Eberhard Karls Universität Tübingen

Faculty of Sciences
Cognitive Neuroscience

Master Thesis in Cognitive Science

Multisensory integration in the working memory - the interplay between perception & working memory in n-back performance

Niclas Renner

July 27, 2018

First Reviewer:
PD Dr. rer. nat. Gregor Hardieß
Cognitive Neuroscience
Department of Biology
Universität Tübingen
Auf der Morgenstelle 28
72076 Tübingen

Second Reviewer:
Prof. Dr. Hanspeter A. Mallot
Cognitive Neuroscience
Department of Biology
University of Tübingen
Auf der Morgenstelle 28
72076 Tübingen

ERKLÄRUNG

Hiermit erkläre ich, dass ich diese schriftliche Abschlussarbeit selbständig verfasst habe, keine anderen als die angegebenen Hilfsmittel und Quellen benutzt habe und alle wörtlich oder sinngemäß aus anderen Werken übernommenen Aussagen als solche gekennzeichnet habe.

July 27, 2018, Tübingen

Niclas Renner

Acknowledgments

First of all I like to thank PD Dr. rer. nat. Gregor Hardieß for accepting my request to supervise my master thesis and being the first reviewer. I like to especially thank him for the great mentoring, the continuous support and time he invested in numerous productive discussions with me.

I also like to thank Prof. Dr. Hanspeter A. Mallot for being my second reviewer and for the time he spent to discuss several problems that occurred during my thesis.

Furthermore, I want to thank the whole staff of the Cognitive Neuroscience Lab for the great working atmosphere and their cordial support and many productive discussions.

At last I want to thank all the people who took the time to participate in the experiments, especially friends who sacrificed their free time to participate in a preliminary experiment with no compensation.

Contents

List of Figures

List of Tables

Abstract

What do we remember if see a dog, only hear a dog or both, see and hear a dog. It is not clear whether information of different sensory systems gets integrated in working memory and stored as one memory trace or if two modality specific traces are stored in working memory. To explore this question I conducted a working memory task (2-back task) with nine different conditions, three modalities (visual, auditory, bimodal) and three detection rates (50%, 75%, 99%) to investigate memory performance of unimodal stimuli compared to bimodal stimuli at different stimulus detectabilities. Furthermore, I suggest a theoretical measurement based on signal detection theory to separate detection benefits from working memory performance benefits for bimodal stimuli presentations. The results show no significant improvement of memory performance for bimodal stimuli at either detection rate. To examine this unexpected result I take a close look at the underlying experimental setting and potential confounding variables.

1 Introduction

For years research on working memory was heavily focused on unisensory[1] processes, in most cases either on visual or auditory processes. But in recent years researchers shifted their focus more on the aspects of multisensory processes but many questions remain unanswered. For example, it is not clear how multisensory representations are handled in the working memory. The question whether multiple unimodal stimuli get integrated into one multisensory representation or not is still to debate. If the later is the case the question arises at which processing stage the integration takes place. Is it in an early or at a later, higher order-processing stage. For an overview take a look at Quak, London, and Talsma (2015).

There are several studies that show a positive effect on memory when stimuli presentation was bimodal compared to unimodal presentations. In a study by

[1] In accordance with Stein et al. (2010) and most of the recent publications the term *sensory*, as in unisensory and multisensory, is used for the description of neural or behavioural processes like detection, perception or memory. The term *modal*, as in unimodal and bimodal, is used for the description of physical characteristics of objects and stimuli.

Thompson and Paivio (1994) participants had to memorize different types of stimuli, visual, auditory and audiovisual. The results showed an increase in recall when the learned stimuli where presented for both modalities. To rule out, that it is not a redundancy effect - solely produced through the availability of two target stimuli - Goolkasian and Foos (2005) showed that learning two stimuli within one modality, like a picture and a written word (e.g., a picture of a bike + the word "BIKE"), did not lead to better recall than just presenting the picture alone. But recall performance was increased by presenting two stimuli of different modalities, i.e., pairing a printed word or a picture with a spoken word.

In a recent study Brunetti, Indraccolo, Mastroberardino, Spence, and Santangelo (2017) could show the influence of congruent vs. incongruent bimodal stimuli in a working memory task. Participants were shown a stream of bimodal (visual+auditory) stimuli in a 2-back task. In a 2-back task a participant has to decide whether or not a shown stimulus (target stimulus) in the same as the one two steps before (sample stimulus). In their task only the stimuli of one modality were of relevance, i.e., participants had only to concentrate on visual stimuli and ignore the auditory ones. The presented stimulus of the task irrelevant modality could be either congruent or incongruent. There were two different sets of stimuli that were used, quantities and digits. Quantities were 1-4 black disks as visual stimuli and 1-4 sinus bursts as auditory stimuli. Digits were the written (visual) and spoken (auditory) numbers 1,2,3,4. For example a participant had to concentrate on the black disks. The congruence could either occur at sample and target stimulus (e.g. three disks paired with three sinus bursts at target and sample stimulus), only at the sample stimulus, only at the target stimulus or at neither one. If the target stimulus was paired with a congruent stimulus in the non-relevant modality participants had significantly shorter reaction times compared to incongruent stimuli (regardless how the sample stimulus was paired). But the better RT performance occurred for auditory targets only for quantities and for visual targets only for digits.

