Development of new procedures for the Digit Triplets Test as a reliable screening tool

Löw, Thomas; Bach, Jörg-Hendrik; Buschermöhle, Michael; Bisitz, Thomas

HörTech gGmbH, Oldenburg

Cluster of Excellence Hearing4all, Oldenburg, Germany

Keywords: Digit Triplets Test, screening, speech in noise test

Introduction

The Digit Triplets Test (DTT) is a speech in noise hearing test using three-digit sequences as speech material. It was first published in 2004 (Smits et al, 2004) and is available in a variety of languages. Since speech reception threshold (SRT) estimation in speech-in-noise hearing tests is relatively robust against changes in presentation level (Plomp and Mimpen, 1979), no exact calibration or sound proof booth is needed. Small training effect and the closed response setup allow the DTT to be used as a self test screening tool with consumer equipment (e. g. via telephone or with headphones via the internet). The most prevalent implementations of the test procedure take about five minutes per ear. While appropriate in a scientific setting, this may be too long in a setting aiming at a quick screening of hearing function. We introduce two new test procedures to reduce the measurement time: An adaptive procedure using maximum likelihood estimation, which results in an SRT estimate, and a second procedure consisting of a starting phase followed by repeated measurements at a fixed signal-to-noise ratio, which gives a simple pass-or-fail result.

Description of test procedures

The default test procedure (TP I) for the DTT is an adaptive one-up-one-down procedure with 2 dB step width. The German implementation uses triplet scoring and 27 trials (Zokoll et al, 2012). This means that the signal-to-noise-ratio (SNR) is decreased by 2 dB, if all three digit responses are correct, and increased by 2 dB otherwise. The start SNR is set to 4 dB. The measurement result is the mean value of the SNRs of the last 20 presentations, assuming convergence after the first 8 presentations.

The first proposed test procedure (TP II) is also an adaptive procedure. It uses a maximum likelihood estimator (MLE), to find the test subject’s most probable psychometric function with respect to their responses. The psychometric function calculates the probability p for a correct digit response at a given SNR and is given by

$$p = c + (1 - c) \cdot \left(1 + \exp\left(4s(\mathrm{SRT} - \mathrm{SNR})\right)\right)^{-1} \tag{1}$$

where c = 0.1 is the chance level to guess the right digit, s describes the slope of the psychometric function and SRT is the signal-to-noise ratio, for which p = 0.5. The slope parameter was fixed at s = 0.145 dB^{-1}, which corresponds to the slope of the psychometric function of a normal hearing subject for the German DTT measured via telephone (Zokoll et al, 2012). TP II uses the MLE for two purposes. The first is level control. For each trial the SNR is set to the maximum likelihood SRT as computed from the three previous trials. For this purpose possible SRT values on a grid with 0.5 dB resolution in a range of 4 dB around the SNR of the previous trial (maximum value of +10 dB) are used. The second is the SRT calculation for the measurement result. Here all trials and possible SRT values on a grid with 0.1 dB resolution in the range between -14 dB and +10 dB are used. The start SNR for this procedure was set to 0 dB.

The second proposed test procedure (TP III) is not an adaptive procedure. Instead the main idea is very similar to the one introduced by Smits (2017). The procedure does not estimate the SRT. Instead it determines, if the SRT is below or above a critical value, producing a pass-or-fail result, respectively. The SNR of the first presentation is 8 dB above this threshold. As long as not all three digit responses are correct, the SNR is increased by 4 dB to a maximum of 16 dB above threshold. After that the SNR is decreased by 4 dB per presentation up to the threshold value. From then on the presentation level remains constant until either the maximum number of trials (set to 20) is reached or the abort criterion for pass or fail is fulfilled. The abort criterion for pass consists of two parts and only the trials with SNRs at threshold are considered. At least four of these trials have to be performed and the probability for a subject with SRT at threshold to perform as good as or worse than observed has to be smaller than a given uncertainty level (set to 5 %). The probability for a correct digit response of a subject with SRT at threshold is 55 %. The probability for an arbitrary number of correct digit responses is given by a Binomial distribution with parameters $p_0 = 0.55$ and N = 3 * number of considered trials. From this the desired probability can be computed.

