Linguo-Acoustic Aspects of Speech

Marianna Hudcovičová[1], Božena Petrášová[1], Monika Rychtáriková[2]

[1]Department of British and American Studies, University of Ss. Cyril and Methodius, Trnava, Slovakia.

[2] Department of Architecture, KU Leuven, Leuven, Belgium.

marianna.hudcovicova@ucm.sk, bozena.petrasova@ucm.sk, monika.rychtarikova@kuleuven.be

Abstract
Speech is the ability making human beings unique. Linguistic diversity reflecting wide variety of speech segments does not only bring about different speech sound production mechanisms, but because of lack of appropriate acoustic research, reminds us of the need to document acoustic properties of individual sounds, as well as their possible combinations in speech in various languages. The paper presents the issues connected with the process of preparing a digital database of Slovak sentences that will be used for testing speech intelligibility in various acoustic conditions. The sentences produced as spoken utterances will be selected according to the most significant linguistic criteria to meet all the requirements for creating the relevant database. The standardised sentences will be recorded and subsequently tested in simulated laboratory conditions, while paying attention to variables in different acoustic environments. The phonetic features of the material analysed will be evaluated by acoustic analyses using the computer program PRAAT (Praat software Version 6.1.47; Boersma & Weenink) which is also used to create auditory spectrograms. The program helps to test speech intelligibility, as it measures pitch, formants, intensity and duration of sounds. These aspects are fundamental for perceiving speech sounds and decoding them properly. Based on the rules of phonotactics, combinations of vowel and consonant phonemes (in open/closed syllables, mono-/polysyllabic words) reflecting higher or restricted level of speech intelligibility in different acoustic environments will be studied. The linguistic data will be collected via the Slovak National Corpus, version prim 9.0. In order to construct the sentences from the most frequent words, the electronic statistical tools will be applied: ARF (average reduced frequency) and potential occurrence in one million words. For detecting the co-occurrence of the words, the statistical tools of the electronic corpora will be applied: MI Score and T Score.

Keywords: speech, intelligibility, digital database, acoustic conditions, linguistic criteria

1 Introduction

Communication is a fundamental and most natural way of interaction between living creatures. Not only people communicate together and exchange messages, animals or even plants can interact like this as well. The spoken form of language is dominant and prevails over the written form. Its historical, structural, functional and biological priority is inevitable and as such also proved by many scholars. The process of exchanging information could be presented by words (verbal communication) and/or by gestures or other body movements representing non-verbal communication. Speech makes use of the smallest units of language which are manifested as concrete, audible realisation of phonemes.

2 The sound system

The sound system of languages is studied by phonetics and phonology. These linguistic disciplines analyse speech properties from various points of view. Phonetics is more practical, it deals with a concrete realisation of sounds in speech, phonology is more theoretical and studies an abstract side of a language. Three main branches of phonetics - articulatory, auditory and acoustic – study phases of communication, as it is depicted in a speech chain:

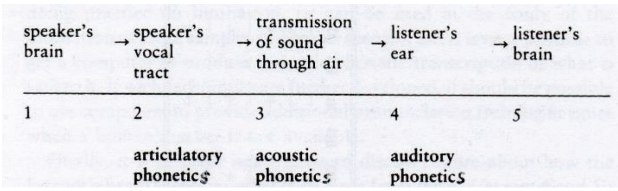


Figure 1: Speech chain [10]

In phonemic inventories of English and Slovak, vowel and consonant phonemes are manifested. The symbols used in transcription are based on the International Phonetic Alphabet (IPA).

