ASSESSING ROOM ACOUSTIC MEMORY USING A YES/NO AND A 2-AFC PARADIGM

Madalina NASTASA[1], **Nils MEYER-KAHLEN**[1], and **Sebastian J. SCHLECHT**[1,2]

[1]*Acoustics Lab, Department of Signal Processing and Acoustics*, **Aalto University**, Otakaari 5, Espoo, 00250 Finland
[2]*Media Lab, Department of Media*, **Aalto University**, Otakaari 5, Espoo, 00250 Finland

ABSTRACT

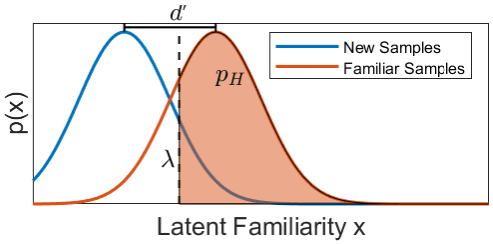
We present a study that tests the ability to remember room acoustics – a cognitive skill that is one of the guiding mechanisms behind plausible virtual acoustics for extended realities. Room acoustic memory was tested by assessing a person's ability to recognise sound samples, convolved with room impulse responses of everyday rooms presented in a preceding training session. To test a common assumption of detection theory, we conducted two listening tests using both a yes/no and a 2AFC paradigm. Results show that subjects can recognise different rooms above chance level, but even with relatively large differences between the rooms, the accuracy is low in general. Furthermore, the relation between the two test paradigms follows the prediction of detection theory when averaging over all participants, but less so for individual participants.

1. INTRODUCTION

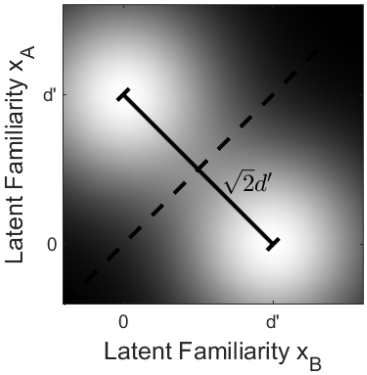
Memory for room acoustic is a cognitive ability that plays an essential role in virtual acoustics for extended realities (XR). The central question is: If a particular room was heard before, either in the real or the virtual world, would a user be able to recognise its acoustic?

Our experimental starting point towards answering this interdisciplinary question assessed how well participants could remember the sound of everyday rooms over a short time span. In this preliminary study, we focus on testing two paradigms for assessing memory or similarity and an important decision theoretical assumption connecting them. It relates the results of two-alternative forced choice (2-AFC) to yes/no tasks. We expect that such tasks will be of increasing practical importance in virtual acoustics, as they are the best choice for evaluating the plausibility of acoustic rendering in VR. In [1, 2], the theoretical relation between the two tasks was used to make an informed choice about the performance threshold of a virtual acoustic rendering system.

In Section 2, we provide some background about acoustic memory and then describe the theoretical decision foundations and the assumption we test. In Section 3, we describe our memory task. In Section 4, we discuss the results, draw conclusions both in terms of decision theory and memory, and summarize the implications for future room acoustic memory tests.



(a) The assumed model of detection theory for a yes/no memory task



(b) The assumed decision space for 2-AFC

Figure 1. Detection Theory Models for the yes/no and the 2AFC task.

2. BACKGROUND

2.1 Room Acoustic Memory

Cognitive psychologists have studied auditory memory for many years [3]. A classical model of the auditory memory mechanism involves a first stage, the auditory sensory memory (ASM), which serves as a signal buffer and is subject to fast decay in the range of a few seconds. As long as a signal is present in ASM, encoding can take place, in which information is extracted and passed on to the working memory. Here, the information can be processed, rehearsed, or compared to information from long-term memory. While such a model seems appropriate for verbal sounds that need to be classified into phoneme categories to understand speech, the exact encoding, and maintenance strategies are less clear for non-verbal sounds.

In terms of room acoustics, one hypothesis could be that listeners encode sound by extracting the magnitude of several acoustic attributes and then remember this information. To our knowledge, there is no work on room acoustic memory yet. However, there might be a strong connection to recent work on the memory for timbre [4]. For example, [5] has presented results that show an increased memory capacity for familiar timbres. Also, it indicates that memory for timbre might not depend simply on rehearsing verbal labels assigned to each sound during encoding because a visual suppression condition had a similar effect as articulately suppression. Our experiment does not go as far as to test specific mechanisms. Still, it represents the first step in testing memory for acoustics, with the aim of trying methodology and evaluation using signal detection analysis.

