AJA

Research Forum

Speech-in-Noise Tests for Multilingual Hearing Screening and Diagnostics[1]

Melanie A. Zokoll,[a,b] Sabine Hochmuth,[a,b] Anna Warzybok,[a,b] Kirsten C. Wagener,[b,c] Michael Buschermöhle,[b,d] and Birger Kollmeier[a,b,c,d]

Purpose: New complementary multilingual speech-in-noise tests in Russian, Turkish, and Spanish for hearing self-screening purposes and follow-up hearing diagnostics are compared to the speech tests of the European project, HearCom (Hearing in the Communication Society).
Method: The tests consist of spoken numbers (Digit Triplet Test; Smits, Kapteyn, & Houtgast, 2004) or sentences (Matrix Test; e.g., Hagerman, 1982) presented in a background noise and estimate the speech reception threshold, which is the signal-to-noise ratio that yields 50% speech intelligibility. All tests were developed according to the HearCom minimum quality standards for speech intelligibility tests. This report presents a cross-language comparison of reference speech intelligibility functions for monaural headphone measurements with normal-hearing listeners. The same model function was employed to describe the speech intelligibility functions for all of the tests.
Results: Reference speech intelligibility functions of the new versions of the Digit Triplet Test and Matrix Test show high comparability to the HearCom tests. In order to achieve the highest possible comparability across languages, language- and speaker-dependent factors in speech intelligibility should be compensated for.
Conclusion: To date, several complementary tests for screening and diagnostics have been developed in several languages. Adhering to the HearCom standards, the tests are highly comparable across languages. For the Matrix Test, equal syntax and linguistic complexity were maintained across languages due to common methodological standards.

Key Words: hearing screening, speech-in-noise test, HearCom, speech perception, Digit Triplet Test, Matrix Test, speech reception threshold (SRT)

Speech audiometry for hearing screening, diagnostics, and rehabilitation needs to be performed in the listener's native language. If tests are measured in noise to obtain results in authentic listening situations, nonnative listeners typically perform worse than native listeners with the same hearing ability (e.g., van Wijngaarden, Steeneken, & Houtgast, 2002; Warzybok, Wagener, Brand, & Kollmeier, 2010). Multilingual versions of a test provide the possibility to test nonnative listeners more exactly and, additionally, to compare results internationally. Thus, we wanted to develop comparable versions of two speech audiometric tests in different languages. We focused on (a) the Digit Triplet Test as a fast, reliable, and anonymous self-screening hearing test (Smits, Kapteyn, & Houtgast, 2004), and (b) the Matrix Test for use by professionals (e.g., Hagerman, 1982; Wagener, Brand, & Kollmeier, 1999a, 1999b; Wagener, Kühnel, & Kollmeier 1999).

[a]Carl von Ossietzky University Oldenburg, Oldenburg, Germany
[b]Centre of Excellence "Hearing4all", Oldenburg, Germany
[c]Hörzentrum OldenburgGmbH, Oldenburg, Germany
[d]HörTech gGmbH, Oldenburg, Germany

Correspondence to Melanie Zokoll-van der Laan: melanie.zokoll@uni-oldenburg.de

Editor: Larry Humes

Received November 2, 2012
Accepted March 5, 2013
DOI: 10.1044/1059-0889(2013/12-0061)

Digit Triplet Test

The Digit Triplet Test was originally developed for screening hearing via telephone by Smits et al. (2004). It can also be administered via Smartphone (e.g., in German: HearContrOl, HörTech gGmbH) or Internet (e.g., Smits, Merkus, & Houtgast, 2006; see also www.hearcom.eu). The test consists of the announcement "the digits" (except the Dutch, Greek, and Polish version) followed by pseudorandom digit triplets (e.g., 1-3-6) presented in a background of noise with the same long-term spectrum as the speech material. The participant's task is to select the digits heard on a number pad (i.e., the administration can be automated). Using a 1-up/1-down adaptive procedure with triplet scoring (i.e., all three digits have to be correct for a correct response), the speech reception threshold (SRT; i.e., the signal-to-noise ratio [SNR] that yields 50% speech intelligibility) is estimated. However, the procedural details vary slightly between languages (e.g., noise or speech level fixed,