The English language - phonemes:
English short vowels: */ʌ e ɪ ɒ ʊ ə æ/*
English long vowels: */a: i: ɔ: u: ɜ:/*
English diphthongs: */eɪ aɪ ɔɪ əʊ aʊ ɪə eə ʊə/*
English triphthongs: */eɪə aɪə ɔɪə əʊə aʊə/*
English consonants: */p t k b d g f v θ δ s z ʃ ʒ h ʧ ʤ l m n η r j w/*

The Slovak phonemic inventory comprises 11 vowels (6 short and 5 long vowels), 4 diphthongs and 27 consonants. Their pronunciation is notated in slant brackets:

Slovak short vowels: phonemes */a e i o u/*, two more graphemes are used: *y* (pronounced */i/)* and *ä* (pronounced /*æ*/ - this phoneme is very rarely used in common speech, it is present in a higher style only)

Slovak long vowels: */a: e: i: o: u:/*, one more grapheme is used. It is written *ý* but pronounced the same way as */i:/*

Slovak diphthongs: */ĭa ĭe ĭu ŭo/*

Slovak triphthongs: none

Slovak consonants: */p t ť k b d ď g f v s z ʃ ʒ c dz ʧ ʤ h x l ľ m n ň r j l: r: /* (ť, ď, ľ, ň are palatalised t, d, l, n)

In English, each short vowel and its long counterpart differ in both quality and quantity. In Slovak, the only difference between short and long vowels lies in their quantity.

The English and the Slovak language differ from various perspectives. Regarding their origin and classification, English comes from the family of the Germanic languages and Slovak is a Slavic one. Another difference lies in their structural point of view. English is analytical, while Slovak is a synthetic language. In other words, grammar structures, owing to either morphological or syntactic issues, are quite different. Describing the sound level, English, unlike Slovak pronunciation, is based on a phonological principle between the written and spoken form of words. It means that in Slovak, in each word a specific grapheme is always pronounced the same, i.e. represented by the same phoneme, whereas in English, no exact rules for pronunciation of single graphemes or their clusters are stated. Thus the already mentioned correspondence between the written and spoken form is a matter of convention. A typical example which can be found in many English textbooks is the pronunciation of English *oo*. Following the occurrence of this vowel cluster in English words, its different pronunciation can be really confusing for learners. Although the written form is identical, spoken equivalents of the written form vary, e.g. *blood /blʌd/, book /bʊk/, door /dɔ:/, cool /ku:l/*, etc. Despite the graphical parallelism, i.e. occurrence of the identical combinations of the graphemes, the spoken equivalents are different. In Slovak, no such discrepancies between the written and the spoken form of words exist.

2.1 Speech intelligibility

Speech intelligibility in architectural and/or civil engineering practice is generally represented by STI (speech transmission index) parameter which can be measured or simulated in different acoustic conditions [1]. The intelligibility of natural speech can be explored in various ways. For example, noise can be added to a speech signal at various pitch levels with variable bandwidth and varying intensity. Some parts of the speech signal can be filtered out, some sounds can be deleted, speech intensity can be changed. The basic results of natural speech intelligibility research prove the following facts: human natural speech is intelligible in the sound level range of about 80 dB, it is fully intelligible even at very low sound levels down to 40 dB, depending on Signal to noise ratio. According to some already done experiments, if natural speech is mixed with noise of the same intensity over the entire frequency range of hearing, noise has no effect on speech intelligibility when the ratio of speech intensity to noise is 100: 1. Speech can remain intelligible even when the noise is more intense than speech or if the speech signal comes from a different direction than the noise [2]. It is a consequence of directional hearing. Other results were obtained by speech filtering. Speech is intelligible when it passes through a low-pass filter (it means that it passes only in low frequencies), but its upper limit must be so high that the most important frequency components of speech pass through the filter. It has also been found that speech is 67% intelligible either when only low frequencies up to 1800 Hz are transmitted, or it is 67% intelligible when only frequencies above 1800 Hz are transmitted. Other experiments have shown that a 1000 Hz wide band in the region of about 1500 Hz is sufficient for 90 % speech intelligibility. The rest of low and high frequency speech sounds may be filtered out. Remarkable speech intelligibility is achieved if only the 100 - 3000 Hz band is passed. Speech remains intelligible even in adverse conditions and when an appropriate signal is not generated [3]. Recent studies have also shown the effect of sound filtering on speech intelligibility, when wearing the face masks during pandemics of COVID-19 of speech as well [4, 5].

