The interactions between Signal-to-noise ratio and Reverberation time in speech intelligibility and learning

Staffan Hygge[1]

[1] University of Gävle, Faculty of Engineering and Sustainable Development, Sweden

Corresponding author's e-mail address: Staffan.Hygge@hig.se

ABSTRACT

Signal-to-noise ratio (SNR) and Reverberation time (RvT) are the main acoustical factors for determining speech intelligibility (SI) and learning in classrooms. From the theoretical perspective of working memory (WM) and its limited capacity, I will address and review two issues: (1) Are the effects of SNR and RvT of different importance for SI and memory of what was said, and (2) Are the effects of SNR and RvT independent of each other or do they interact in their the effects on SI and memory? Studies from our own lab have indicated that a low SNR and a long RvT more strongly affects learning of a spoken text than the SI. As WM has a limited capacity, it follows that that SNR and RvT they should interact, e.g. when a low intelligibility of the spoken message exhausts the limited WM-resources in the process of identifying what is said. I will present studies that indicate such an interaction, and also that a long RvT sometimes improves, rather than impairs, memory and learning in certain conditions.

INTRODUCTION

In a classroom, the teacher's speech signal is distorted or degraded by the acoustic properties of the classroom. This reduces speech intelligibility, which in turn makes learning more difficult. Government agencies and professional societies have established building codes, recommendations and standards for acceptable background noise levels and reverberation times (RvT) in classrooms and other premises where hearing and understanding auditory information is important (American National Standards Institute, 2002 [1]; Byggnadsstyrelsen, 1975 [2]; Swedish Standards Institute, 2007 [3]; Swedish Work Environment Authority 2006) [4], 2006; Swedish Work Environment Authority, 2011 [5]; Vallet and Karabiber, 2002 [6]). These codes and standards are based on what is known to be required for correct identification of spoken words or isolated sentences, i.e., speech intelligibility (SI), that is typically defined as percentage or probability of correct identifications of words or short sentences. Recommended values for reverberation times are often given as < 0.5 - 0.6 s and for background noise ≤ 35 dBA. For an ordinary classroom where the teacher is standing at the front of the classroom and speaking at 66 dBA, the sound level of the speech decreases to ≈ 52 dBA at a distance of ≈ 6 m from the speaker out into the classroom, resulting in a +12 dB signal-to-noise ratio (SNR) above a very low 40 dBA background noise level where the pupils' activity noise level is assumed to add only 5 dB to the recommended 35 dBA.

A critical implicit assumption in the recommendations and norms for speech intelligibility (SI) in e.g., a classroom, is that they regard Signal-to-Noise Ratio (SNR) and Reverberation time (RvT) as basically additive and independent of each other. That is, improving on a long RvT by shortening it, is expected to improve SI to about the same magnitude regardless of whether that shortening of the RvT occurs in a context where the SNR is favorable or less favorable.

This additive assumption can be questioned from a theoretical point of view. On the assumption that the degradation of the speech signal from low SNRs and long RvTs is handled by overlapping cognitive structures/processes in working memory (WM), it follows that if a low intelligibility of the spoken message exhausts the limited WM-resources in the process of identifying what is said, very little cognitive capacity, if any, is left for further elaboration, decoding and understanding of the message.

In studies in our research group, we have seen indicators that question the assumption of additivity of SNR and RvT, and indicate synergistic or even crossover antagonistic interactions between RvT and SNR in how learning is affected. A shared feature of the studies is that the participants had Swedish as their first language (L1) but they listened to learning material spoken in English (L2). We also have seen that choosing learning as an endpoint, rather than SI, improves the Cohen effect size.

SUMMARIES AND RESULTS FROM SOME RELEVANT STUDIES

Are the effects of SNR and RvT of different importance for SI and memory of what was said?

SI, SNR and recall. In a study by Hygge, Kjellberg and Nöstl [7] 48 College students (mean age 27.1 yr, equal number of men and women) listened to 12 different word lists with 12 words each in Swedish and 12 in English. The lists were played back over headphones with a SNR of +3 or +12 dB. For the half of the word lists that made up the SI shadowing task, twelve in each language, the participants were instructed to repeat aloud the words they heard (shadowing), and their response was recorded. The presentation orders of the SI-list was counterbalanced. For the wordlists, the subjects had 1 min after the end of the list to type in the words they remembered.