The abort criterion for fail follows the same idea: At least four trials have to be considered and the probability for a subject with SRT at threshold to perform as good as or better than observed has to be smaller than a given uncertainty level (set to 5 %). This results in a Binomial distribution as above but with $p_0 = 0.45$ and counting incorrect digit responses. Again, only trials with SNR at threshold are considered. If the abort criterion is not fulfilled, earlier trials are added to consideration successively starting with the last one and the abort criterion is evaluated each time again. The SNR value of these trials is ignored. If the maximum number of trials is reached, but no abort criterion is valid, there are two possibilities. The test result can either be "uncertain" or the percentage of correct digit responses at threshold is compared to the reference value of 55 %, resulting in a "pass" or "fail" result.

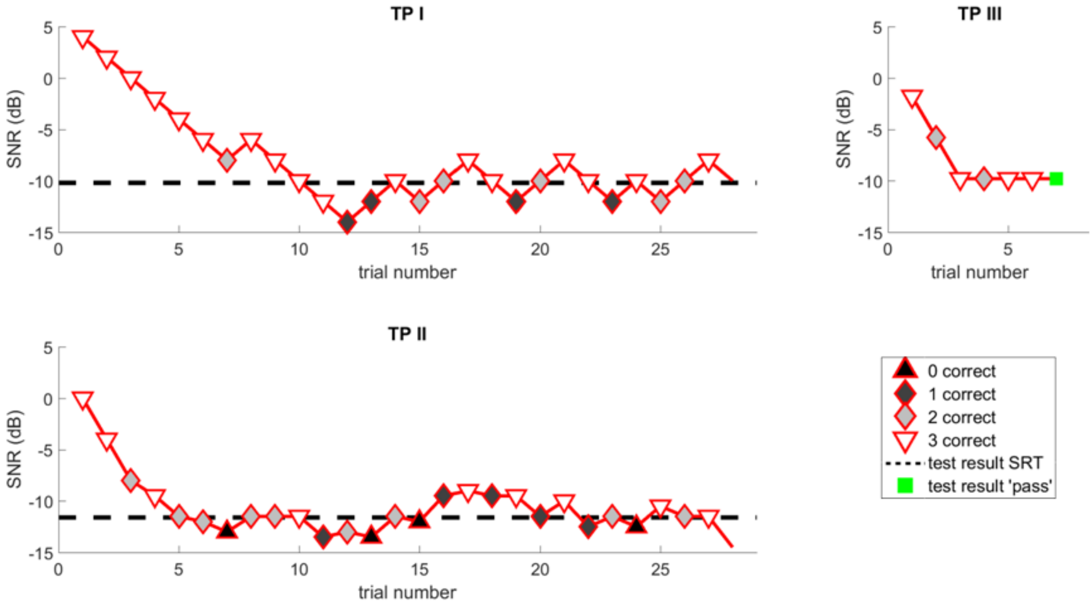

Figure 1: Possible trends of measurements with different test procedures (simulations).

Discussion of test procedures

TP I has the advantage that it is very easy to implement, but it has three major drawbacks regarding precision and efficiency:

i. The first trials do not contribute to the measurement result.
ii. The test does not differentiate between trials with no, one or two correctly recognized digits. This additional information is ignored.
iii. Fixed step size. A larger step size would minimize the number of trials needed, until the SNR has reached a value near the subject's SRT. A smaller step size later on would lead to a more precise SRT measurement.

TP II tries to overcome these disadvantages.

i. The usage of the MLE instead of calculation of the mean of presentation SNRs allows exploiting all test data.
ii. The test takes into account each single digit, so no data is ignored.
iii. The step size ranges between 0 dB and 4 dB. This will usually lead to a larger step size in the beginning of the measurement, when the SNR is far from the subject's SRT, and to a smaller step size, when the SNR has reached values near the SRT.

The reduction of step size has to be treated with caution. If the step size is reduced early during the measurement, and the SNR is not yet near the SRT, and is never increased again, the measurement will yield a poor result. This happens for example, if the test subject needs time to adapt to the task and underperforms in the first trials. This is the reason why the MLE of TP II uses only the data of the last three trials for level control. The step size is not only decreased, but dynamically decreases or increases again, which makes the test procedure more robust. Using three trials results in a good compromise between reduction of step size and therefore higher precision on the one hand and robustness on the other.