Owing to linguistic aspects, the intelligibility of sentences is greater than the intelligibility of words, and the intelligibility of words is consequently greater than the intelligibility of syllables. The context and situation in which speech takes place is of great importance. The specific communicative situation that corresponds with the given information is also crucial, because in certain situations we expect a specific response as well as an adequate reaction. Therefore, the expected word is much easier to recognize than the unexpected word. The key point of the information is carried by the phonetic context, knowledge of grammar and vocabulary of the language, as well as possible combinations of its segments (e.g. phonotactics which means possible speech sound combinations in a certain language). Speech recognition is therefore based on a combination of acoustic, grammatical, semantic and situational information sources. Speech intelligibility is based on many sources of information, not just acoustic information itself. Everything that happens in spoken communication is derived from acoustic information, which is the basic stimulus for recognizing the speech mechanism. In natural speech, there are also information signal components that can complement and replace each other [3].

2.2 Digital database

One of the ways how to test speech intelligibility is to provide acoustic and linguistic tests focused on listening comprehension of a set of carefully constructed sentences which are recorded first in silence in laboratory conditions and then in changed acoustic environments with increased noise components of various intensity. The selected standardised sentences form a so-called digital database which is prepared for a specific language by experts. To prepare the database as well as possible, the experts are supposed to specialise both in acoustics as well as in linguistics. The databases already exist and are used for testing speech intelligibility in many languages, e.g. the Dutch language, the Danish language, the Finnish language or the German language [6]. Our prior objective of the research is to prepare a digital database for testing speech intelligibility of such selected sentences in the Slovak language [7, 8].

3 Methodology

The research material has already been partially evaluated through description, observation, statistic methods, and synthesis. Linguistic data collection was done via the Slovak National Corpus, version prim 9.0. In order to create sentences, the electronic statistical tools were applied: ARF (average reduced frequency) and potential occurrence in 1 million words [9]. In order to detect co-occurrence of the words, the statistical tools of the electronic corpora were used: MI Score and T Score. The sentences comprised in a Slovak digital database were selected and recorded. The research sample used for this paper includes 760 Slovak sentences which differ from the structural and functional point of view. The sentences were recorded by a native Slovak speaker and then tested for their intelligibility. The whole procedure of this test is in [8]. Recordings were performed in silent environment of semi-anechoic room at the Faculty of Civil Engineering STU Bratislava, Slovakia. The intention of the performed research is to verify the STI parameter for the Slovak language using developed standardized sentences. In the following research phase, the phonetic features of analysed material will be evaluated by acoustic analyses using the computer program PRAAT (Praat software Version 6.1.47; Boersma & Weenink) which can be also used to create auditory spectrograms (or cochleagrams). Praat is a free software created by Paul Boersma and David Weenink from the University of Amsterdam. The software allows the user to do an accurate analysis of spectrograms, cochleagrams, pitch, formants, intensity, along a multitude of other functions. The software is used to recognize F1 and F2 values of vowel sounds, which furthermore allows the identification of the tongue position during the pronunciation of vowels. The program helps to test speech intelligibility by measuring pitch, formants, intensity and duration of sounds. These aspects are fundamental for perceiving speech sounds and decoding them properly.

Regarding issues of acoustic phonetics, the basic acoustic parameters of individual vowel sounds are frequency, intensity, duration and quality, i.e. wave structure determined by formants in vowels. Spectrograms allow the visualisation of three main properties of sounds: frequency, length, amplitude. The F1 (i.e. *formant 1*) and F2 values are directly connected to the height and location of vowels: for high (close) vowels, low F1 values are recognised, low (open) vowels are defined by high F1 values. On the other hand, for front vowels, it is typical to have high F2 values and back vowels have low F2 values.