In a study by Heikkilä, Alho, Hyvönen, and Tiippana (2015), participants had to memorize different sets of unimodal (visual or auditory) stimuli, that were shown simultaneously with a stimulus of the other modality. Participants were asked only to concentrate on one modality and to ignore the second one. The stimuli of the second modality could either be congruent (e.g. a picture and the 'baa' of a sheep), neutral (e.g. white noise or meaningless written stimuli, like *XXXXXXX*) or incongruent (e.g. a picture of a cow and the sound of a foghorn). Thus the stimuli of the second modality only sometimes bared meaning (congruent) and the other times where meaningless (neutral, incongruent). After the encoding phase the recognition phase followed. In contrast to the encoding phase stimuli were shown unimodal. The learned stimuli were mixed with novel ones and participants had to decide which stimuli had been shown in the encoding phase. There were five different conditions (the italic written presentation form was the one to be learned). *Pictures* with sounds (1), *sounds* with pictures (2), *written words* with spoken words (3), *spoken words* with written words (4) and *pictures* with written words (5). The results show a significant better memory performance for congruent stimuli compared to neutral stimuli in condition (2), (3) and (4). In condition (1) and (5) there were no significant differences. The incongruent stimuli were not different from the neutral stimuli in either condition. The authors interpreted these results as evidence, that the audiovisual integration of bimodal stimuli result not only in better perception but also in better memory performance and "that congruent multisensory stimuli may receive more effective encoding than unisensory stimuli" (Heikkilä et al., 2015).

In a recent study, Xie, Xu, Bian, and Li (2017) showed a significant reaction time (RT) benefit in working memory in bimodal trials. In this study participants were shown either a picture, a tone or a bimodal and congruent presentation of a picture and a tone, for 0.6 *s*. Then the screen was black for 2 *s* and then again a stimulus was shown. Participants had to decide as fast as possible, whether the stimulus was the same or a different one compared with the stimulus that had been shown before the black screen.

In 2015 Hardiess, Erhardt, and Mallot conducted a 2-back task in which stimuli were either presented unimodal (visual or auditory) or bimodal. They showed that participants performed significantly better in the bimodal condition compared to the unimodal conditions. In that experiment the visual stimuli were 12 (6+6) random dot patterns, each consisting of six black dots. Auditory stimuli were six different chords played either on a piano or with a guitar. The stimuli were different in the bimodal condition compared to the unimodal conditions (e.g. visual: 6 dot patterns, auditory: 6 piano chords; bimodal: 6 different dot patterns + guitar chords).

1.1 Working memory

The traditional and most widely-used working memory model is the multicomponent model of working memory introduced by Baddeley and Hitch (1974). In its original form it consists out of three parts, the central executive and the two domain[2] specific slave systems, the phonological loop and the visuo-spatial sketchpad. The central executive is the control centre, coordinating cognitive processes, attention and directing information between the two slave systems. The task of the phonological loop is to maintain acoustic and speech-based information (e.g. a phone number or a list of words). The visuo-spatial sketchpad is responsible to maintain visual and spatial information (e.g. the location of an item in a room). According to the original multicomponent model, information is stored domain specific in the correspondent subsystem. Twenty-six years after its introduction a third slave system, the episodic buffer was introduced (Baddeley, 2000; Baddeley, Allen, & Hitch, 2011). One reason it was added was to account for the growing evidence of interactions between phonological and visual processes, that could not be accounted for with purely domain-specific subsystems. The episodic buffer is a limited capacity, passive system. It integrates visual, spatial and verbal information and links it with time (e.g., the scene of a movie). It also has links to the central executive and long term memory (Baddeley, 2012). A schematic representation of the multicomponent model of working memory can be seen in Figure 1.

[2] The term domain has to be distinguished from the term modality. As mentioned above modality describes the physical properties of a stimulus. On the other hand, domain as used by Baddeley describes the way a stimulus is encoded in the working memory. For example a written word is a visual stimulus and a spoken word is an auditory stimulus. But both stimuli are probably maintained within the phonological loop and are therefore within the phonological domain.

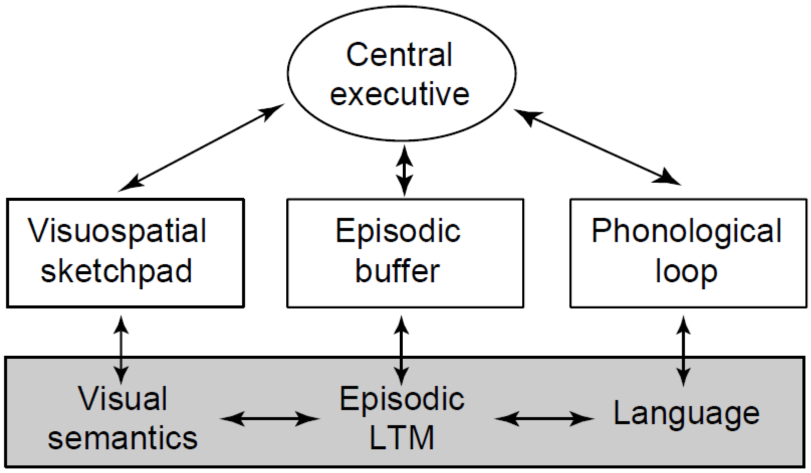

Figure 1: Multicomponent model of working memory. Picture taken from Baddeley (2000).