The main findings were that both SI and recall was impaired in the unfavorable listening condition (+3 dB), but the *F*-ratios and the effect sizes were larger for recall than for SI. Language also had a main effect on recall, with a medium effect size, but Language did not have any significant effect on SI. Further, the effect of shadowing on recall was negative for the first two parts of the list, but positive for the last part.

The results indicated that SNR had strong effects on both SI and recall, but also that the effect size was larger for recall than for SI.

Thus, recall appear to be a more sensitive indicator than SI for the acoustics of learning, which has implications for building codes and recommendations concerning classrooms and other workplaces where both hearing and learning is important.

RvT and listening comprehension. In a study Hurtig et al. [8] pupils in grade 9, with Swedish as their first language (L1) (N = 125, age 15-16 yr) took two parallel versions of an English (L2) listening comprehension test, which is one of the Swedish National Tests for that grade. Two classrooms were used for the test, one with a short and one with a long RvT (0.33 and 1.07 s, means 250 Hz - 4 kHz). The listening test was played back from only one loudspeaker in the front of the classroom. All pupils took two counterbalanced versions of the test, one at the front and one in the middle of the classroom (1 and 6 m). We also had their scores on an independent English reading test. The results indicated that there was a significant effect of position, meaning that sitting at the front of the classroom resulted in higher scores on the English listening test. There was also an unexpected crossover interaction between RvT and how good the pupils were at reading English. See Figure 1! The participants who scored low on the English reading test did unexpectedly benefit from a longer RvT, while those with a higher reading score did benefit from the shorter RvT. The simple main effect of RvT were significant at both levels of English Reading Comprehension.