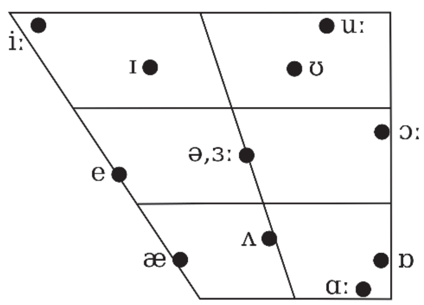


Figure 2: English distinctive vowels [10]

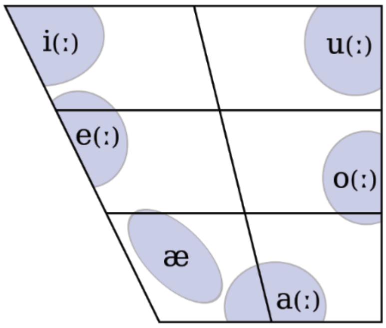


Figure 3: Slovak distinctive vowels [3]

F3 parameters reveal the level of nasality in vowel pronunciation. For consonants, the manner and place of articulation are the most relevant factors for their classification. According to the F1 value, the size of the constriction gives a clue to the manner of articulation, F2 and F3 values are related to the place of articulation. Formant structure is not recognized for voiceless consonants. Frequency corresponds to resonance in the vocal tract. Its length determines the pitch, i.e. the value of the frequency. The average formant frequencies vary when it comes to differences in a person´s gender or age. The pitch of a sound is of the highest value in children's voice (300 Hz) and lowest in adult males (110 Hz). Female voice (120-180 Hz) occurs in the middle of the frequency span. The reason is length of the vocal tract, women generally have a shorter vocal tract than men. A child's vocal tract is undoubtedly shortest causing the highest values of formant frequencies.

Table 1: Average formant frequencies of English vowels pronounced by an adult male [12]

Vowel	F1 (Hz)	F2 (Hz)	F3 (Hz)
i:	280	2620	3380
ɪ	360	2220	2960
e	600	2060	2840
æ	800	1760	2500
ʌ	760	1320	2500
ɑ:	740	1180	2640
ɒ	560	920	2560
ɔ:	480	760	2620
ʊ	380	940	2300
u:	320	920	2200
ɜ:	560	1480	2520

Table 2: Average formant frequencies of Slovak vowels [3]

Vowel	F1 (Hz)	F2 (Hz)	F3 (Hz)
i/ i:	285	1916	2656
e/ e:	452	1718	2365
a/ a:	682	1315	2293
o/ o:	481	1084	2194
u/ u:	326	967	2059
æ	700	1510	2300

Acoustic properties of speech are traditionally depicted by spectrograms which show three dimensions of the sound: time, amplitude and frequency. "A spectrogram is a graphical analysis of a sound and offers an analysis in three dimensions. The first (horizontal) dimension represents time, the second (vertical) dimension represents frequency, and the third dimension (distinguished according to the colour intensity of the vertical lines) represents amplitude" [5]. "The intensity of the speech sounds is shown by the relative darkness of the marks. The vowels and vowel-like sounds are darkest, and the different vowel qualities can be clearly seen in the changing pattern of black bands (formants), which represent varying concentrations of acoustic energy in the vocal tract" [6]. Frequency which is measured in Hertz depicts the number of cycles per one second. According to the information mentioned by R.Pavlík, "the human ear is capable of hearing the sounds ranging from 20 to 20 000 Hz, however, most human sounds are recognizable within the frequency scope as narrow as 80 – 350 Hz" [10].

The other way how to depict speech parameters is using oscillograms. In the specialised literature, the term *waveform* is sometimes used instead. There is the horizontal axis, representing time, that directs from left to right. The curve presents the pressure increasing and decreasing in the speech signal.

Regarding linguistic criteria, acoustic parameters of speech sounds, especially vowels, were crucial for us. We stated the hypotheses:

H1: High vowels in Slovak sentences (i/i:, u/ u:) lower speech intelligibility.
H2: Polysyllabic words lower speech intelligibility of the sentences.

