
gstanchev@af-acad.bg

Abstract. **As voice communication increasingly rely on technological devices, understanding the intricate relationship between equipment properties and communication outcomes becomes imperative. The purpose of this study is to investigate the impact of technical condition of communication equipment on the intelligibility of voice communications. To achieve this goal, methods and results of a qualitative analysis of the voice signals were used, as well as statistical methods for establishing the relevant dependencies. The importance of technical characteristics of voice signals for speech intelligibility was considered. The influence of the technical parameters of electronic communication equipment and the working environment on the technical characteristics of voice signals and thus on the intelligibility of speech was discussed. On the basis of the research, conclusions were drawn highlighting the serious influence of the communication equipment characteristics and the expertise of the service personnel on speech intelligibility.**

Keywords: communication equipment, speech intelligibility, technical characteristics, voice.

I. INTRODUCTION

Humans are social beings – we feel most comfortable when we live in groups of different formats – families, working teams, friendship circles, interest clubs etc. And communication is the most essential part of our existence within these social formations. It is so important to us that we do our best to keep it, even when our direct presence in the social group is impossible. Then, technical means come to help, through which we can be heard at great distances or receive a message that is sent to us at a convenient time.

There are countless examples when technical means of communications enable us to accomplish a certain work activity – one such example is air traffic management. Here, the impact of the characteristics of the technical means used in aviation and air traffic control, in particular on the quality of voice communications and speech intelligibility, will be discussed.

The advantages and opportunities provided by the technical advancement in voice communications are undeniable. Of great importance is the fact that in this way other human channels for receiving information are freed from the load, especially the most important one – the eyesight.

Through technical radio devices, we communicate by voice, overcoming distances and time differences. However, it is important to consider, if the speech that passed through the processing of the equipment and then through the radio channel will pay a high price to reach far or fully retain its authentic form.

In-depth information and studies focused on individual aspects of the interrelationships explored here can be found in literary sources. These include investigations into the statistical dependencies between different types of speech intelligibility, examinations of the technical parameters and condition of communication equipment, as well as analyses of the spectral and amplitude profile of voice signals. However, information and research on the complex influence of technical parameters of the electronic communication equipment on speech intelligibility are difficult to find and scarce.

In order to consider the influence of the technical characteristics of electronic devices for voice communications, it is only natural to pay attention to the more measurable characteristics of speech.

Using a combined analytical and statistical approach, those characteristics of speech that would be influenced by the characteristics of the electronic devices carrying it can be determined.

II. MATERIALS AND METHODS

A. Technical Characteristics of Speech

Combined quantitative and qualitative analysis methods, as well as stochastic methods of Probability Theory, were used to explore sound signals as the basis of speech communications. These signals can be represented

Print ISSN 1691-5402
Online ISSN 2256-070X
https://doi.org/10.17770/etr2024vol3.8129
© 2024 Georgi Stanchev. Published by Rezekne Academy of Technologies.
This is an open access article under the Creative Commons Attribution 4.0 International License.

as complex random processes occurring in the medium between their source and receiver. Voice audio signals are changes in the pressure of the medium, most often air, which have different amplitudes and frequencies.

More precise definitions characterize these random processes with parameters that are common to sounds of any nature, such as: sound pressure level, sound level, dynamic range, frequency range, shape of the frequency spectrum, duration, echo and reverberation (for sound propagating in closed spaces) and others.

The purely physical characteristics of speech sounds can be objectively determined, measured and described. Variations in these descriptions may occur because of differences in the speaker's sex, age and speech apparatus, but may also be due to peculiarities of the language used for voice communications.