[1]Part of this work was presented at AHS 2012, the 2nd International Conference on Adult Hearing Screening, June 7–9, 2012, in Cernobbio (Lake Como), Italy.

presentation level, step size rules in the 1-up/1-down procedure, and the SRT estimation), which was shown to have only marginal influence on SRT estimates (Hagerman, 1982; Morgan, 2010; Wagener, 2004). One measurement (i.e., one list of 21 to 30 digit triplets depending on the language) lasts 3–6 min.

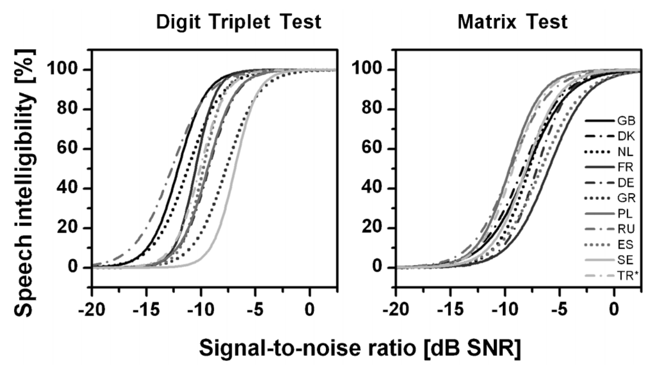
After HearCom (Hearing in the Communication Society), a European research project with 28 partners cooperating on several topics of research in auditory communication (e.g., diagnostics, screening, and web services for people with hearing impairment; Vlaming et al., 2011), the Digit Triplet Test was made available in several European languages (e.g., British English: Hall, 2006; Dutch: Smits et al., 2004; French: Jansen, Luts, Wagener, Frachet, & Wouters, 2010; German: Zokoll, Wagener, Brand, Buschermöhle, & Kollmeier, 2012; Polish: Ozimek, Kutzner, Sek, & Wicher, 2009; Greek and Swedish: unpublished data). Tests were either selected or developed to fulfill minimum quality standards regarding recording, optimization, and evaluation defined by the HearCom project members in order to reach the highest comparability across countries (e.g., HearCom Deliverable D-1-3, 2006). Those requirements were partially implemented in the International Standards Organization (ISO) standards for speech audiometry (ISO, 2012). The current contribution adds test versions for three additional languages in accordance to these HearCom standards: Russian (Warzybok et al., 2011), Spanish (Hochmuth, Zokoll, Brand, & Kollmeier, 2010), and Turkish (Zokoll, Hochmuth, et al., 2012).

Matrix Test

The Matrix Test is a (professional) speech audiometric tool that further assesses hearing deficiencies encountered in noisy environments. Using an adaptive procedure and word scoring (for details, see Brand & Kollmeier, 2002), the Matrix Test determines the patient's SRT in noise (or quiet). The test consists of syntactically fixed but semantically unpredictable sentences (e.g., "Steven prefers three large windows," see also Hagerman, 1982; Wagener, Kühnel, & Kollmeier, 1999). Sentences are generated from words taken in a seemingly random fashion from an inventory of 50 words (10 alternatives per word group, i.e., name, verb, numeral, adjective, and object). The phoneme distribution corresponds to the respective language. After a training session of two measurements, the Matrix Test can be used repeatedly with the same patient because the sentences are difficult to memorize. One measurement (i.e., one list of 20 sentences) lasts 2–4 min. In addition to the standard open format where listeners repeat what they understood, the test can be performed in a closed format where listeners select the words they understood out of the inventory of 50 words displayed on a screen, making it suitable for use with patients of a different native language than the audiometrist or audiologist, but also for automating.