The other presumptions took into consideration different types of vowels/ consonants (segmental problems), tempo, rhythm, sentence stress, intonation (prosodic features), whether the words in the sentences were mono- or polysyllabic and if the words included open/ closed syllables. The other relevant criteria were their position in sentences, classification of words according to their word classes, as well as their frequency in the Slovak lexicon and sentence form from the functional and structural point of view – statements, questions, commands, exclamations, simple or multiple sentences. According to the rules of phonotactics (a branch of phonology), combinations of vowel and consonant phonemes (in open or closed syllables, mono- or polysyllabic words) reflecting higher or restricted level of speech intelligibility in different acoustic environments will be described. From the technical point of view, attention was paid on variables in different acoustic environments simulated in laboratory conditions. It was proved that selection of the sentences into the Slovak digital database required meeting both linguistic and technical criteria and parameters.

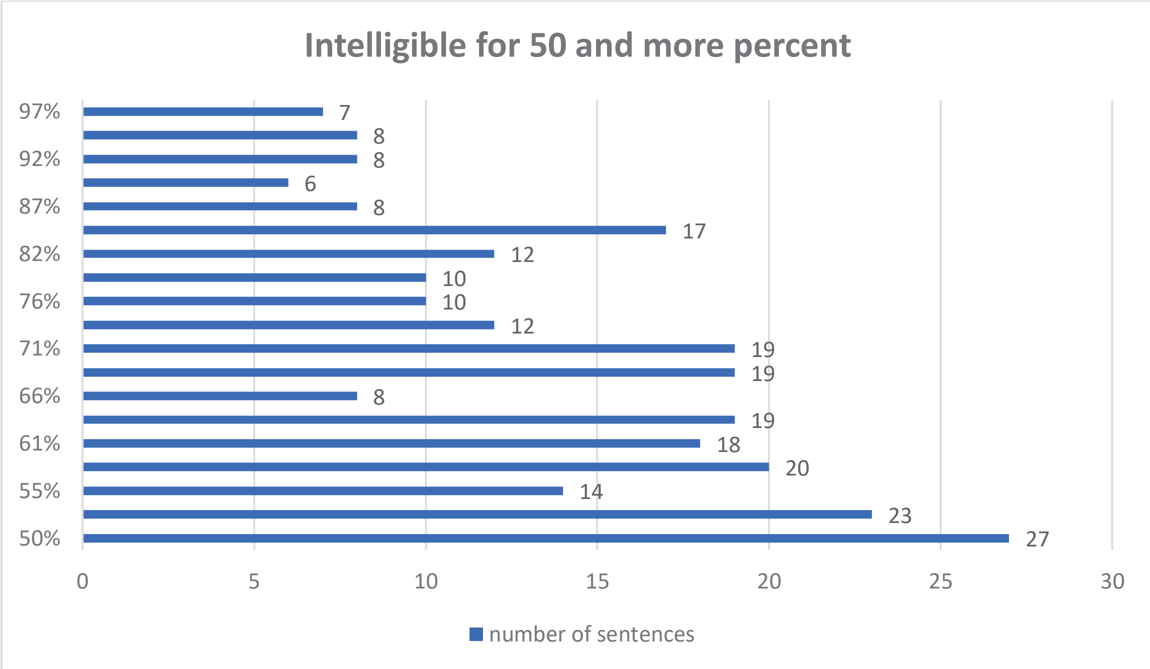
It does not matter whether the acoustic or linguistic criteria for describing the aspects of the analysed sentences in the research are used, it is crucial to use correct and adequate terminology in all disciplines of science. I.Cíbiková speaks about *terminology literacy* in this context and defines it as "the ability of a language users and specialists to use consistent terminology in functional specialised communication" [13, p.6]. According to this specialist, mastering terminology thoroughly must follow eight interrelated levels: 1) basic, 2) functional, 3) conceptual oriented, 4) defining, 5) interdisciplinary, 6) comparative, 7) documentary and 8) perspective. She emphasises the fact that scientific and terminology literacy must be consistent and reflecting a systematic documentation of human knowledge [13].