The propagation of sound in the air causes areas of condensation and rarefaction, in which the pressure is respectively higher or lower than its average value. Sound intensity level (***N***) can be described numerically as shown in (1), where ***N*** is a relative logarithmic measure of sound power flow density ***I*** relative to the standard value of I_0=10^{-12}W/m^2 (sound pressure p_0=20µPa), which corresponds to the lower audibility threshold of a sound wave with a sinusoidal shape at frequency of 1 kHz [1]. The dynamic range ***D*** defines the difference in the levels of the lowest intensity sounds I_{min} and the highest intensity sounds I_{max} (1), for sounds of a certain nature – as an example, normal conversation, industrial noise or office noise.

$$N = \frac{I}{I_0}, dB \quad ; \quad D = \frac{I_{max}}{I_{min}}, dB \qquad (1)$$

Another very inherent and informative characteristic of sound is its spectrum. Pure tones are almost never found in human speech – i.e. tones which can be described mathematically as a simple harmonic oscillation. The sounds in human speech are mostly non-harmonic complex oscillations, which can be described fairly accurately as a linear combination of harmonic oscillations of different frequencies and amplitudes, in the form of an acoustic spectrum. The vibration with the lowest frequency of this combination is the fundamental tone, and the others are usually multiples of the fundamental frequency and are called overtones or harmonics.

Moreover, different sounds are characterized and distinguished from each other by the distribution of energy in their spectrum. The same applies to the sounds produced by the human speech apparatus and especially to vowels and voiced consonants – as the example, shown on Fig.1.

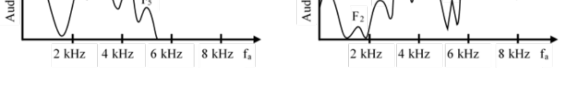

Fig. 1. Formants in spectrums of sounds ***i*** and ***s*** in ***i-n-s*** sequence.

Thus, acoustic energy is concentrated in certain areas along the frequency axis. The specificity of the sound is determined not only by the absolute position of these areas on the frequency axis, but also by the ratio between the power values concentrated in these zones. These areas are called formants – a more general term defining the zones along the linear spectrum in which a rise in vocal energy is observed. The central spectral line of such an area, where the highest value of the local peak is observed, is called the formant frequency [2]. According to their position on the frequency axis, in ascending order, the formants are labelled as ***F1***, ***F2***, ***F3***, ***F4***, ***F5***, etc. The distribution of acoustic energy along the frequency spectrum shows significant differences between vowels and voiced consonants, on one hand, and voiceless consonants and especially the fricative ones, on the other.

The example on Fig. 1 shows the acoustic energy distribution in the sound sequence {i n s}, as an extracted fragment from a longer one [3].

In order to characterize speech more fully, it is necessary to know its frequency spectrum as a whole – i.e. to work with the ***Long-Term Average Speech Spectrum*** (***LTASS***). Even with this representation of the frequency spectrum of speech, by averaging over time, different profiles can be observed – by gender, age, language and so forth – Fig. 2.

The minimum threshold at which sound remains audible depends on the frequency. The graph of this dependence is called the ***Absolute Threshold of Hearing*** (***ATH***).

Human hearing is most sensitive in the range between 1 to 5 kHz, approximately, although with age sensitivity decreases significantly in the range above 4 kHz. Here, the influence of the distinct characteristics of each individual person – age, state of health, professional and domestic hearing trauma and even taking some medications, is quite large.

And although it is strictly individual, the relationship between the sound pressure level, the frequency and a person's perception of loudness, through numerous studies and measurements, on large groups of people, averaged, is indicated in a publication of the International Organization for Standardization – ISO 226:2003 [4], as it is shown on Fig. 3.

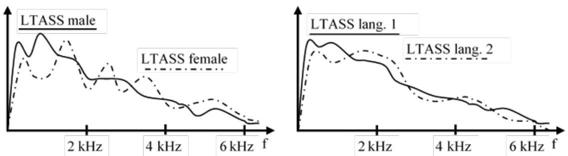

Fig. 2. Different forms of LTASS.