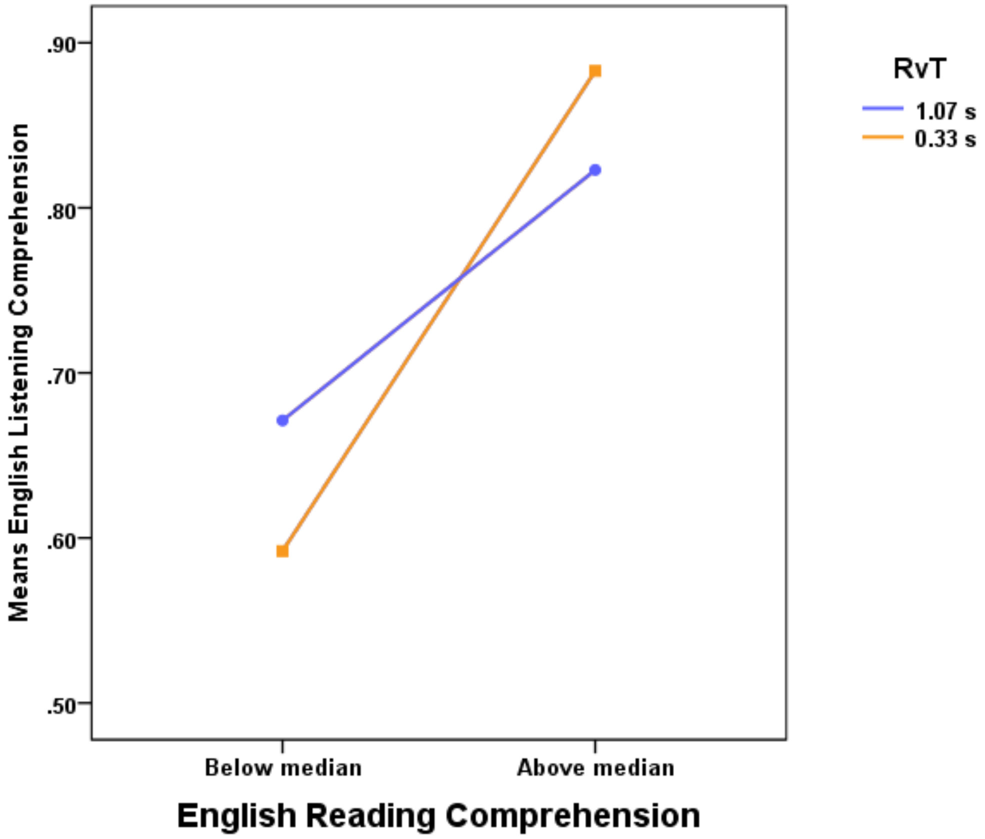

Figure 1: Means English Listening Comprehension for Grade 9 by their English Reading Comprehension at RvTs of 0.33 and 1.07 s. Significant simple main effects of RvT at both levels of English Reading Comprehension.

In MacCutcheon et al. [9] 57 Swedish college students (mean age 25.3 yr, 23 male) were presented with 18 audio clips (1.5-2.0 min) in English. Half of the clips were presented with a RvT of 0.3 s and the other half (counterbalanced) with a RvT of 0.9 s. The students worked individually with a computer and the audio clips presented over headphones. Comprehension of the text was assessed with multiple-choice questions directly after each clip. All participants were also pre-tested on a WM updating test and on the Boston Naming Test to determine their baseline second-language (L2) naming ability. There was a significant synergistic between RvT and how good they were at the Boston test (median split). The high scorers did benefit from a short RvT, while those with a low score did not. See Figure 2! There was no significant interaction between RvT and above/below median on the WM test. Also there was no three-way interaction RvT*Boston naming test*WM. This lack of interaction may indicate that the effects of RvT primarily were mediated by or routed by language competence processes rather than by WM-updating processes.

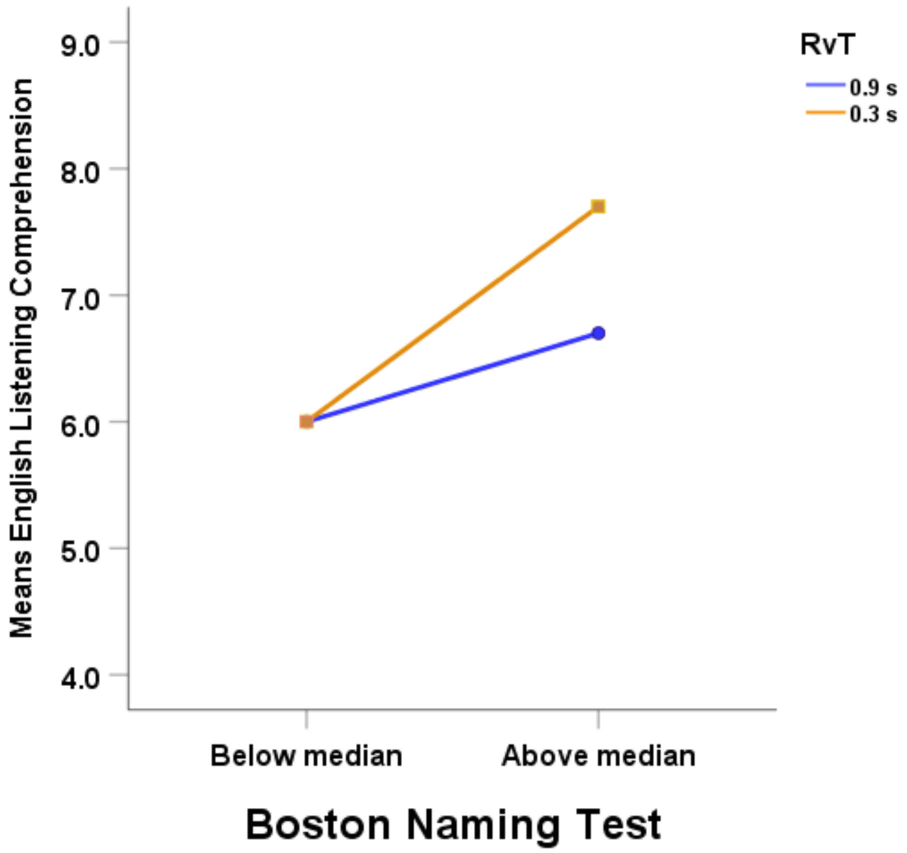

Figure 2: Means English Listening Comprehension for College students by RvTs of 0.9 and 0.3 s and their scores on the Boston Naming Test. Synergistic interaction significant.

