# DESIGNING THE SLOVAK MATRIX SENTENCE TEST

Renáta Panocová, Faculty of Arts, Pavol Jozef Šafárik University in Košice, Slovakia,
E-mail: renata.panocova@upjs.sk
Renáta Gregová, Faculty of Arts, Pavol Jozef Šafárik University in Košice, Slovakia,
E-mail: renata.gregova@upjs.sk

**Abstract.** This paper presents partial results of a larger-scale project of designing the matrix sentence test for Slovak. The main aim is presentation and detailed discussion of linguistic aspects of Slovak matrix sentence test. First, morphosyntactic criteria are outlined. These are followed by description of problematic issues and the solutions proposed. Second, phonological criteria are given and discussed. In the next step, the matrix test will be optimized and evaluated in order to measure speech intelligibility function and to establish the correct reference data for listeners with normal hearing.

***Keywords:*** *Slovak matrix test, audiometry, speech perception, speech reception threshold.*



## INTRODUCTION

Speech audiometry is a standard method used in the diagnostics of hearing impairment. A number of standardized speech tests for individual languages have been designed over the past decades in order to determine the degree and nature of the impairment. Two main types of speech tests can be distinguished. One is based on meaningful, everyday sentences with a variable grammatical structure (e.g. Plomp & Mimpen, 1979; Nilsson et al, 1994; Kollmeier & Wesselkamp, 1997; Versfeld et al, 2000; Wong & Soli, 2005; van Wieringen & Wouters, 2008; Luts et al, 2008; Ozimek et al, 2009; Nielsen & Dau, 2011). The advantage of this type of test is that it accurately reflects everyday language in common communicative situations. On the other hand, its main disadvantage is that the sentences can be easily memorized. The other type of speech tests is a so-called matrix test. A matrix test is characterized by a fixed order of items, proper name, verb, numeral, adjective, noun (object), which produces grammatical sentences with an unpredictable meaning (Hagerman, 1982; Wagener, 1999a, b, c; Ozimek et al, 2010; Hochmuth et al, 2012; Jansen et al, 2012; Dietz et al, 2014; Houben et al, 2014; Kollmeier et al, 2015).

Corresponding Author
Renáta Panocová, Faculty of Arts, Pavol Jozef Šafárik University in Košice, Slovakia,
E-mail: renata.panocova@upjs.sk



The very first matrix test was developed by Hagerman (1982) for Swedish. His matrix consisted of 10 first names, 10 verbs, 10 numerals, 10 adjectives and 10 nouns. This represents a corpus of 50 distinct words. Test sentences are generated randomly from the matrix in a way that all sentences have an identical syntactic structure, e.g. *Kathy sees nine small chairs*. Alan gives eight dark toys, that is, SVO. The matrix serves as a basis for the set of 10 sentences in which a word can occur only once in a set. This results in 105 = 100 000 sentences or in other words, 10 000 sets with 10 sentences each.

In the past years, the matrix test developed in Oldenburg, Germany (Oldenburger Satztest, OLSA) has become popular and widely used. Gradually, matrix tests started to be designed for individual languages. At the moment, there are nine matrix tests available as a medical device for German, American English, Spanish, Finnish, Italian, Polish, Russian, French and Turkish. Eleven matrix tests are officially under development including British English, Swedish, Danish, Norwegian, Hebrew, Arabic, Persian, Dutch, Japanese, Chinese, Hindi (Hörtech, Oldenburg, 2015). It is expected that the number of languages with matrix tests readily available as a medical device will continue increasing.

There are at least three strong points of matrix tests. Undoubtedly, the main advantage is that the meaning of the sentences cannot be predicted. Given the total number of possible generated sentences (see above) and their low meaning predictability, it is unlikely that patients will memorize them. An advantage following from this is that the matrix test can be conducted repeatedly with the same patient without negative influence on the test results.

Another advantage is that a patient can be tested in any language for which a standardized matrix test is developed. In addition, it is not necessary that an audiometrist speaks the language the patient is tested. Matrix Tests can be carried out in a so-called closed test format. This means that the patient sees the matrix of possible words on a computer screen and can select the words that he or she just heard. Last but not least, individual language versions of matrix tests can be easily compared thanks to their standardized structure.