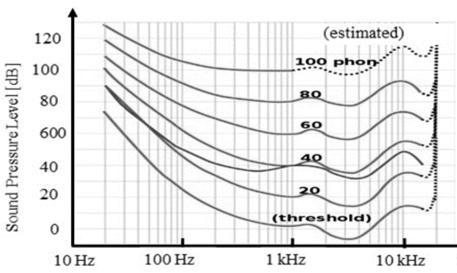

Fig. 3. Equal-loudness contours by ISO 226:2003.

Another statistical means of describing speech characteristics is based on its instantaneous values.

The ***Probability Density Function*** (***PDF***) $f(a_v,t)$ of acoustic signal values can be represented with sufficient accuracy as that of a non-stationary normal random process (2). This random process has zero mathematical expectation of the amplitudes a_V, it is compact, but has a time-varying value of the variance $D_V(t)$ [5] - Fig. 4.

$$f(a_V,t) = \frac{1}{\sqrt{2\pi D_V(t)}} \exp\left\{-\frac{a_V{}^2}{2D_V(t)}\right\} \qquad (2)$$

In other words – in the flow of speech, quiet sounds, such as voiceless consonants, are much more likely to occur than loud sounds, such as vowels and voiced consonants.

The large differences in the variances of this process, when it represents voiced sounds, on the one hand, or consonant sounds, on the other hand, requires the communication channels to have a fairly wide dynamic range, on the order of 50 ÷ 60 dB.

Voice communications in a real-world environment always occur in the presence of external acoustic noise. In an airplane cabin, the width of the noise spectrum measured at the 0.5 power level is about 10 kHz.

The power of acoustic noise, in some cases, can be comparable to that of speech.

The speech level is usually about 60 ÷ 70 dB above the threshold of audibility. The acoustic noise level in the cabin of a jet fighter during engine afterburning can reach 120 dB, where the noise power exceeds speech by 50 ÷ 60 dB. To protect against acoustic noise, noise-resistant laryngophones and microphones are used, which have a special construction, directivity diagrams and generally provide a better acoustic signal to acoustic noise ratio. The noise can also be significantly reduced by using an airtight flight helmet.

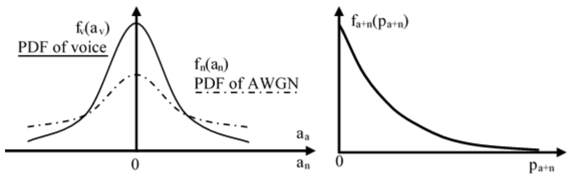

Fig. 4. Probability density functions of voice and noise.

B. Speech intelligibility - types and general definitions

Speech intelligibility is defined as the relative share of correctly perceived speech elements: sounds; syllables; words and phrases; to the total number of submitted ones.

Depending on the elements of the speech used for evaluation, sound, syllabic, word (single-word), phraseological (semantic) intelligibility of the speech is determined accordingly. In this arrangement, with the same gradation, it affects how much the listener is prepared for the topic [6], or at least how much meaningful content is embedded in what is being said to them. The statistical relationship between different types of speech intelligibility has been proven through numerous experiments and studies [5], [7].

One very convenient and informative statistical relationship is that between phonetic intelligibility ***S*** [%] and word intelligibility ***W*** [%], i.e. seamlessly passing or skipping through syllabic intelligibility – Table 1. This is so because the sounds are the main carriers of the technical parameters of the voice – fundamental frequency, dynamics, formants, etc. In turn, words are the primary conversational fragment that is a meaningful carrier [5].

TABLE 1. CLASSES OF COMMUNICATIONS.