2.2 Detection Theory

In one of the applied paradigms under evaluation in this study, participants decide whether samples are either "familiar" or "new." Such a paradigm is generally referred to a yes/no paradigm [6]. The analysis used for the results of such a test can not only be applied to memory studies (like [5]), but to many different questions, for example, regarding the plausibility of virtual acoustic rendering [1, 2, 7]. There, a participant is presented with a real source, often reproduced using a loudspeaker, and a virtual source reproduced over headphones. Then, the participant is asked whether the source is real (corresponding to "yes") or virtual ("no").

The result of a yes/no experiment is a 2×2 table consisting of four different response classes, which are then converted to relative response frequencies. In the case of our memory task, the table may be labeled with "familiar" and "new." In the terminology of detection theory, the four categories are named as shown in Table 1. If a familiar sample was detected, it was a hit; it is a miss if a familiar sample was not detected as such. Further, if a new sample was recognized as new, it is a correct reject, and a familiar sample marked as new is a false alarm. The terminology is borrowed from radar technology.

One way of analyzing this data would be to add all the diagonal elements and divide them by the total number of trials. This results in the percentage correct p_{C}. While this gives a first intuitive impression of the results, it is not the most common measure used in detection theory. Instead, the theory offers separate measures of sensitivity ("How large was the perceptual distance?") and bias ("Did a subject tend towards one of the answers?"). Sensitivity can be calculated by

$$d' = \Phi^{-1}(p_{\mathrm{H}}) - \Phi^{-1}(p_{\mathrm{F}}), \tag{1}$$

where p_{H} = #Hits/#Familiar and p_{F} = #False Alarms/#New and $\Phi^{-1}(x)$ is the inverse Gaussian cumulative density function. The model assumes that the decision is based on a latent, continuous variable in an abstract decision space, with one distribution for the new and one for the familiar samples. These distributions

		Given Answer	
		"Familiar"	"New"
True Answer	Familiar	Hits	Misses
	New	False Alarms	Correct Rejects

Table 1. The result table of a yes/no experiment.

are assumed to be Gaussian and have equal variance. Subjects are then thought to have an internal "criterion" or "decision boundary" λ. If a sample exceeds this criterion on the decision axis, the response will be "familiar". The sensitivity corresponds to the distance of the means in these assumed distributions. From the various forms of measuring bias [6], we have chosen the likelihood ratio because it is the most general measure, also applied in other fields. The likelihood ratio, usually called β, is a dependent on the sensitivity, d', and the criterion location λ, the basic bias measure for detection theory, i.e.,

$$\beta = e^{\lambda d'}. \tag{2}$$

A listener bias ratio of 1 indicates an unbiased response, while a ratio < 1 shows a tendency to report "yes" to familiar responses, and a ratio > 1 is a tendency towards "no".

The most important feature of a yes/no task is that the subject is presented with one item per trial. In our experiment, we compare this paradigm to a 2-AFC task. In such a design, participants are presented with two alternatives ("A", "B"), one of which always is familiar and one is new, and participants are asked to decide which is which. Performance is expected to be better in this task. Hence a larger sensitivity should be measured, which is calculated just as before, with the arbitrary definition of A as yes and B as no, so that a hit is a case in which A was selected and indeed the familiar sample. Detection theory predicts the sensitivity difference between the tasks by

$$d'_{\mathrm{AFC}} = \sqrt{2} d'_{\mathrm{YN}}. \tag{3}$$

This is based on a striking assumption: if one new and one familiar sample are presented in the same trial, the latent decision space becomes two dimensional, with one axis corresponding to each sample. The 2D distance between the means is $\sqrt{2}$ longer than the distance of the means on one axis, hence the assumed factor. [1] uses this factor to derive a performance limit of his virtual acoustic system. The percentage of correct answers in a hypothetical, unbiased 2AFC task was set to be 5% over guessing probability ($p_{\mathrm{C,2AFC}} = 0.55$). The equivalent d' in the yes/no task is converted by

$$d'_{YN} = \sqrt{2}\Phi^{-1}(0.55). \tag{4}$$

Our memory test gives us a chance to test the applicability of the $\sqrt{2}$ conversion factor in practice.

(a) Training (b) Yes/No Task (c) 2 AFC Task

Figure 2. GUI of the training session and the interfaces used for the two paradigms. Also a confidence rating was provided on each trial (currently not used for analysis).