4 Research results

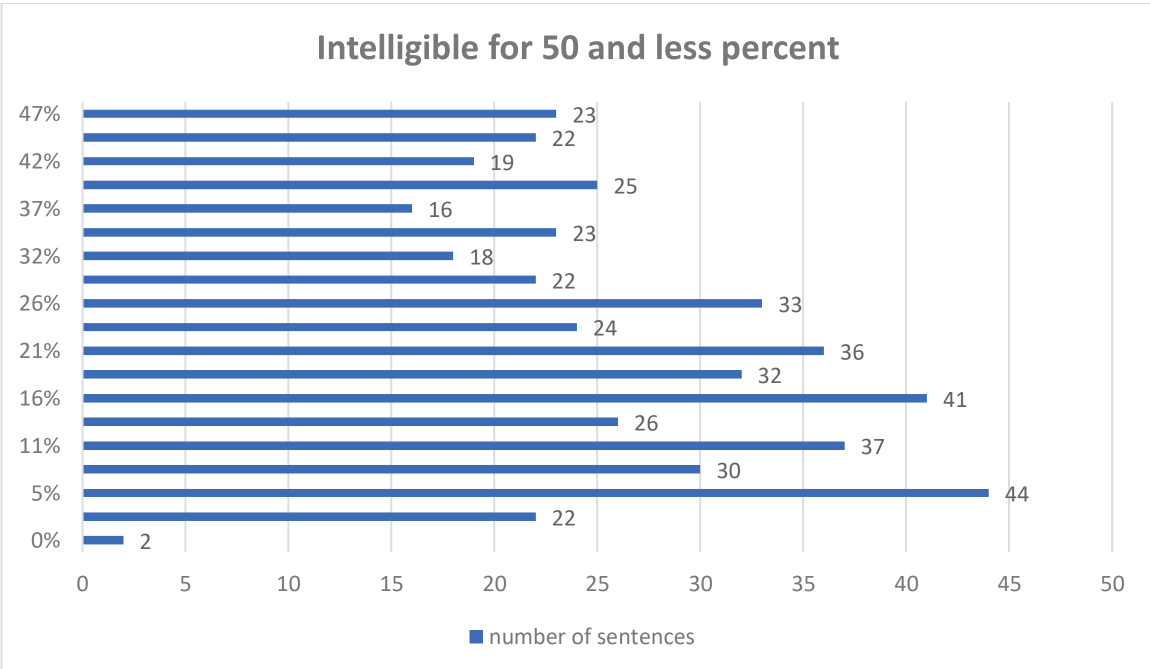
The research sample included 760 Slovak sentences which differ from the structural and functional point of view. They were recorded by a Slovak native speaker and tested in laboratory conditions with the added noise component during comprehending their intelligibility in simulated changed acoustic conditions. Their intelligibility has been evaluated in percentage:

7 sentences were intelligible for 97%, 8 sentences for 95%, 8 ones for 92%, 6 sentences for 89%, 8 for 87%, 17 for 84%, 12 for 82%, 10 for 79%, 10 for 76%, 12 for 74%, 19 for 71% as well as 19 for 68%, 8 for 66%,

19 for 63%, 18 for 61%, 20 for 58%, 14 for 55%, 23 for 53%, 27 for 50%, in TOTAL: **265 sentences** out of 760 were **intelligible for 50 and more percent**:



The rest (495 sentences in TOTAL) were intelligible for less than 50%: 23 sentences for 47%, 22 for 45%, 19 for 42%, 25 for 39%, 16 for 37%, 23 for 34%, 18 for 32%, 22 for 29%, 33 for 26%, 24 for 24%, 36 for 21%, 32 for 18%, 41 for 16%, 26 for 13%, 37 for 11%, 30 for 8%, 44 for only 5%, 22 for 3% and – what was surprising - 2 sentences were intelligible for 0%. It means that **495 sentences** were **intelligible less than for 50%:**