For only two Slavic languages, Russian and Polish, a matrix test has been developed so far. The aim of this paper is to outline the process of designing a matrix speech test for Slovak.

## MATRIX TEST SPEECH MATERIAL

Slovak belongs to the West Slavic group of Slavic languages together with Czech and Polish. It is an inflectional language with elaborated declension and conjugation systems. The morphological type of Slovak played important role in the selection of linguistic material for the matrix sentence test given in Table

**Table 1.** Fifty-word base matrix of the Slovak matrix test

| Name | Verb | Numeral | Adjective | Noun |
|---|---|---|---|---|
| Jano | chce ‘*wants*‘ | veľa ‘*many/much*‘ | ďalších ‘*other*‘ | domov ‘*houses*‘ |
| Peter | čaká ‘*waits*‘ | tristo ‘*three hundred*‘ | nových ‘*new*‘ | lavíc ‘*benches*’ |
| Martin | dáva ‘*gives*‘ | sto ‘*hundred*‘ | celých ‘*whole*‘ | mostov ‘*bridges*‘ |
| Jožo | vidí ‘*sees*‘ | štvoro ‘*four*’ | veľkých ‘*big*‘ | lámp ‘*lamps*‘ |
| Pavol | hľadá ‘*looks for*‘ | dvesto ‘*two hundred*‘ | malých ‘*small*‘ | vedier ‘*buckets*‘ |
| Mária | drží ‘*holds*‘ | pár ‘*a few*‘ | starých ‘*old*‘ | lyžíc ‘*spoons*‘ |
| Viera | pozná ‘*knows*‘ | sedem ‘*seven*‘ | dobrých ‘*good*‘ | okien ‘*windows*‘ |
| Anna | má ‘*has*‘ | osem ‘*eight*‘ | zlých ‘*bad*‘ | budov ‘*buildings*‘ |
| Jana | berie ‘*takes*‘ | málo ‘*little/few*‘ | pekných ‘*nice*‘ | nožov ‘*knives*‘ |
| Eva | nechce ‘*doesn’t want*‘ | mnoho ‘*many/much*‘ | iných ‘*different*‘ | izieb ‘*rooms*‘ |

Table 1 presents the fifty-word base matrix of the Slovak matrix sentence test we developed. It includes ten words of five syntactic categories: personal names, verbs, numerals, adjectives and nouns. The main factor influencing the selection of lexical items was the underlying assumption that each sentence must be syntactically correct when randomly generated at the test. The words were selected on the basis of the frequency lists in the Slovak National Corpus (SNC). The frequency lists are available for individual word classes. This is one of advantages of SNC which makes SNC a well-designed, balanced and user-friendly corpus. For the matrix design the top 1000 most frequent lemma lists were used to ensure that the words are general and commonly used in basic communicative situations. Another criterion was that all selected words were semantically neutral. Stylistically and emotionally marked words were excluded.

### *Personal Names*

The first five names in Table 1 are male names and the remaining ones are female personal names. In their selection, the main criteria applied include absolute frequency value and length. Only names with maximum two syllables[1] were selected. Male names appear first in Table 1 and the reason is that their absolute frequencies were higher than female personal names, for instance, the value for *Jano* is 458 882, for *Peter* 433 677, for *Martin* 342 806, for *Jozef* 317 557, and for *Pavol* it is 240 186 whereas the frequency score for the female name Mária is 180 829 and for Eva the value is even lower, 75 727. Still, absolute frequencies were significantly higher than in the Russian matrix where the absolute frequency

1 Although the female personal name Mária with its three syllables (Má-ri-a) is an exception to this criterion, it has been included into the matrix due to its high frequency of occurrence (see below in the running text).

threshold was 2034 (Warzybok et al., 2015: 2). Warzybok et al.'s decision was based on the recent frequency dictionary of modern Russian (Sharoff, 2002). Given that Russian has a much larger word stock, defining frequent words on the basis of the threshold value of 2034 seems surprising.