3. EXPERIMENT

To test the memory ability for room acoustics, ten different environmental monaural impulse responses measured in daily life locations were chosen from the dataset described in [8]. The locations vary from medium-sized rooms (T_{30} = 0.29 s to 0.49 s), an office or classroom, to larger spaces (T_{30} = 1.02 s to 6.49 s), a garage and a train station hallway, and even outdoor spaces (T_{30} = 0.07 s to 0.49 s), a house balcony or tram stop shelter. In this way, it was ensured that there were enough differences between the room characteristics, such that all pairs would be distinguishable in direct comparison. Three different sound sources were used in the experiment, namely conga drums, speech, and classical guitar. The sound sources were chosen so that they would possess different frequency and temporal characteristics. The test signals were created by convolving the impulse responses of the various locations with the sound sources. [1]

For each category of sound source the test consisted of a session with a yes/no task (see GUI in Figure 2b), and a session with a 2AFC task (Figure 2c), each preceded by a training session (Figure 2a). In the training session, each subject was presented with five room renderings, randomly selected from the set, and the task was to memorize them as well as possible. The training session was not time-limited. When the subject was ready to proceed, there was a break of 15 s in which the original sound source without any room response was played. Then, the test session started. The break was included to minimize the effect of auditory sensory memory, which may allow participants to store signals themselves for several seconds.

In the yes/no task, the subjects were presented with ten renderings, and their task was to choose whether the test signal was familiar from the training session or not. In the 2AFC test, subjects went through five trials in which two sound sources were presented; their task was to decide which of the two was familiar to them. 2AFC had half as many trails as the yes/no task, as two signals are required for comparison, and we wanted to avoid repetitions that might boost retention, and limit ourselves to the same data as in the other task. After ten, respective five questions, the next training set was presented. The order of presenting the two test tasks was randomised across subjects so that four of the subjects started with the yes/no task, while the other six started with the 2AFC task. This random presentation allows us to investigate whether the order in which the tests are presented has any effect on the participants' sensitivity.

In total, the test consisted of (yes/no task trials, stimuli, 2AFC task trials) 10 x 3 + 5 x 3 = 45 trials. The experiment was implemented in Matlab and conducted in designated listening booths over Sennheiser HD650 headphones at the Aalto Acoustics Lab with ten participants (8 male, 2 female). Overall, the average age was of 29 years old (SD=3.81). All of the participants were Master's or Ph.D. students of the Lab and had experience participating or designing listening tests themselves. Therefore they can be categorised as experienced listeners.

4. RESULTS

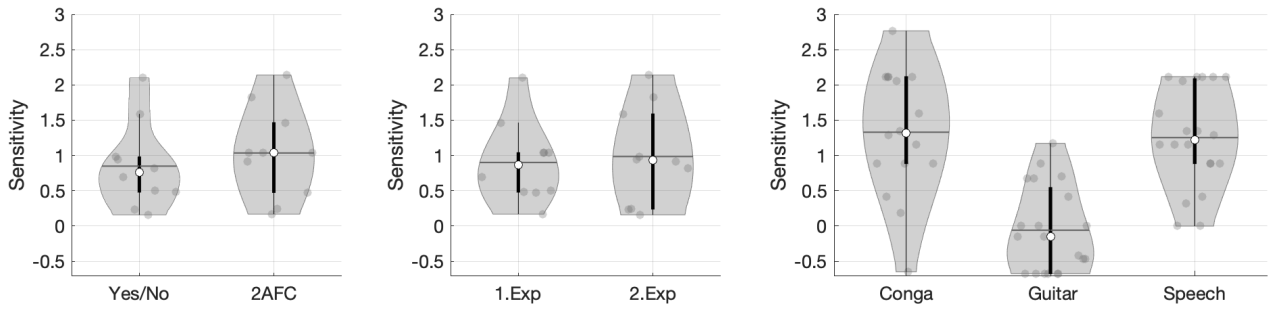
The responses resulting from the two experiments were separated into hit and false alarm rates per individual, which allows for the calculation of the sensitivity measure, d', and the listener bias, β.

One participant recognized all the familiar samples for all stimuli cases, while another recognized both the familiar and the new samples in case of the conga stimulus. In both cases, d' suggests a perfect accuracy by taking an infinite value. To avoid infinite values in the figure, we adjusted the results by adding 0.5 to the hits and misses cells and 1 to the number of familiar and new cells [6].

Based on the d' values, we conducted a mixed ANOVA with the order of the two paradigms as between-subject factor and the paradigm and sound source as within-subject factors. There was no effect of the presentation order, $F(1, 6) = 0.0029$ $p = .96$, indicating that subjects did not perform better in the second test, because they fad familiarized themselves with the task.

