Outline 3 Discussion - Short answers **Hanneke Snieders**

Working title: Using a simulation test to assess warning signal detection in the real-world noise environment of Dutch train cabins

Discussion

What are the most important findings? Are the two hypotheses confirmed or not?

The aim of this study was to develop a signal detection test, based on the real-world acoustic environment of Dutch train cabins and describe its reproducibility. The overall reproducibility was found to be good, but varied widely when using the background noise of different train types. The first hypothesis was that the association between tone audiometry and the signal detection test would be the highest for pure tone thresholds at the frequencies at which the warning signal shows the most prominent signal peaks, which turned out to be the case. The second hypothesis was that a larger percentage of variance could be explained when the pure tone thresholds would be combined with the results of the hearing-in-noise test, but this turned out not to be the case in most test conditions.

Regarding the first experiment (reproducibility of the test), how can the findings be compared with what others have found?

To our knowledge, this is the first study that describes the reproducibility of a simulation test that assesses the detectability of warning signals in the acoustic environment of a specific workplace.

- **How is the reproducibility of our test compared to the reproducibility of tone audiometry.** Tone audiometry might be more reliable, but also perhaps less valid.
- **How is the reproducibility of our test compared to the reproducibility of standardized signal-in-noise tests?**
- **In what way did the reproducibility differ between the different train types? Are there explanations for it?**
- **Has the test sufficient reliability to continue further development.** Yes, for 4 out of the 6 trains.
- **What is the clinical implication of the described smallest detectable change?** In the future, the SDC might be valuable for assessing the auditory fitness for job of individual engineers, for example when the use of hearing aids is needed for safe and effective job performance. What change in detection threshold indicates a relevant change.

Regarding the first hypothesis of the second experiment (highest association for the frequencies at which the signal shows it most prominent signal peaks), how can the findings be compared with what others have found?

- **Who have indicated that frequency specificity might be relevant?** In general e.g. Tufts et al., specifically for locomotive engineers it was suggested by Hoek-Snieders (2020) who performed a study based on a computational prediction model of warning signal detectability. Our findings are in line with these studies.
- **It has been argued that a warning signal should not contain signal peaks above 3100 Hz, why not? Does our study underline this recommendation?** >3100 Hz has been indicated to be more sensitive to hearing loss. In train cabins, the ATP signal has all supra threshold signal peaks above 3100 Hz, DSD has prominent signal peaks below and above 3100 Hz. In our study, we see that the difference between the hearing impaired and the normally hearing individuals is way bigger for the ATP signal. This might be due to the frequency specificity.
- **What is a difference between our study and the statement of others about frequency specificity?** Other guidelines mainly speak about the frequency of the signal spectrum, but our study shows that it's more about the frequencies at which the signal has supra threshold signal peaks, which not only depends on the signal, but also on the level and spectrum of the ambient noise.

Regarding the second hypothesis of the second experiment (adding speech in noise would increase the explained variance of the test), how can the findings be compared with what others have found?

- **Who have indicated that tone audiometry and speech in noise would be both relevant? Were they relevant?** Although the results of the signal detection test were associated with the outcome of the pure tone audiometry and the speech in noise the percentage of variance was small regarding the DSD signal and moderate regarding the ATP signal.
- **Why is the explained variance (much) higher for the ATP signal detection test compared to the ATP signal detection test?** In this study, the percentage of explained variance was higher for the ATP signal than for the DSD signal. This might be explained that the masked thresholds of the ATP signal showed more variation than the masked thresholds of the DSD signal test. Also, a larger variation in measured pure tone thresholds was observed in the higher frequencies. Further, the masked threshold of the DSD signal did not differ that much (6 dB and 3 dB) from the mean masked threshold in the normal hearing population. Apparently, the engineers can compensate pretty good for their hearing loss when detecting the DSD signal, but not when detecting the ATP signal.
- **What can explain the fact that the hearing-in-noise tests have a higher association with the aided masked thesholds, than for the unaided measurements?** It has been argued that two auditory skills are required for adequate signal detection in noise. First detecting the sound, and second, distinguishing the sound from the ambient noise. Tone audiometry assesses the detection of a sound in quiet and speech in noise tests assesses the distinguishing of sounds. When hearing loss is present, both auditory skills are affected. Hearing aids can mostly restore the ability to detect a signal, but distinguishing sounds often remains difficult.
- **What does this imply?** The ability to detect a warning signal depends on the hearing acuity, but also on the level and spectrum of the signal and acoustic environment in which the signal needs to be detected.
- **What factors could improve the regression model?** Validated signal detection tests in noise, subjective outcomes??

What are the limitations of the study? What other factors could have influenced my findings. Have I reported everything that could make my findings invalid?

- **Is there a possible source for measurement bias?** The data of the hearing impaired engineers was collected during regular fitness for job assessments, as a result only two trains were included (otherwise it would cost too much time). Also, the aided measurements were performed with experienced hearing aid users as well as with engineers that did not wear hearing aids until they did not meet the hearing requirements of their job.
- **Is there a possible source for selection bias?** FIRST, We only included engineers who performed a hearing in noise test, this might have resulted in selection bias, since hearing in noise test might not have been performed in engineers with the highest or the lowest hearing degree of hearing loss. SECOND, Our normally hearing individuals were not locomotive engineers. adequate signal detection might be facilitated when in engineers who are familiar with the noise environment and the warning signals, which might implicate that normal hearing engineers would have even lower masked thresholds. However, our participants performed a practice round to get familiar with the noise environment and the signals. Further, all normal hearing subjects performed the test two times in 12 driving conditions. It can be concluded that there was no learning effect, since there was no systematic difference between the first and the second time
- **What are the implications of using a simulation test as the functional ability to detect warning signals?** We use a simulation test to assess the functional ability to detect warning signals in the acoustic environment of a Dutch train cabin. The real world situation is simplified – for example because an engineer only needs to detect the acoustic signals during the test, instead of combining the signal detection with driving safely, communicating with others and so on. Also, our simulation test used the masked detection threshold as an outcome, which results in a test procedure of decreasing and increasing the level of the warning signal, whereas in the real work situation, the test needs to be detected at a certain, predefined level. Assessing if an engineer is able to detect signals at this predefined level would not be an appropriate method, because of the simplified task that is performed in the simulation test.
- **How was the power?** We analyzed the reproducibility results of the different train types separately, which reduces the study power.

Do my interpretations contribute some new understanding of the problem that I have investigated? In which case do they suggest a shortcoming in or an advance on the work of others?

- **What are the advantages of using a simulation test rather than the outcomes of conventional hearing tests?** To our knowledge, this is the first study that describes to use a simulation test to assess the functional ability of locomotive engineers to detect acoustic warning signals the work environment of a train cabin. This has the advantage..
- **So, can the simulation test already be used to assess auditory fitness in locomotive engineers?** No, we did not assess appropriate cut offs for locomotive engineers, further research on this subject is required.
- **Could a simulation test be useful for other professions too?** In what professions is detecting warning signals important? Having multiple simulation tests, in multiple work settings, is required to come to general recommendations.
- **What future research would be needed?** Further validation of our test & determination of cut off values for locomotive engineers. AND developing train simulation tests for other professions AND investigation and validation of outcome measures that might be associated with the ability to detect warning signals in order to investigate if the ability to detect warning signals in simulated acoustic environments can be explained with other tests or subjective outcomes.