TP III follows a completely different approach. The main reason for adaptive procedures is that SRT estimation is more precise the closer SNR and SRT. But in order to get a pass-or-fail result no precise SRT estimate is needed. Instead the repeated measurement at threshold increases efficiency in determining, if the SRT is below or above threshold. The downside is, that very little information is gained about how much the SRT differs from threshold.

The first few trials in TP II above threshold do in fact decrease the efficiency of the test procedure. But they allow the subject adapting to the task. By provoking a three-digits-correct response in the beginning it is ensured, that the subject has understood the task. Another difference to the previous test procedures is the variable measurement duration. The introduction of the third category "uncertain" reduces false-positive or false-negative test results and its usage as well as the other test parameters can be adjusted to the purpose of application.

Performance of test procedures in computer simulations

In order to investigate the efficiency and accuracy of the test procedures, Monte Carlo simulations were applied. For this purpose test subjects were modeled by parameters SRT and s. The probability for a correct digit response was then given by equation (1). 50,000 Monte Carlo runs were performed. The parameters were set to SRT = -12 dB and s = 0.17 dB^{-1}, which correspond to the mean values of the German digit triplets test for normal hearing listeners (Zokoll et al, 2012).

As a measure for the precision of the test results for the adaptive procedures the standard error of SRT results was calculated. As both procedures did not show any bias, this equals the standard deviation. To analyze the efficiency, the number of trials per test procedure was varied between 10 and 27. Then the number of trials was determined, after which TP II reached the same accuracy as TP I after 27 trials. The results are shown in figure 2.

With the new test procedure the measurement can be completed after 16 instead of 27 trials without loss of accuracy, i.e. the measurement time is reduced to approximately 60 percent.

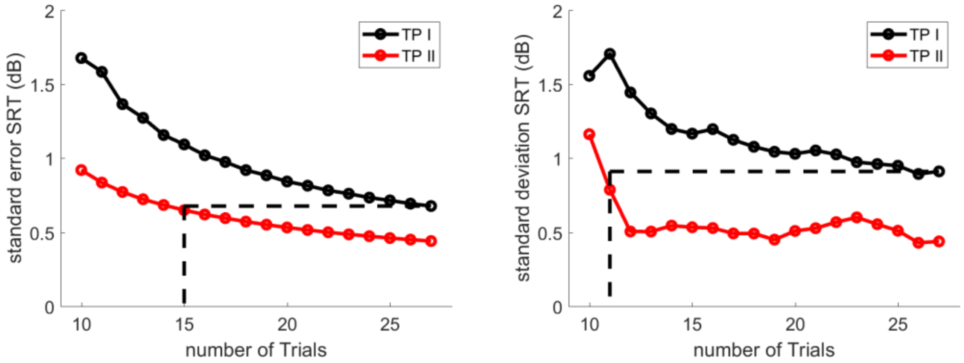

Figure 2: Performance of adaptive procedures in Monte Carlo simulations (left) and measurements (right.) The possible reduction of trial number without loss of accuracy is indicated.

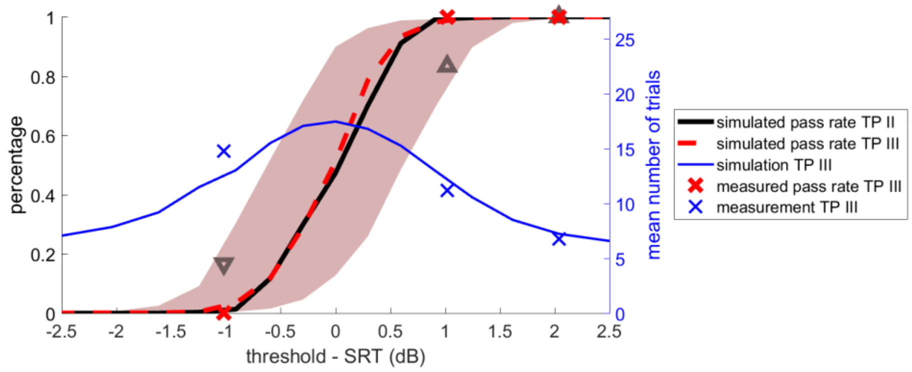

Figure 3: Performance of pass-or-fail test procedures in Monte Carlo simulations and measurement. Pass rates and in the case of TP III mean trial number are indicated. The hatched area indicates the rate of result "uncertain"*, the triangles corresponding measurement data.*