Are the effects of SNR and RvT independent of each other or do they interact in their effects on SI and memory?

SNR, RvT and recall. Free recall of words in wordlists were studied for children in Grade 4 [10] and for College students. [11]. The two studies had same independent variables: SNR +3 and +12 dB, RvTs 1.2 and 0.3 s, and Language (Swedish-English) in a crossed within-person design.

In Grade 4 (10-11 yr) a total of 72 pupils in 4 classes were run as a group in their classrooms. The 48 College students (24 males) were run individually in a lab with headphones and a computer. There were 12 wordlists in English and 12 in Swedish. There were 12 words in the lists for the College students, but the lists for Grade 4 had 8 words each, the 8 easiest words from the lists for the College students.

Words with high ranks were taken from language specific category norms for 24 categories, which were sorted by Graeco-Latin squares into 24 lists with equal average ranks of the words.

For the children in grade 4 there was no main effect of RvT but a significant unexpected crossover interaction SNR*RvT. At both SNR levels the simple main effects of RvT were significant and in different directions. That is, at the lower +3 dB SNR level, a short RvT impaired recall, but at SNR +12 dB the shorter RvT improved recall. See Fig 3!

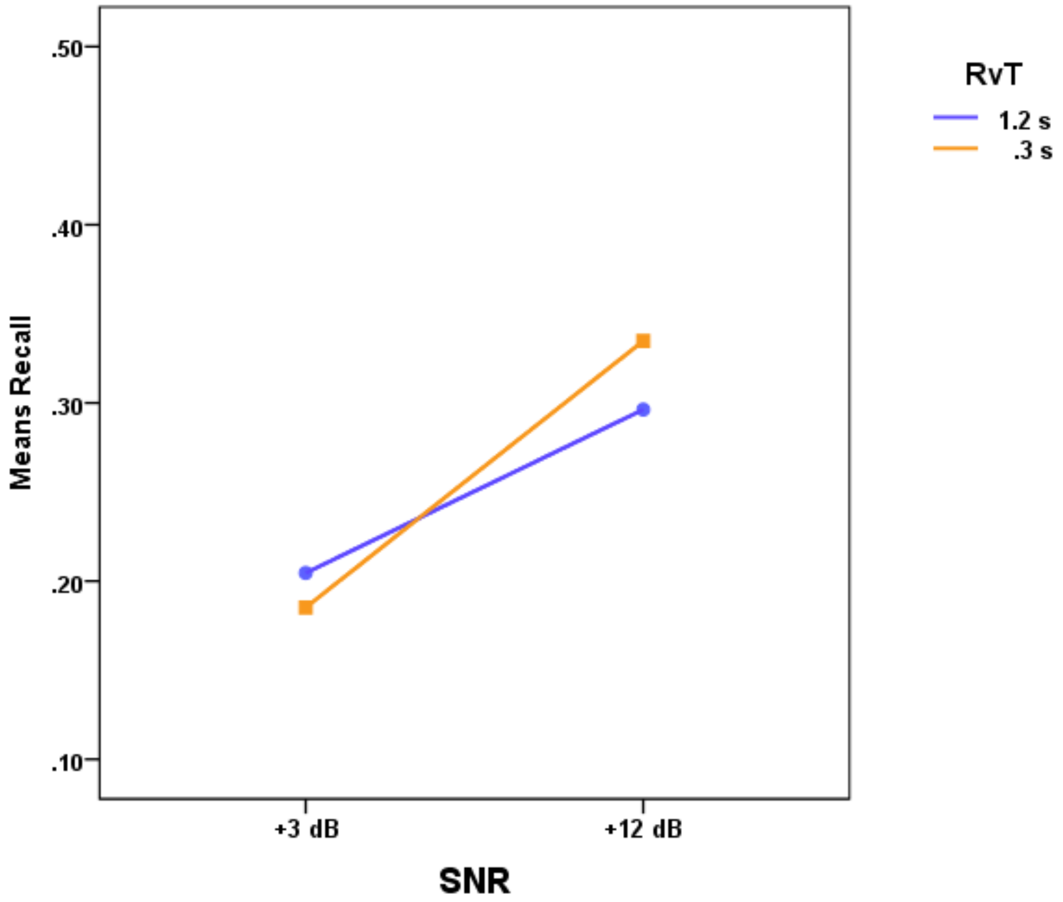

Figure 3: Means Recall for Grade 4 by SNR and RvT. Crossover interaction significant. Significant simple main effects of RvT at both levels of SNR.

For the College student there was a strong main effect of SNR, but no main effect of RvT, or any interaction SNR * RvT.

However, when the College students were sorted into three groups of how skilled they were in English, based on their recall scores for the English words, a crossover interaction turned up for RvT*Engskill ($p < .05$). Se Fig 4! There were also significant simple main effects of RvT at the low and high levels of Engskill.

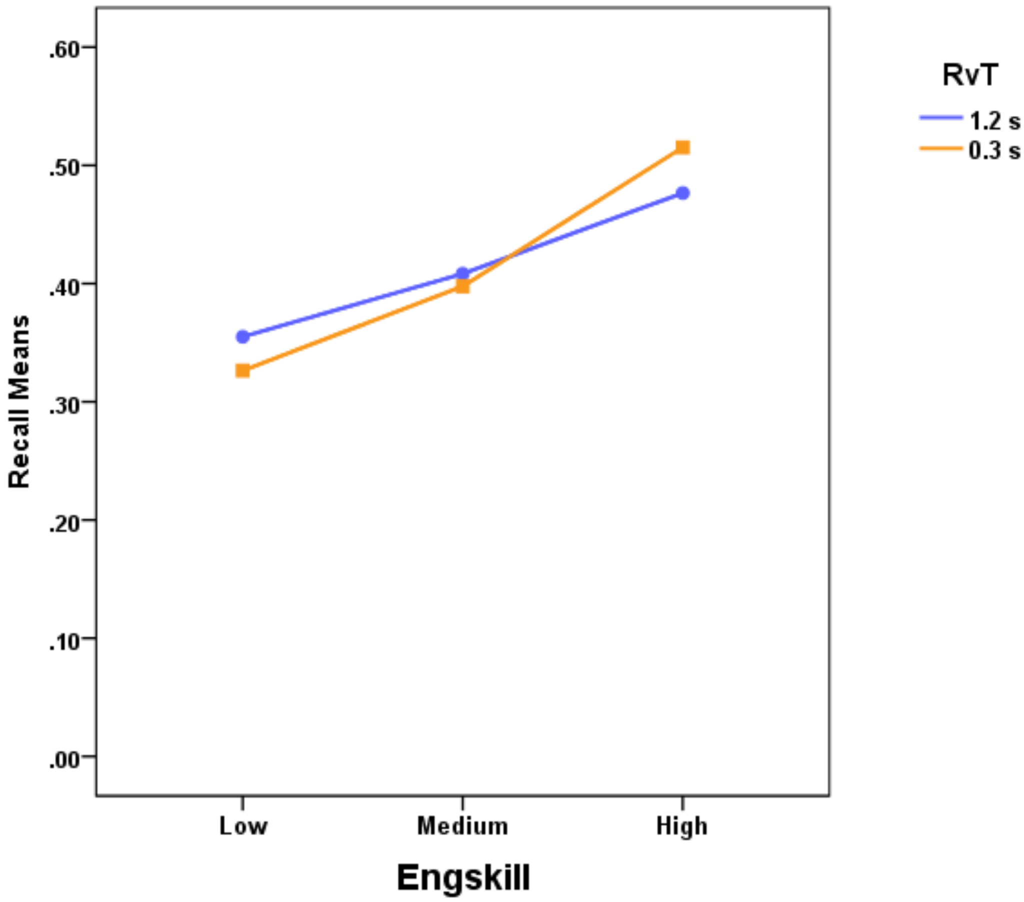

Figure 4. College students. Crossover interaction. Significant simple main effects of RvT at the low and high levels of Engskill.