In analogy to the Digit Triplet Test and in accordance with the HearCom standards, the current work added test versions for Russian (Warzybok et al., 2011), Spanish (Hochmuth et al., 2012), and Turkish (Zokoll, Hochmuth, et al., 2012) to the collection of existing Matrix Tests in British English (Hall, 2006), Danish (Wagener, Josvassen, & Ardenkjaer, 2003), Dutch (Houben et al., 2013), French (Jansen et al., 2012), German (Wagener et al., 1999a, 1999b, 1999), Polish (Ozimek, Warzybok, & Kutzner, 2010), and Swedish (Hagerman, 1982).

Figure 1. Reference speech intelligibility functions for the Digit Triplet Test (left graph) and the Matrix Test (right graph) in several languages obtained in headphone measurements with normal-hearing listeners. The languages are labeled with their international acronym. For Danish, no version of the Digit Triplet Test is available. For Greek, no version of the Matrix Test is available. The asterisk for the Turkish tests (TR) indicates preliminary data.



Cross-Language Comparison

Within the limits of comparability across languages and speakers, all speech tests considered here provide a highly comparable test format, recording and optimization guidelines, and evaluation standards. In the following paragraph, different versions of the tests are compared by examining the reference speech intelligibility functions determined with normal-hearing listeners in speech intelligibility measurements at several (fixed) SNRs. The logistic function[2] (Equation 1) is fitted to the measured intelligibility scores, and the parameters describing the fitted function (SRT and slope) act as the reference values. The logistic model function is as follows:

$$SI(SNR) = \frac{1}{1 + e^{4s(SRT-SNR)}} \qquad (1)$$

where SI = speech intelligibility, SNR = the signal-to-noise ratio of the presentation, SRT = the mean speech intelligibility threshold, and s = the slope at the SRT.

The left graph of Figure 1 depicts reference speech intelligibility functions of the Digit Triplet Test for Russian (RU), Spanish (ES), and Turkish (TR) in comparison to the HearCom tests for British English (GB, Morgan, 2010),

[2]For the Greek and Polish versions of the Digit Triplet Test, the cumulative normal distribution function was used instead of the logistic model function. However, both functions lead to similar parameter estimations.

Dutch (NL, Smits et al., 2004), French (FR, Jansen et al., 2010), German (DE, Zokoll, Wagener, et al., 2012a), Greek (GR, Vlaming, personal communication, 2009), Polish (PL, Ozimek et al., 2009), and Swedish (SE, Hällgren & Larsby, personal communication, 2008). All data were obtained monaurally with normal-hearing listeners using different types of free-field equalized headphones.

The Russian, Spanish, and Turkish versions of the Digit Triplet Test show comparable speech intelligibility functions to the HearCom tests. In general, the speech materials for all languages show steep speech intelligibility functions with slopes in the range of 14.8 (RU) to 27.1%/dB (FR; mean 20%/dB), permitting precise SRT estimates. SRTs range from –6.9 (SE) to –12.7 dB SNR (RU; mean = –10 dB SNR). The British English and Russian versions of the Digit Triplet Test have considerably higher speech intelligibilities in comparison to the other versions at comparable SNRs. For the Swedish and Greek versions, the opposite is true. Because the slopes of the speech intelligibility functions for those tests are comparable to the other tests, these differences could be compensated for by corrections.

The right graph of Figure 1 displays reference intelligibility functions of the Matrix Test for Russian (RU), Spanish (ES), and Turkish (TR) versions in comparison to the existing Matrix Tests in British English (GB, Hewitt, 2008), Danish (DK, Wagener et al., 2003), Dutch (NL, Houben et al., 2013), French (FR, Jansen et al., 2012), German (DE, Wagener et al., 1999b), Polish (PL, Ozimek et al., 2010), and Swedish (SE, Hagerman, 1982, data reanalyzed by Wagener, 2004). Reference speech intelligibility functions for the Russian, Spanish, and Turkish versions of the Matrix Test are very similar to the speech intelligibility functions of the HearCom tests. Generally, the slopes are less steep than found for the Digit Triplet Test and range from 13.0 (GB) to 17.1%/dB (DE and PL; M = 15%/dB).