To investigate the performance of TP III Monte Carlo simulations were run using different threshold values. In figure 3 the relation between SRT-to-threshold difference and the percentage of "pass" results is shown. The test result "uncertain" was not used in this case. This is compared to the pass-or-fail result determined by comparing threshold value to the test results of TP II. Both procedures produce very similar results. But in TP II 27 trials are performed, while in TP III only 6 to 15 trials are needed. Therefore, the measurement time can be reduced to 20 – 74 percent.

Performance of test procedures in measurements

Measurement data were obtained in 12 normal-hearing listeners. Measurements were made in a sound-proof booth using Sennheiser HDA200 circumaural headphones. The test subjects were aged between 19 and 28. They were tested for normal-hearing and their thresholds of all octave frequencies between 250 and 8000 Hz had been below 20 dB HL. Subjects received an hourly compensation for taking part in the study. The three test procedures were employed in pseudo randomized order and repeated twice. In these measurements the threshold for TP III was set to -9.8 dB. This value is the normal hearing mean SRT plus three standard deviations, as determined in pilot measurements. Directly afterwards a maximum likelihood estimator was used to determine the subject‘s SRT taking into account all data points of all six measurements, in order to get the best possible SRT estimation. Two more measurements were taken using TP III with thresholds 1.0 dB below and above the subject’s SRT. This corresponds to intelligibility levels of 33 % and 67 %, assuming a slope parameter of $s = 0.17\ dB^{-1}$.

To analyze precision and efficiency of the adaptive test procedures we proceeded as in the previous section. In this case the variability in test results consists of the intra-subject variability and the inter-subject variability. The former corresponds to the measurement precision. The latter is independent of the test procedure and therefore does not distort the results. The results are shown in figure 2 and do not contradict the conclusions regarding the reduction of measurement time drawn in the previous section.

All measurements with procedure TP III with threshold at -9.8 dB and at 67 % intelligibility level yielded the result “pass”. All measurements at 33 % intelligibility level yielded the result “fail”, as expected. This is in good agreement with the computer simulations. The mean number of trials is shown in figure 3. For the measurements with threshold value below SRT it is slightly higher than the computer simulations predict. But because of the small number of test subjects this difference is not significant.

Discussion

As Smits and Festen have pointed out (2011), the test precision alone may not be a meaningful factor when comparing SRT measurements with different scoring methods (as with TP I and TP II). However, we can easily modify TP II to measure the triplet SRT by parametrizing the psychometric function not by the digit SRT (i. e. the 55 % correct response level), but by the 79 % correct response level (79 % of correct digit responses result in 50 % of correct triplet responses). Although this will have an effect on SNR placement and result calculation, first examination show that it has a negligible effect on test precision. In this regard further study is required.

Both in computer simulations and measurements only normal-hearing subjects were considered. Further study regarding subjects with different levels of hearing impairment is required.

Conclusion

We have introduced two new test procedures for the digit triplets test. An adaptive procedure, that reduces measurement time to about 60 percent while maintaining the same accuracy. The procedure using a fixed SNR (after a starting phase) gives a simple pass-or-fail result and reduces the number of trials even more, which makes it suitable as a quick and reliable screening tool.

Literature

Plomp R. & Mimpen A. M. (1979). Speech-reception threshold for sentences as a function of age and noise level. J Acoust Soc Am, 66(5), 1333-42

Smits C. (2017). Improving the Efficiency of Speech-In-Noise Hearing Screening Test. Ear & Hearing, 38(6), 385-388

Smits C. & Festen J. M. (2011). The interpretation of speech reception threshold data in normal-hearing and hearing-impaired listeners: Steady-state noise. J Acoust Soc Am, 130(5), 2987-98

Smits C., Kapteyn T. & Houtgast T. (2004). Development and validation of an automatic speech-in-noise screening test by telephone. Int J Audiol, 43, 15-28

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. Int J Audiol, 51, 697-707