Class	Communications by Speech Intelligibility		
	Characteristic of the class	***S [%]***	***W [%]***
I	Speech intelligibility without the slightest strain on attention.	> 91	> 98
II	Speech intelligibility without difficulty.	85 ÷ 90	94 ÷ 97
III	Speech intelligibility with attentional tension, without further inquiry and repetition.	78 ÷ 84	89 ÷ 93
IV	Intelligibility of speech with high attentional tension, further inquiry and repetition.	61 ÷ 77	70 ÷ 88
V	Connection lost.	< 60	< 69

C. Influence of the technical parameters of the equipment and working conditions on speech intelligibility

The microphone, which converts acoustic signals into electrical signals, is essentially a complex mechanical system, as it can be seen on Fig. 5, with its own frequency and transient characteristics [8].

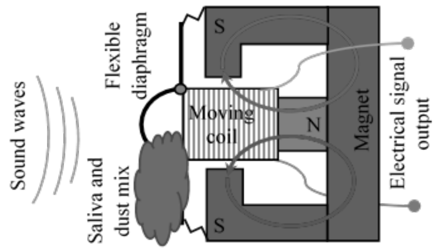

Fig. 5. Microphone construction.

The microphone also has a specific directivity diagram, which suggests that it should be properly positioned in relation to the sound source. In voice communications, this source is a person's mouth, and therefore a microphone placed on the larynx (laryngophone) or chest cannot be expected to reproduce the human voice with the same quality as one placed in front of the speaker's mouth. A good idea of the influence of the position of the microphone on the spectrum of the converted acoustic to electrical signals can be found in posters published by established microphone manufacturers [9].

With the ***PDF*** of speech sounds shown above according to their strength, and indirectly according to their belonging to different groups – vowels, voiced or voiceless consonants, the different noise reduction methods have a strong influence on speech intelligibility. One such approach to reducing the influence and fatigue caused by noise is squelching, where amplifiers only pass signals through themselves if they are above a certain level, which is assumed to be the noise floor. Unfortunately, the weak acoustic signals of voiceless consonants often fall below the squelch level. This way of combating the influence of noise leads to a significant deterioration of speech intelligibility – Fig. 6.

Limiting the influence of side noises through using of squelch circuits is also applied in the transmitting part of the radio equipment. This is the case, for example, in the use of laryngophones and automatic switches of the transmitters (voice operated exchange - VOX), which work when they detect a signal from the microphone with a level higher than set lower limit (voice presence).

Limiting acoustic signals from above (clipping) has little effect on speech intelligibility, but listening to such speech is unpleasant, as the natural ratios between the strengths of vowels, voiced and voiceless consonants are disturbed [5].

A widely known circumstance is that in non-tonal languages, such as Western ones, including English, as well as in Slavic languages, a large part of the semantic information is placed in the consonants. Applying different types of filters – ***Low Pass Filters*** (***LPF***); ***High Pass Filters*** (***HPF***); ***Band Pass Filters*** (***BPF***) and ***Band Stop Filters*** (***BSF***) – in order to reduce the general level of noise passing through the transmitting and receiving equipment and the communication channel, affects speech intelligibility – Fig. 7 [5].

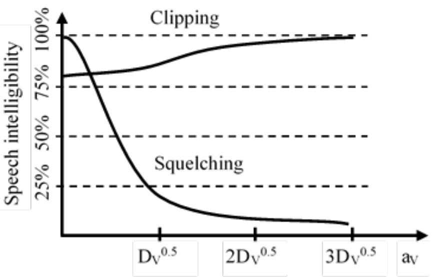

Fig. 6. Influence of clipping and squelching on intelligibility.

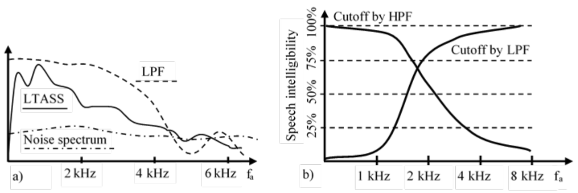

Fig. 7. Influence of cutoff on speech intelligibility.

While in the low-frequency range of voice signals (up to about 1.5 kHz) where most of their energy is concentrated, the limitations in this range set by a ***HPF*** cutoff do not lead to a significant reduction in speech intelligibility.