Differences between the slopes obtained for the different languages might result from both differences in linguistic structure and across speakers. The SRTs of the functions show fewer differences and range from –6.0 (FR) to –9.6 dB SNR (PL; M = –8.0 dB SNR). Again, Russian shows higher speech intelligibilities compared to most of the other versions of the Matrix Test at comparable SNRs, but is in line with findings for Polish, which is the other Slavic language. Due to the methodological similarities of these versions of the Matrix Test (similar to the Digit Triplet Test), these language-specific differences in reference functions must be considered language (and also speaker) dependent.

Conclusion

The collection of speech-in-noise tests selected or developed within the HearCom project was extended by three versions of the Digit Triplet Test and Matrix Test, respectively. All tests fulfill the HearCom standards and thus allow for precise SRT estimates in noise as well as high comparability across languages. We suggest using the standards whenever developing similar screening or diagnostic tests in other languages. This leads to good test–retest reliability (or within-subject variability) calculated as root mean square of the within-subject standard deviations of repeatedly measured adaptive SRTs (≤1 dB, following two training lists if using the Matrix Test; e.g., Jansen et al., 2012, Wagener et al., 1999b). Additionally, there seems to be a high correlation of Digit Triplet and Matrix Tests (Jansen et al., 2012, Pearson product–moment correlation of e.g., 0.81 for listeners with hearing impairment), which suggests that the tests are indeed complementary tools for hearing screening and follow-up diagnostics.

Acknowledgments

This work was supported by the EFRE-project HurDig. The contributions to the test arsenal extension of the scientists of the following partner sites are gratefully acknowledged: Germany, University of Oldenburg, Institute of Slavonic studies; Spain, Clínica Barajas; and Turkey, Ankara University, Department of Linguistics.

References

Brand, T., & Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *Journal of the Acoustical Society of America, 111,* 2801–2810.

Hagerman, B. (1982). Sentences for testing speech intelligibility in noise. *Scandinavian Audiology, 11,* 79–87.

Hall, S. (2006). *The development of a new English sentence in noise test and an English number recognition test* (Unpublished master's thesis). University of Southampton, UK.

HearCom Deliverable D-1-3. (2006). *Protocol for implementation of communication tests in different languages*. Retrieved from http://hearcom.eu/about/DisseminationandExploitation/deliverables.html.

Hewitt, D. R. (2008). *Evaluation of an English speech-in-noise audiometry test* (Unpublished master's thesis). University of Southampton, UK.

Hochmuth, S., Brand, T., Zokoll, M. A., Zenker Castro, F., Wardenga, N., & Kollmeier, B. (2012). A Spanish matrix sentence test for assessing speech reception thresholds in noise. *International Journal of Audiology, 51,* 536–544.

Hochmuth, S., Zokoll, M., Brand, T., & Kollmeier, B. (2010). Closed-set intelligibility tests for a multilingual society: A new Spanish Digit Triplet and Sentence Speech Test [Abstract]. *Association for Research in Otolaryngology*, *33*, 149. Abstract retrieved from http://www.aro.org/abstracts/2010_Abstract_Book_REV.pdf.

Houben, R., Koopman, J., Luts, H., Wagener, K. C., van Wieringen, A., & Dreschler, W. A. (2013). *Development of a Dutch matrix sentence test to assess speech intelligibility in noise*. Manuscript submitted for publication.

International Standards Organization. (2012). *Acoustics: Audiometric test methods: Part 3: Speech audiometry* (ISO 8253-3:2012). Genève, Switzerland: Author.

Jansen, S., Luts, H., Wagener, K. C., Frachet, B., & Wouters, J. (2010). The French Digit Triplet Test: A hearing screening tool for speech intelligibility in noise. *International Journal of Audiology, 49,* 378–387.