When the scores for the medium skilled group in the College group was analyzed by itself, a borderline significant interaction between RvT*SNR ($p < .10$) was found. See Figure 5! There were also borderline significant simple main effects of RvT at both levels of SNR.

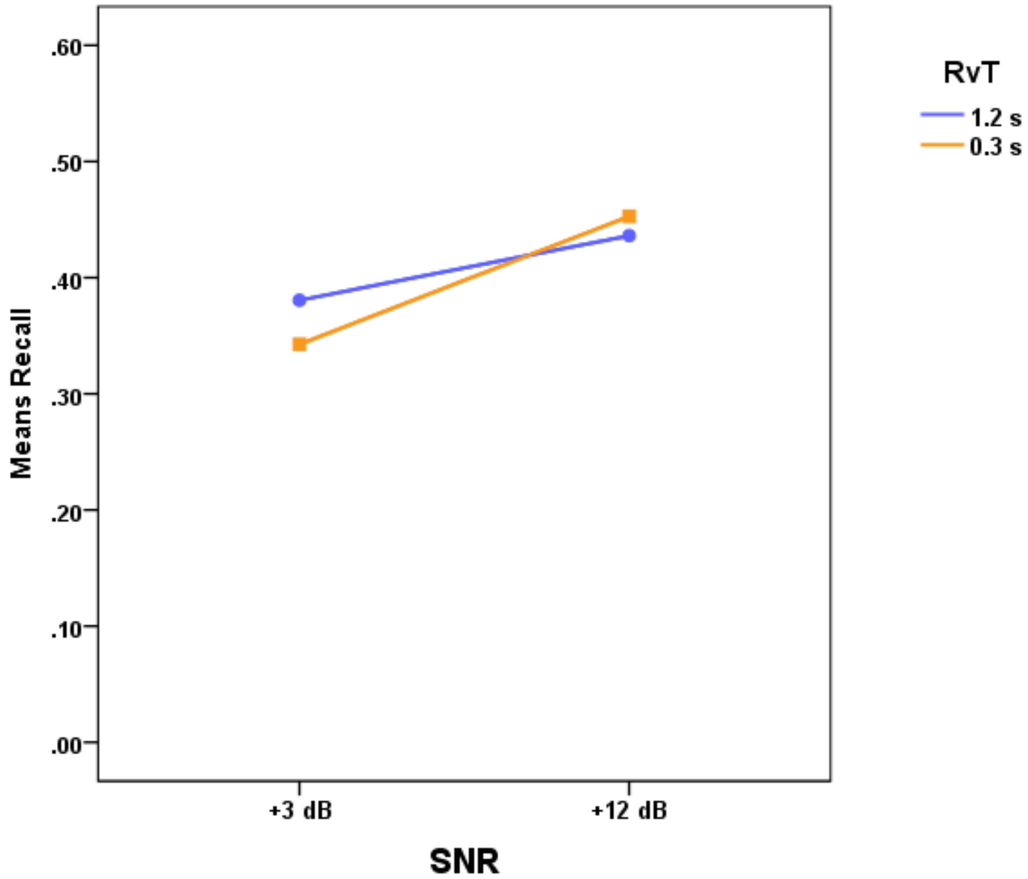

Figure 5. College students, medium skilled. Borderline (p < .10) crossover interaction, border line significant simple main effects of RvT at both levels of SNR.

DISCUSSION

At the beginning of my presentation of results from relevant studies, it was indicated that that free recall of wordlists is more sensitive to a favorable SNR than SI is. Thus, SI and speech perception may not be the best suitable criterion for class classroom contexts where good memory and recall have a high priority.

In the second part it was indicated that for listening comprehension a favorable RvT it not always a good predictor. Rather, the benefit a short reverberation time is restricted to students who are good at reading the language (L2 English) for the auditive test material. To the students less able in English, a long RvT was beneficial for recall and comprehension.

In the third part a crossover interaction SNR*RvT was indicated for free recall of word lists in Grade 4 pupils to the effect that at the lower +3 dB SNR level, a short RvT impaired recall, but at SNR +12 dB the shorter RvT improved recall.