Limitations in the high-frequency range set by a ***LPF*** cutoff (below 2 kHz), however, lead to a drastic drop in speech intelligibility [5], [9].

The other place where bandwidth limitation can occur is in radio channels. There, the fight against noise and interference from adjacent channels in the receivers is carried out with the help of ***BPF***s.

III. Results and Discussion

The presented in previous section features are exemplary and serve only to illustrate the idea. This work is not intended to prove that a particular language or voice type is more suitable than any other is for aviation's or any other type of voice communications.

As can be seen on Fig. 1. the moment distributions show significant differences in the distribution between the formants of the sound { i..} and the sound {..s }.

It can be seen that for the vowel sound {i…} the audio energy is concentrated mostly around the first four formants. With the fricative consonant {…s}, not only are more formants ***F1*** ÷ ***F9*** formed, but also the sound energy is mainly concentrated around those with larger numbers – ***F4*** and up. Thus, it can be assumed that the energy of this sound is mostly located in the frequency band above 4 kHz, keeping a significant value around 8 kHz.

The situation is similar with the other voiceless consonants: f, k, p, t, etc., whose energy is mostly concentrated in the range around 4 kHz.

It should be noted that along with the purely physical characteristics of speech sounds, and the technical characteristics of electronic equipment, the intelligibility of speech is influenced by the psychophysiological characteristics of sound audibility.

The psychophysiological characteristics of sound reflect a person's subjective perception and feeling of perceived sounds. Some of the effects, such as masking of sounds and hearing phantom sounds, are related to the way a person's brain and nervous system work. These features of human hearing are the basis of the psychoacoustic model which provides many of high quality irreversible sound compression methods (mp3 , aac, Vorbis, etc.). All of these forms involve voice digitization, after which non-essential to quality components are removed [10].

It can be seen that while the language with its rules and constructions is a kind of general contract for communications in social groups, the physical and psychophysiological characteristics of the vocal sounds that are used in communications through this language are quite diverse and strictly individual.

For example of the statistical relationship between different types of speech intelligibility, 80% syllabic corresponds to 98% word intelligibility, or 40% syllabic intelligibility corresponds to 90% phraseological one, the latter being a rather low and unsatisfactory value. In dispatch communications, due to the operators' and users' prior knowledge of the set of possible phrases, fully acceptable speech intelligibility is achieved with syllabic intelligibility between 40% and 50%, which corresponds to word intelligibility between 87% and 93% [7].

Speech intelligibility is also affected by many environmental factors, such as the ratio of average speech power to average ambient noise power. Also important are the attenuation time of power-significant reflections of sounds from surrounding walls and objects (reverberation time), the frequency profile of absorption of sound waves by surrounding surfaces and other similar factors. Naturally, the intelligibility of speech depends mostly on the characteristics of the media carrying the voice information. The influence becomes especially complex when the media is mixed: e.g. speech apparatus - air environment - electronic equipment - radio channel - electronic equipment - air environment - hearing apparatus.

The closest to the technical parameters and limitations of the equipment are sounds with their parameters and the corresponding sound intelligibility. In this way, through the statistical relationship between the different types of speech intelligibility, the influence of purely technical characteristics of the equipment can be traced, as an example – the influence of frequency or impulse characteristics on the most significant, highest form of speech intelligibility, namely semantic intelligibility.

In dispatch communications, when operating conditions are around the normal for the scenario, conversations are fairly formalized and lines, phrases and words are largely predictable [6]. This applies both to the tone and to the speed of their speech – they are calm and close to those of a normal conversation. This is not the case in emergency situations – when, under the pressure of extraordinary circumstances, unexpected, and in some such cases, thoughtless words and phrases appear in voice communications, indistinctly spoken, at high speed and in a high tone. This shifts the spectrum of voice signals upward, the influence of high-frequency suppression by low-pass filters becomes even greater and, as a result, already difficult-to-understand phrases become even more unclear. If in these situations the crew or the control staff also perform any physical actions, these can obstruct normal breathing or lead to the microphone being moved from its optimal position – the result is further deterioration of intelligibility.