*Verbs*

In Slovak, the system of conjugation is complex and often determined by gender of the subject of the sentence. This is similar in other Slavic languages. Past tense forms were excluded, because they are marked for gender. Therefore, present tense verb forms were used instead. Only disyllabic present tense verb forms were selected from the top 1000 most frequent verbs in SNC. The cut-off point was an absolute frequency higher than 25 000. The most challenging task was to select verbs with the neutral and sufficiently general meaning to ensure meaningful combinability with the numerals, adjectives and nouns. It is also interesting to note that verbs are listed in the second column although they were selected after nouns, adjectives and numerals.

*Numerals*

A basic criterion for the selection of numerals was that they had to be higher than 5. The reason is that only 5 and higher combine with nouns and adjectives in the genitive plural, for example *5 domov* '5 housesGen Pl'. The numerals 2, 3, and 4 combine with nominative plural, for instance, *3 domy* '3 housesNom Pl'. Indefinite numerals are also included. In fact, these had to be taken into account to make a list of ten numerals, which would meet the condition of being maximum disyllabic while simultaneously being higher than 5. The absolute frequency threshold for numerals was 8000.

*Adjectives*

The selection of adjectives was a challenging task. All adjectival word forms are disyllabic in genitive plural, semantically neutral yet possible in combination with nouns, resulting in grammatically correct, although not entirely predictable phrases. In addition, the set of adjectival forms had to be phonologically balanced. The absolute frequency threshold was 15 000. It is interesting to note that the above mentioned Russian matrix includes two colour adjectives, *krasnyj* 'red' and *seryj* 'gray' (Warzybok, 2015: 2). In Slovak, the equivalent for 'red' is *červený*, which consists of three syllables and therefore falls outside the criteria. The Polish matrix lists three colour adjectives, *biały* 'white', *żółty* 'yellow', and *czarny* 'black' (Ozimek et al., 2010), all similar to Slovak in their genitive plural. However, frequent disyllabic colour adjectives were excluded in Slovak either due to their difficult consonant clusters, for example *žltý* 'yellow', or because they resulted in meaningless combinations with nouns.

*Nouns*

Nouns taking the object position in generated sentences were selected prior to adjectives. Concrete and countable nouns were considered as appropriate candidates. Only disyllabic forms in the genitive plural were included. The nouns are of all three genders, three masculine nouns (*domov* 'houseGen Pl', *nožov* 'knifeGen Pl', *mostov* 'bridgeGen Pl'), five feminine (*lavíc* 'benchGen Pl', *lámp* 'lampGen Pl', *lyžíc* 'spoonGen Pl', *budov* 'buildingGen Pl', *izieb* 'roomGen Pl') and two neuter nouns (*vedier* 'bucketGen Pl', *okien* 'windowGen Pl'). The frequency threshold was 10 000.

*Phonological criteria*

The selection of words for matrix test has to follow also two phonological criteria. First, the pronunciation of words should be identical in all possible combinations, that is, it is necessary to solve the co-articulation and/or assimilation processes between the neighbouring words in a sentence. Second, the distribution of phonemes in words creating the matrix test should reflect the distribution of phonemes in a given language.

In the Slovak language, the regressive voice assimilation at word boundaries plays very important role in pronunciation. Basically, voice obstruents when followed by a voiceless sound lose their voice character and become voiceless (for example, *pod stromom* /pot stromom/) and voiceless obstruents when followed by a voiced sound gain the voice character and become voiced (for example, *vlak mešká* /vlag mešká/). Mistakes in the assimilation of voice are noticeable and are usually evaluated as errors of orthoepy (for details, see Kráľ, 2005: 53 – 62).

So as to provide the sound form of the matrix test sentences as close to the natural pronunciation as possible, we had to keep in