Thus, the additivity assumption for the effects of RvT on recall and comprehension does not hold up. Both for listening comprehension and free recall of word lists a favorable RvT does not always results in a better performance, and if a favorable RvT improves performance it seems to be restricted to persons who are linguistically competent.

A synergistic interaction involving RvT, showing better performance with a short RvT for persons in one group, and no difference with a long RvT, is logically compatible with the assumption from WM that the degradation of the speech signal with a long RvTs may be sufficient to exhaust the limited WM-resources in the process of identifying what is said. When the RvT is favourably short, there will be less exhaustion of WM capacity, at least for those who are good at the language spoken (L2 English).

However, I see no easy way to incorporate into a WM-model the crossover antagonistic effect of the longer RvT to be beneficial for persons with less linguistic competence.

Further, the crossover interaction SNR*RvT for pupils in Grade 4, in Fig 3, seems even more difficult to incorporate into a WM-model.

In addition, in several of the studies above, there were no indications that WM capacity had a strong impact on recall or listening comprehension.

As it is not clear from WM-model how and why (favorable) short RvTs is of help only for those who have high linguistic skills, while the opposite is true for those with low linguistic skills, or a 3 dB SNR, an alternative to the WM-model is needed.

Such an alternative theoretical approach to WM framework, probably needs to go deeper into details of speech perception/intelligibility issues that include speed of lexical access for correct responses, reading comprehension in Swedish and English, quality of phonological representations, and decoding skills.

Acknowledgements

Both the Swedish Research Council (PI: S Hygge - A0204201) and the Swedish Research Council Formas (PI: S Hygge - 242-2010-1006) contributed to the funding of the studies reported here.

REFERENCES

[1] American National Standards Institute. ANSI S12.60-2002. Acoustical performance criteria, design requirements, and guidelines for schools. 2002.

[2] Board] BTSB. Svensk byggnorm [Swedish Building Code]. The Swedish Planning Board. 1975.

[3] Swedish Standards Institute. SS 25268:2007 Acoustics - Sound classification of spaces in buildings - Institutional premises, rooms for education, preschools and leisure-time centers, rooms for office work and hotels [In Swedish] Swedish Standards Institute2007.

[4] Swedish Work Environment Authority. Rapport – Verksgemensamt projekt – Buller i skolmiljö 2003 [Report – Interdepartmental project – Noise in the school environment 2003].CTO 2003/482 2006.

[5] Swedish Work Environment Authority. Kunskapsöversikt - Yrkesrelaterade röststörningar och röstergonomi [Review – Work related voice problems and voice ergonomics]. Report 2011:6. 2011.

[6] Vallet M, Karabiber Z. Some European policies regarding acoustical comfort in educational buildings. *Noise Control Engineering Journal. 2002*;50:58-62.

[7] Hygge S, Kjellberg A, Nöstl A. Speech intelligibility and recall of first and second language words heard at different signal-to-noise ratios. *Frontiers in Psychology. 2015*;6. doi: 10.3389/fpsyg.2015.01390. PMC4568391.

[8] Hurtig, A., Sörqvist, P., Ljung, R., Hygge, S., & Rönnberg, J. (2016). Student's second-language grade may depend on classroom listening position. *PlosOne*, PONE-D-15-41873R2

[9] MacCutcheon, D., Hurtig, A., Pausch, F., Hygge, S., Fels, J., & Ljung, R. (2019): Second language vocabulary level is related to benefits for second language listening comprehension under lower reverberation time conditions. *Journal of Cognitive Psychology*, doi: 10.1080/20445911.2019.1575387

[10] Hurtig, A., Keus van de Poll, M., Pekkola, E., Hygge, S., Ljung, R., & Sörqvist, P. (2016). Children's recall of words spoken in their first and second language: Effects of signal-to-noise ratio and reverberation time. *Frontiers in Psychology, 6*:2029. doi: 10.3389/fpsyg.2015.02029

[11] Hygge, S. (2020). Recall of words spoken in the first and second language: Effects of signal-to-noise ratio and reverberation time (Manuscript).