In order to properly convert sound vibrations, especially those of higher frequencies, the moving parts of the microphone's overall construction (for the electrodynamic microphone, these are an elastic membrane and an attached coil) must be light and elastic. The accumulation of a mixture of saliva and dust on the internal moving parts of the microphone not only significantly reduces its sensitivity, but also seriously distorts its frequency response, especially by suppressing the high frequencies of the voice, where the voiceless consonants spectrum is mostly concentrated.

Reducing of speech intelligibility happens mainly due to the undesirable reduction in the level of high-frequency voice signals – such as voiceless consonants, which already have a low level. This is because it is impossible to make an ideal filter – with flat frequency pass bands, steep slopes, and frequency bands of full rejection – Fig. 7a. Researches have shown that limiting the bandwidth of voice signals from below with ***HPF*** or above with ***LPF*** affects speech intelligibility differently – Fig. 7b.

This means that in our desire to reduce the level of noise, which compared to voice signals has a wide uniform frequency spectrum, we can cause significant damage to speech intelligibility if we use frequency filtering carelessly.

Moreover, when fighting noise, neither the low-frequency range should be sacrificed, since it provides us with a good signal-to-noise ratio, nor the high-frequency range, since it provides us with good intelligibility – Fig. 8.

The frequency band of radio channels Δf_R, compared to that of speech signals, is much wider. If the transmitting and receiving radio equipment are tuned to the same carrier frequency, there should be no bandwidth limitation of the speech signals due to the ***BPFs*** in the receiving radio equipment.

However, if the carrier frequency or the receiving frequency is unstable, due to damaged and/or outdated equipment, or if the two frequencies differ for some other reason – a Doppler shift, as an example. It is then possible to obtain bandwidth limitation of the speech signals from above due to suppression of one of the sidebands – ***Lower Side Band*** (***LSB***) or ***Upper Side Band*** (***USB***), as it is shown on Fig. 9.

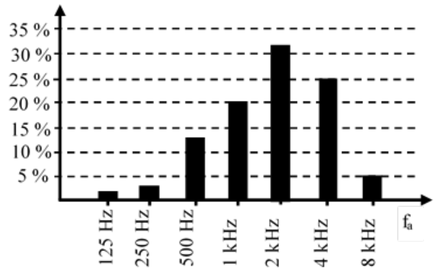

Fig. 8. Importance of spectral regions for speech intelligibility.

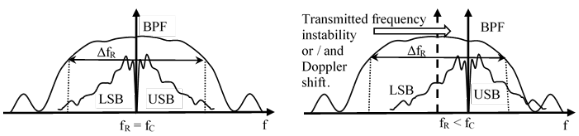

Fig. 9. Spectral cutoff from frequency instability or Doppler shift.

In conversion of analog signals to digital and their subsequent digital processing, there are also potential dangers of inaccurate representation of the frequency spectrum of speech signals. This can be seen in the reverse conversion of the digitized voice signal back into an analog one and its subsequent playback.

Problems here are, as follows – to convert analog signals to digital, they are subjected to the process of time sampling, in which the analog signal is sampled over a period of time.

A basic and important requirement is that these samples are taken at a frequency called the sampling rate f_d, which should not be lower than the doubled highest frequency in the spectrum of the analog signal. For speech signal processing, the sampling rate is standardized on f_d=8 kHz. This means that before being converted to digital, the speech signals should be well limited to a frequency not higher than 4 kHz, using a low pass filter LPF_T.

If, with the idea of keeping the highly informative part of the upper frequency band in the spectrum of the speech signals, it is limited to a higher frequency, then when applying the standardized sampling rate, a rather harmful phenomenon is obtained.