mind that when randomly generating the test sentences, the voiced/voiceless character of the word-final segments may change depending on the voice/voiceless nature of the following sound. Consequently, in our speech material, in the sequence *personal name – verb* no voice assimilation will take place irrespective of the combination of a name and a verb since all personal names chosen for the matrix end either in a vowel or in a sonorant (see Table 1). The voiced character of those sounds is not affected by the following sound. The combination *verb – numeral* does not cause any difficulties too since all present tense verb forms (see above) end in a vowel. Taking into account various semantic and syntactic restrictions accompanying the selection of *adjectives*, we had to find out *numerals* that – except for the combinability possibilities given by the inflectional character of the Slovak language (specified above) – would end either in a vowel or in a sonorant. The voice character of the word-initial segment of the following adjective could then be of any value. All adjectives in the matrix are in the plural accusative form that is characterized by the suffix -ý(í)ch. The phoneme /x/[2] as a voiceless obstruent changes into its voiced counterpart /h/ when followed by a voiced element. In the preliminary version of our matrix when only the semantic and frequency criteria were considered all but three nouns started in a voiced element. To preserve the uniform pronunciation of the adjective final consonant /x/ we had to replace those three nouns with nouns starting in a voiced sound. Then the pronunciation of the adjective in the combination with any noun from our matrix is with /h/.

*Phoneme distribution*[3]

In the Slovak language, there are five short vowel phonemes (i, e, a, o, u), five long vowel phonemes (i:, e:, a:, o:, u:), four diphthongs (ia, ie, iu, uo) and 27 consonant phonemes (p, b, m, f, v, t, d, n, l, r, s, z, ts, dz, c, ɟ, ɲ, ʎ, ʃ, ʒ, ʧ, ʤ, j, k, g, x, h). The graphic representation of the frequency of the occurrence of the Slovak vowel and consonant phonemes can be found in Figure 1.

2 The symbols of the IPA are used for noting down phonemes (see, e.g., Roach 2000).

3 See note 2.

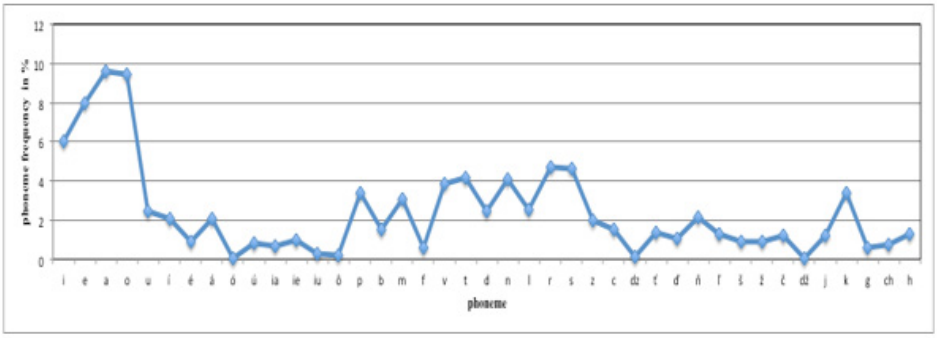


**Figure 1.** The frequency distribution of the Slovak phonemes

The frequency distribution of phonemes in the matrix test designed for Slovak (Table 1) is captured in Figure 2.

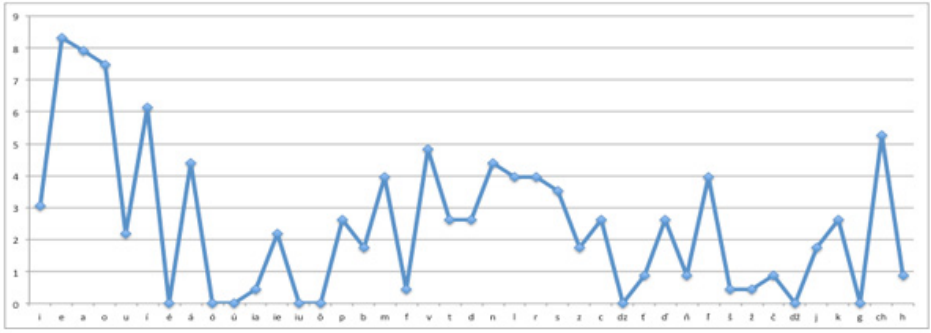

**Figure 2.** The frequency distribution of the Slovak phonemes in the matrix sentences

Figure 3 shows the comparison of the so-called reference distribution of the Slovak phonemes (Fig. 1) with the distribution of phonemes in words we have chosen for the Slovak matrix test.