As expected, participants performed better at the 2-AFC task. While the yes/no experiment had an overall percentage of correct responses of $p_C = 66.07\%$, the 2AFC task had an overall $p_C = 72.14\%$. However, while this trend follows our expectations, the effect of the paradigm on d' is not significant, $F(1, 6) = 0.27$ $p = .61$. This is due to the large variability between participants, see Figure 3a.

[1] Examples can be found at `http://research.spa.aalto.fi/publications/papers/nordicsmc_roommemory/`



(a) All participants, comparison of the two paradigms.

(b) Comparison of experiment order

(c) Comparison of stimuli

Figure 3. Sensitivity d' between paradigms and stimuli.

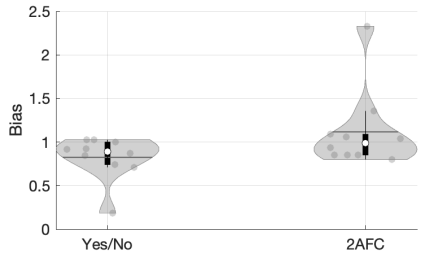


Figure 4. Likelihood ratio (bias) for all participants for both experiments.

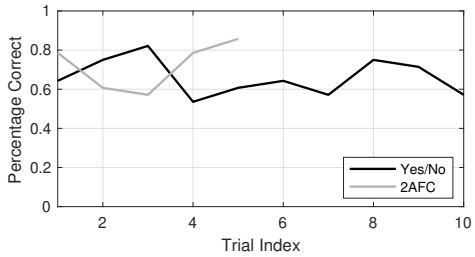


Figure 5. Percentage correct depending on position in the test.

Some performed much better at the AFC task, while one even performed worse. The ratio $d'_{\mathrm{2AFC}}/d'_{\mathrm{YN}}$ is between 0.179 to 6.55 (M=1.80).

However, when taking all the data of each paradigm together, the factor relating (d'_{YN}= 0.84) and (d'_{2AFC} = 1.19) is 1.4192, and thereby curiously close to $\sqrt{2} = 1.4142$. Note however, that this is not a meaningful procedure, as sensitivity and bias are thought to be subject specific.

A statistically significant effect was found when comparing d' between the signal conditions, F(2, 5) = 17.37 p = .006. The effect is due to the low performance during the guitar sample, when compared to conga and speech, see Figure 3c. There was also a significant interaction between paradigm and signal, owing to the fact that the differences were larger in the 2AFC paradigm, F(2, 5) = 47.468 p < .001.

Only one participant showed a totally unbiased behaviour for the yes/no task $\beta = 1$. The mean bias for the whole experiment is at 0.82, representing a slight tendency toward "yes" answers, see Figure 4. For the 2AFC task, where no bias is expected, participants tended slightly toward the “Room 2" answers, with the mean of the whole experiment of 1.11.

We also investigated whether the order of the presented trials affects the percentage of correct answers. Figure 5 shows that such an effect was not present, with subjects' performance rising and falling in no particular order.

5. DISCUSSION

In general, we showed that even for vastly varying rooms and using the same signals for training and testing, some participants' performance was low, especially for the guitar sample. A possible explanation for the low performance of the guitar could be the fact that the sample were more continuous, unlike the transient nature of the conga drums, or the natural breaks of the speech. Such breaks could improve memory because they make it easier to extract the acoustic room characteristics in the reverberant tail. Furthermore, the variation between participants was considerable, indicating that the ability to remember room acoustics might vary strongly. Figure 6 shows that some participants with professional musical experience of 10 or more years performed very well, whereas listeners with no professional musical experience performed bad or on a medium level. However, the sample size is too small to make conclusive statements about such a simple relation.

With regards to the conversion factor $\sqrt{2}$, we have seen that while it holds for the complete dataset, the performance difference between the tasks varies strongly between participants, at least using this relatively low number of decisions (45 per participant). This suggests that using the relation might not always be reliable, which is in agreement with other tests of the assumption using other tasks [9]. More elaborate models of relating tasks have been proposed in the literature [10], for example, incorporating unequal variance. In any case, signal detection theory provides the bias-free measure d', which can be compared between studies, also when evaluating virtual acoustics.