The lower sideband LSB_1 that, in results of digitalization, forms around the sampling frequency f_d enters the area of natural speech bandwidth, with mirror-image spectral components.

A low pass filter ***LPF$_R$*** is applied when restoring the analog signal from the discrete samples obtained by the receiver. Then, the mirror-inverted spectral components cause a false recovery of the frequency components of the original speech signal. This phenomenon looks like a “frequency spectrum bending”, as it can be seen on Fig. 10.

The incorrect restoration of amplitudes and relationships between high-frequency components of the speech signal spectrum, especially in the case when their order is reversed, leads to a serious degradation of speech intelligibility.

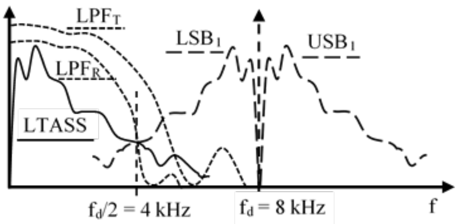

Fig. 10. Frequency spectrum bending.

IV. Conclusions

The technical condition and parameters of the electronic equipment intended for voice transmission significantly affect speech intelligibility. This happens, mostly, due to the impact of these characteristics on the frequency spectrum of speech signals, especially in their high-frequency range.

Therefore, a few key points need to be regarded:

- Electronic equipment should be regularly serviced by qualified technical personnel who are fully informed of the possible effects of their work on speech intelligibility.
- Only well-trained and drilled personnel, with an adequate and sustainable psycho-physiological status, should use the electronic equipment.
- It should be considered to update standardized technical requirements for electronic equipment for voice transmission, which would lead to improvement of speech intelligibility and improvement of the quality of voice communications in general.

Acknowledgments

This work was supported by the NSP SD program, which has received funding from the Ministry of Education and Science of the Republic of Bulgaria under the grant agreement no. Д01-74/19.05.2022.

References

[1] Microphone Handbook. Vol. 1: Theory, Naerum: Bruel & Kjaer, 2019, pp. 2-2 to 2-3.

[2] T. Lubin, Getting Great Sound: The Microphone Book (Third Ed.), London: Rowman & Littlefield Publishers, 2022, p. 51.

[3] S. Wood, What are formants, January 2005. [Online]. Available: *Lund University, Sweden – SOL* https://person2.sol.lu.se/SidneyWood/praate/whatform.html [Accessed: March 13, 2023].

[4] Equal-loudness contour, October 2023. [Online]. Available: *Wikipedia – The Free Encyclopedia*
https://en.wikipedia.org/wiki/Equal-loudness_contour. [Accessed: Jan. 9, 2024].

[5] V. I. Tihonov Ed., Aviation radiocommunication devices (in Russian), Voronezh: Zhukovsky AFEA, 1986, pp. 24-39.

[6] V. Katsarska, “Formulaic Language in Aeronautical English – Friend or Foe”, *The Journal of Teaching English for Specific and Academic Purpose*, vol. 9, no. 2, pp. 239–245, 2021. [Online]. Available: http://espeap.junis.ni.ac.rs/.
[Accessed: Feb. 15, 2024],
https://doi.org/10.22190/jtesap2102239k.

[7] Speech clearency and intelligibility. (in Russian), February 2016. [Online]. Available: *StudFiles*
https://studfile.net/preview/5395324/.
[Accessed: Feb. 15, 2024].

[8] J. Eargle, The Microphone Book (Second Edition). From Mono to Stereo to Surround. A Guide to Microphone Design and Application, Oxford: Elsevier/Focal Press, 2005, pp. 42-45, 340.

[9] Mic Univesity, *DPA Microphones*, [Online]. Available:
https://www.dpamicrophones.com.
[Accessed: Jan. 9, 2024].

[10] M. Bosi and R. E. Goldberg, Introduction to Digital Audio Coding and Standards, New York: Springer Science +Business Media, 2003, pp. 315-367.