Jansen, S., Luts, H., Wagener, K. C., Kollmeier, B., Del Rio, M., Dauman, R., … van Wieringen, A. (2012). Comparison of three types of French speech-in-noise tests: A multi-center study. *International Journal of Audiology, 51,* 164–173.

Morgan, D. (2010). *Normative data for the UK Triplet Test in stationary and time varying noise* (Unpublished master's thesis). University of Southampton, UK.

Ozimek, E., Kutzner, D., Sek, A., & Wicher, A. (2009). Development and evaluation of Polish Digit Triplet Test. *Speech and Communication, 51*(4), 307–316.

Ozimek, E., Warzybok, A., & Kutzner, D. (2010). Polish sentence matrix test for speech intelligibility measurement in noise. *International Journal of Audiology, 49,* 444–454.

Smits, C., Kapteyn, T., & Houtgast, T. (2004). Development and validation of an automatic speech-in-noise screening test by telephone. *International Journal of Audiology, 43,* 15–28.

Smits, C., Merkus, P., & Houtgast, T. (2006). How we do it: The Dutch functional hearing-screening tests by telephone and Internet. *Clinical Otolaryngology, 31,* 436–440.

van Wijngaarden, S. J., Steeneken, H. J. M., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *Journal of the Acoustical Society of America, 111,* 1906–1916.

Vlaming, M. S. M. G., Kollmeier, B., Dreschler, W. A., Martin, R., Wouters, J., Grover, B., … Houtgast, T. (2011). HearCom: Hearing in the communication society. *Acta Acustica united with Acustics, 97*(2), 175–192.

Wagener, K. C. (2004). *Factors influencing sentence intelligibility in noise* (Unpublished doctoral dissertation). University of Oldenburg, Germany.

Wagener, K., Brand, T., & Kollmeier, B. (1999a). Entwicklung und Evaluation eines Satztests in deutscher Sprache Teil II: Optimierung des Oldenburger Satztests [Development and evaluation of a German sentence test—Part II: Optimization of the Oldenburg sentence test]. *Zeitschrift für Audiologie, 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 [Development and evaluation of a German sentence test—Part III: Evaluation of the Oldenburg sentence test]. *Zeitschrift für Audiologie, 38,* 86–95.

Wagener, K., Josvassen, J. L., & Ardenkjaer, R. (2003). Design, optimization, and evaluation of a Danish sentence test in noise. *International Journal of Audiology, 42,* 10–17.

Wagener, K., Kühnel, V., & Kollmeier, B. (1999). Entwicklung und Evaluation eines Satztests in deutscher Sprache I: Design des Oldenburger Satztests [Development and evaluation of a German sentence test—Part I: Design of the Oldenburg sentence test]. *Zeitschrift für Audiologie, 38,* 4–15.

Warzybok, A., Wagener, K. C., Brand, T., & Kollmeier, B. (2010). Influence of linguistic complexity in relation to speech material on non-native speech perception in noise. In *Proceedings of the DAGA 2010, 36th German Annual Conference on Acoustics* (ISBN 978-3-9808659-8-2). Berlin, Germany.

Warzybok, A., Zokoll, M., Brandes, O., Ozimek, E., Brand, T., & Kollmeier, B. (2011). It's all Russian to me: Development of two closed-set Russian speech intelligibility tests that can be administered even by non-Russian audiologists [Abstract]. *Journal of Hearing Science, 1,* 115. Abstract retrieved from http://www.efas2011.org/download.php.

Zokoll, M. A., Hochmuth, S., Fidan, D., Wagener, K. C., Ergenç, İ., & Kollmeier, B. (2012). Speech intelligibility tests for the Turkish language [Abstract]. In *Proceedings of the 15th Annual Conference of the German Audiology Society* (ISBN 978-3-9813141-2-0). Erlangen, Germany.

Zokoll, M. A., Wagener, K. C., Brand, T., Buschermöhle, M., & Kollmeier, B. (2012). Internationally comparable screening tests for listening in noise in several European languages: The German Digit Triplet Test as an optimization prototype. *International Journal of Audiology, 51,* 697–707.