The lack of memory decay in dependence of the trial position shows that the break of 15 s was sufficient to exclude

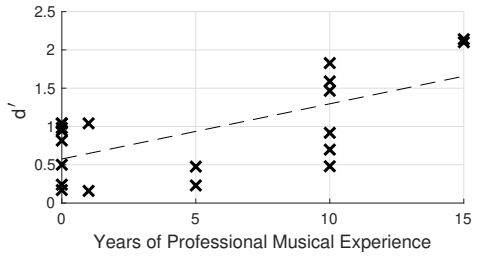


Figure 6. Influence of professional musical experience on the response sensitivity. $R^2 = 0.42, p = .002$

effects of sensory memory, if they exist. If sensory memory were still at work, the subjects would perform best in the trials following the training session, and their performance would be expected to decrease with the trial index.

5.1 Future Research

One interesting aspect to investigate further would be the effect of different time gaps between training and testing. This would allow us to check if ASM plays a role in comparing room acoustics. Furthermore, in this test, we have asked users to recognise rooms convolved with the same sound type per test session. For future work, subjects could be presented with different sounds convolved with the same IR to see how well they can recognise the acoustics of different rooms when the signals for the same RIR (Room Impulse Response) vary. It has been shown that such an added cognitive process of source/room separation makes comparing room acoustics much more difficult [11]. While in this test, the room renderings have been chosen randomly from a relatively small subset, in the future, it would also be beneficial if a set of participants would rate the perceptual similarity of the impulse responses in a prior test. In this way, a clearer relationship between sound similarity and the subject's memory performance can be formed. Also, binaural, or even head-tracked reproduction should be used in the future.

6. CONCLUSION

In this test, we have found that the $\sqrt{2}$ relationship relating the yes/no and 2AFC test is only approximated well when taking the whole dataset with all subjects into account. However, this factor highly varies between participants, which suggests that this assumption does not produce reliable results in all situations. When it comes to room memory, subjects can recognize different rooms above chance level, but with the large differences between the selected rooms, the sensitivity can be considered fairly low. Furthermore, the choice of stimuli seems to play an important part too. For the guitar sample, familiar and new renderings were indistinguishable for subjects. In terms of memory, we demonstrated that after 15 s, there was no decrease in performance with the trial index, which means that such a break is enough to exclude the effects of sensory memory. Now that we have a designated evaluation method using signal detection theory and a test methodology, future work can focus on the specific mechanisms at work for room acoustics memory and the intricacies of room dissimilarities and different stimuli.

Acknowledgments

This research has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 812719.

7. REFERENCES

[1] A. Lindau and S. Weinzierl, "Assessing the Plausibility of Virtual Acoustic Environments," *Acta Acustica united with Acustica*, vol. 98, no. 5, pp. 804–810, Sep. 2012.

[2] C. Pike, F. Melchior, and T. Tew, "Assessing the Plausibility of Non-Individualised Dynamic Binaural Synthesis in a Small Room," in *AES 55th International Conference*, Helsinki, 2014.

[3] M. H. Ashcraft and G. A. Radvansky, *Cognition*, sixth edition ed. Boston: Pearson Education, 2014.

[4] K. Siedenburg and D. Müllensiefen, "Memory for Timbre," in *Timbre: Acoustics, Perception, and Cognition*, ser. Springer Handbook of Auditory Research. Switzerland: Springer Nature AG, 2019, vol. 69, pp. 87–118.

[5] K. Siedenburg and S. McAdams, "The role of long-term familiarity and attentional maintenance in short-term memory for timbre," *Memory*, vol. 25, no. 4, pp. 550–564, Apr. 2017.

[6] N. A. Macmillan and C. D. Creelman, *Detection theory: a user's guide*, 2nd ed. Mahwah, N.J: Lawrence Erlbaum Associates, 2005.

[7] A. Neidhardt and A. M. Zerlik, "The Availability of a Hidden Real Reference Affects the Plausibility of Position-Dynamic Auditory AR," *Frontiers in Virtual Reality*, vol. 2, Sep. 2021.

[8] J. Traer and J. H. McDermott, "Statistics of natural reverberation enable perceptual separation of sound and space," *Proceedings of the National Academy of Sciences*, vol. 113, no. 48, pp. E7856–E7865, Nov. 2016.

[9] M. M. Langley, "d' is not appropriate for contrasting yes-no and forced-choice recognition," Retrospective Theses and Dissertations, Iowa State University, Ames, 2006.

[10] Y. Jang, J. T. Wixted, and D. E. Huber, "Testing signal-detection models of yes/no and two-alternative forced-choice recognition memory." *Journal of Experimental Psychology: General*, vol. 138, no. 2, pp. 291–306, 2009.

[11] A. Kuusinen and T. Lokki, "Recognizing individual concert halls is difficult when listening to the acoustics with different musical passages," *The Journal of the Acoustical Society of America*, vol. 148, no. 3, pp. 1380–1390, Sep. 2020.