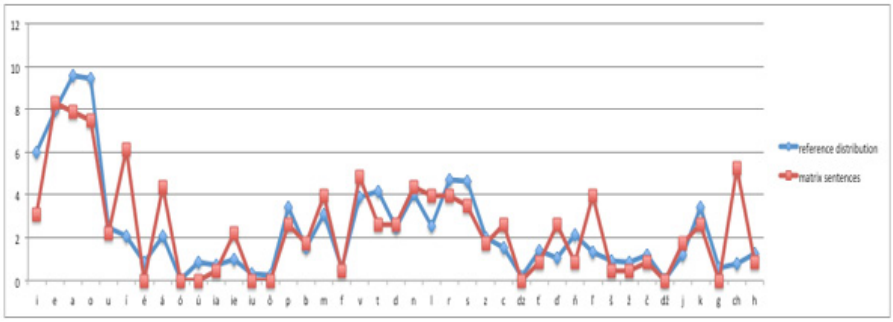


**Figure 3.** The reference distribution of the Slovak phonemes and the distribution of the Slovak phonemes in matrix sentences

As its follows from Figure 3, the occurrence of phonemes in the matrix test designed for Slovak corresponds with the general (reference) frequency distribution of Slovak phonemes. The Figure indicates higher occurrence of the phonemes /i:/, /a:/ and /x/ in the matrix sentences. This discrepancy can be easily explained by the structure of the sentences: each verb is in the 3rd person singular present tense where the suffix -á /a:/ dominates. Each adjective is in the plural accusative form ending in, as already mentioned, the suffix -ý(í)ch /i:x/[1].

## CONCLUSION AND IMPLICATIONS FOR FURTHER RESEARCH

The paper outlines the process of creating the matrix test – a diagnostic method of hearing impairment – for the Slovak language. The matrix tests for 20 languages have been developed until the present-day or are still being developing. Slovak is a language with very rich inflectional morphology and thus the preparation of a matrix consisting of 10 proper names, 10 verbs, 10 numerals, 10 adjectives and 10 nouns that, when selected randomly, would provide a meaningful and grammatically correct sentences with the structure SVO was the real challenge.

First, the selection of the proper linguistic material was based on the frequency of the occurrence in everyday communicative situations and semantic neutrality of the selected words.

Then, the matrix consisting of 50 words falling into five syntactic categories was re-evaluated so as to fulfil also phonological criteria. The result is the matrix (Table 1) enabling to produce grammatically correct and semantically unpredictable sentences whose sound form respects the natural Slovak pronunciation and the frequency distribution of the phonemes included in the selected words corresponds to the general distribution of Slovak phonemes.

Our matrix test for Slovak thus follows the recommendations of the International Collegium of rehabilitative Audiology as specified in Akeroyd et al (2015). The recommendations supplement the norm ISO 8253-3: 2012 *Acoustic-audiometric test methods*, part 3 *Speech audiometry*. In the next step, optimization and evaluation measurements will have to be carried out to measure speech intelligibility function and to establish the correct reference data, that is, the data for listeners with normal hearing. Then, we hope, the matrix test can be used in audiometry practice.

1 See Warzybok et al. 2015 for similar results in Russian.

## ACKNOWLEDGEMENTS

This work was supported by the Slovak Research and Development Agency under the Contract No APVV-15-0492 and the Contract No APVV-0077–11.

## REFERENCES

Akeroyd M. A, Arlinger S., Bentler R. A. et al. 2015. ICRA recommendations for the construction of multilingual speech tests. Int j Audiol, 54(S2), DOI 10.3109/14992027.2015.1030513.

Boboshko M, Warzybok A, Zokoll MA, Maltseva N (2013) RUMatrix test: construction, evaluation and clinical validation. Otorhinolaryngologia Hungarica. Vol. 59, N 2.-P. 49 (Abstracts of the 11th EFAS Congress Budapest).

Dietz A., Buschermöhle M., Aarnisalo A.A., Vanhanen A., Hyyrynen T. et al. 2014. The development and evaluation of the Finnish matrix sentence test for speech intelligibility assessment. Acta Otolaryngol, 134(7), 728–737.