5 Conclusions

Partial analysis has already been done focusing on intelligibility of the Slovak sentences in a digital database. The total number of 760 sentences were evaluated according to their acoustic and linguistic properties, so the research results can be considered relevant owing to a high amount of the sentences analysed. All of the sentences were tested for intelligibility in changed, simulated acoustic laboratory conditions with the added background noise component, so the conditions were not ideal for comprehending every word comprised. The two hypotheses were stated, both were proved. Slovak high vowels (i/i:, u/ u:) were mostly present in sentences with the level of intelligibility less than 50%, so they really lower speech intelligibility. The sentences that were better understood and the meaning was clearly deciphered, included vowels a/ a:, o/ o: mainly. These vowels are considered low in the Slovak language.

Regarding Hypothesis 2: Polysyllabic words lower speech intelligibility of the sentences was proved as well. The sentences with the level of intelligibility less than 50% included mainly polysyllabic words. The sentences with a higher degree of comprehension comprised predominantly shorter words than the sentences with a lower degree of intelligibility. It is a great challenge to explore also the other aspects supporting or breaking intelligibility of Slovak sentences, specifically the ones involved in the already prepared database. The research is still ongoing.

References

[1] HOUTGAST, Tammo – STEENEKEN, Herman. Past, present and future of the Speech Transmission Index. Publ. TNO Human Factors, Soestenberg, 2002. ISBN 90-76702-02-0.

[2] RYCHTÁRIKOVÁ, Monika, et al. (2011): *Perceptual Validation of Virtual Room Acoustics: Localization and Speech Understanding.* Applied Acoustics 72 (4), p. 196-204, ISSN 0003-682X.

[3] KRÁĽ, Ábel – SABOL, Ján. Fonetika a fonológia. Bratislava: SPN, 1989. ISBN 80-08-00036-8.

[4] BOTTALICO, Pasquale, et al. Effect of masks on speech intelligibility in auralized class-rooms. J. Acoust. Soc. Am. 2020, 148, 2878–2884.

[5] CHMELÍK, Vojtech, et al. Effect of Mouth Mask and Face Shield on Speech Spectrum in Slovak Language. *Appl. Sci.* 2021, *11*, 4829. https://doi.org/10.3390/app11114829

[6] VAN WIERINGEN, Astrid – WOUTERS, Jan. Sentences and numbers for quantifying speech Understanding in severely impaired listeners for Flanders and the Netherlands. Publ. International journal of Audiology. London, 2008.

[7] PANOCOVÁ, Renáta – GREGOVÁ, Renáta. Designing the Slovak matrix sentence test, International journal of applied languages and cultural studie. Volume 2, Issue 2, 2019.

[8] HÚDOKOVÁ, Dominika et al. Conception of the listening test procedure for quantifying speech intelligibility in Slovak language – a preliminary study, *EUROREGIO/BNAM2022.* Aalborg, Denmark, 2022.

[9] Slovak National Corpus, version prim 9.0

[10] PAVLÍK, Radoslav. Phonetics and Phonology of English. Bratislava: PdF UK, 2000. ISBN80-88868-67-X.

[11] CRYSTAL, David. The Cambridge Encyclopedia of the English Language. New York: Cambridge University Press, 2011. ISBN 978-0-521-53033-0.

[12] WELLS, John Christopher. Sounds Interesting. CUP. 2014. ISBN 978-1107074705.

[13] CÍBIKOVÁ, Ingrid – SIANTOVÁ, Gabriela – MITAĽOVÁ, Katarína. Specialised communication. Scientific and terminology literacy. In: International Masaryk Conference for Ph.D. Students and Young Researchers. Volume XII. Hradec Králové: Magnanimitas akademické sdružení. 2021. ISBN 978-80-87952-35-1.