Hagerman B. 1982. Sentences for testing speech intel-

ligibility in noise. Scandinavian Audiology, 11, 79–87.
Hochmuth S., Brand, T., Zokoll M.A., Zenker Castro F., Wardenga N. Et al. 2012. A Spanish matrix sentence test for assessing speech reception thresholds in noise. Int J Audiol, 51, 536–544.
Houben R., Koopman J., Luts H., Wagener K.C., van Wieringen A. et al. 2014. Development of a Dutch matrix sentence test to assess speech intelligibility in noise. Int J Audiol, 53, 760–763
Jansen S., Luts H., Wagener K.C., Kollmeier B., Del Rio M. et al. 2012. Comparison of three types of French speech-in-noise tests: A multi- center study. Int J Audiol, 51, 164–173.
Kráľ, Á. Pravidlá slovenskej výslovnosti. Systematika a ortoepický slovník. Matica slovenská 2005.
Kollmeier B. & Wesselkamp M. 1997. Development and evaluation of a sentence test for objective and subjective speech intelligibility assessment. J Acoust Soc Am, 102, 1085–1099.
Kollmeier B., Warzybok A., Hochmuth S., Zokoll M., Uslar et al (2015). The multilingual matrix test: Principles, applications and comparison across languages – a review. Accepted for publication in Int J Audiol, DOI 10.3109/14992027.2015.1020971
Luts H., Boon E., Wable J. & Wouters J. 2008. FIST: A French sentence test for speech intelligibility in noise. Int J Audiol, 47, 373–374.
Nielsen J.B. & Dau T. 2011. The Danish hearing in noise test. Int J Audiol, 50, 202–208.
Nilsson M., Soli S.D. & Sullivan J.A. 1994. Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. J Acoust Soc Am, 95, 1085–1099.
Plomp R. & Mimpen A.M. 1979. Improving the reliability of testing the speech reception threshold for sentences. Audiol, 18, 43–53.
Roach, P. 2000. English Phonetics and Phonology. Cambridge: Cambridge University Press.
Ozimek E, Warzybok A, Kutzner D (2010). Polish sentence matrix test for speech intelligibility measurement in noise, Int. J. Audiol. 49:444-454.
Slovenský národný korpus – prim-7.0-public-all. [Slovak national corpus – prim-7.0-public-all.] Bratislava: Jazykovedný ústav Ľ. Štúra SAV 2015. Available at WWW: http://korpus.juls.savba.sk.
Štefánik, J. – Rusko, M. – Považanec, D. 1999. Frekvencia slov, grafém, hlások a ďalších elementov slovenského jazyka. Jazykovedný časopis, 50, 1999, No. 2, pp. 81 – 93.
Versfeld N.J., Daalder L., Festen J.M. & Houtgast T. 2000. Method for the selection of sentence material for efficient measurement of the speech reception threshold. J Acoust Soc Am, 107, 1671–1684.
Wagener K., Brand T. & Kollmeier B. 1999a. Entwicklung und Evaluation eines Satztests in deutscher Sprache Teil II: Optimierung des Oldenburger Satztests (in German). (Development and evaluation of a German sentence test, Part II: Optimization of the Oldenburg sentence tests). Z Audiol, 38, 44–56.
Wagener K., Brand T. & Kollmeier B. 1999b. Entwicklung und Evaluation eines Satztests für die deutsche Sprache Teil III: Evaluation des Oldenburger Satztests (in German). (Development and evaluation of a German sentence test – Part III: Evaluation of the Oldenburg sentence test). Z Audiol, 38, 86–95.
Wagener K., Kühnel V. & Kollmeier B. 1999c. Entwicklung und Evaluation eines Satztests in deutscher Sprache I: Design des Oldenburger Satztests (in German). (Development and evaluation of a German sentence test – Part I: Design of the Oldenburg sentence test). Z Audiol, 38, 4–15.
van Wieringen A. & Wouters J. 2008. LIST and LINT: Sentences and numbers for quantifying speech understanding in severely impaired listeners for Flanders and the Netherlands. Int J Audiol, 47(6), 348–355
Warzybok, A., Zokoll, M., Wardenga, N.,, Ozimek, E., Boboshko, M. & Kollmeier, B. (2015) Development of the Russian matrix sentence test, International Journal of Audiology, 54:sup2, 35-43, DOI: 10.3109/14992027.2015.1020969
Wong L.L.N. & Soli S.D. 2008. The Cantonese version of the hearing in noise test (CHINT). Int J Audiol, 47(6), 388–390.