The second most used and second best known model is the embedded processing model of working memory by Cowan (1999), as can be see in Figure 2. In contrast to Baddeley's model, working memory and long term memory are parts of the same system. In this model working memory is a subset of the long term memory. At each given point in time a subset of memory traces are active. The part of this subset which is in the so called focus of attention and awareness is what Cowan defines as working memory. The activation of memory is time-limited and the focus of attention is capacity limited.

Both working memory models have in common that the capacity of the working memory is limited (Baddeley, 2000; Cowan, 1999). The mean capacity limit is usually denoted as four objects (Cowan, 2010; Luck & Vogel, 1997) but differ between individuals from 1.5 to 5 objects (Vogel & Machizawa, 2004) and can be dependent on the difficulty of the task (Turner & Engle, 1989).

It is not clear what causes the capacity limit. It could be that the holding device, like the episodic buffer, is limited itself. Another equal probable explanation is that an unlimited holding device is limited by attentional resources (Morey &

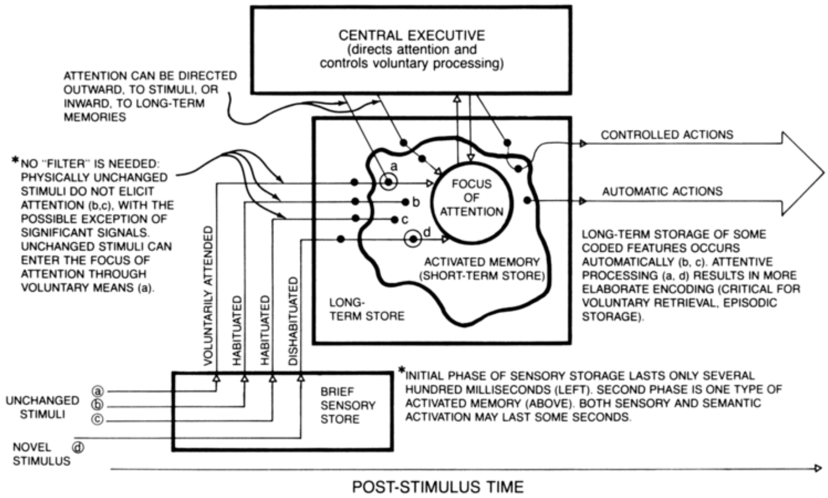

Figure 2: Embedded processing model of working memory. Picture taken from Cowan (1988).

Cowan, 2004).

It is to debate whether there are shared attentional resources, that have to be divided between different sensory modalities, or different sensory modalities have distinct attentional resources. Wahn and König (2017) showed in a recent review that resource allocation might be task dependent. For tasks where object-based attention is necessary, there seems to be distinct attentional resources for vision and audition but not for vision and touch. In other tasks (spatial attention, object based + spatial attention) attentional resources seem to be shared or at least partially shared depending on the sensory modalities.

As Quak et al. (2015) note, it is not clear how multimodal information is encoded in working memory. It could be possible that unisensory representations are stored separately and are integrated at a later point or that the information of stimuli of different modalities gets integrated and stored as a multisensory representation instead. Depending on the specific task and the minute circumstances there could be argued that both ways of integrating information are mechanisms found in the working memory.

In the framework of the multicomponent model of working memory the modalities, or rather domains, are separated by definition and integration has to be at a later point in time.

The embedded processing model of working memory works with both alternatives. An early integration could lead to one multisensory memory trace and a later integration to two separate memory traces that are in some way connected. The model can even account for both mechanisms being used at the same time.

1.2 Integration on neuronal level

On neuronal level there is evidence for different intensities of multisensory integration. With the integration of different senses the detection of an event is more likely because the evidence from multiple sources can be combined. Stanford and Stein (2007) give a review over the topic. They define the different intensities of integration on neuronal level. Depending on the strength of the input signals the strength of the output signal changes. If the input signals are weak, then the integrated output signal is greater than their sum, which they called superadditive. If the two input signals are of medium strength than the output signal is the sum of the two input signals, called additive. If the input signals are strong, then the output signal is not much stronger than either input signal, called subadditive. This behaviour is illustrated in Figure 3.

Superadditive integration could be shown in in different areas of the brain of the cat and of the monkey (Angelaki, Gu, & DeAngelis, 2009).

Giard and Peronnet (1999) showed that participants were faster and more accurate in an categorizing task, when objects where presented bimodal. They also showed that multisensory integration of audio-visual stimuli can take place as early as 200 ms post stimulus presentation.

Werner and Noppeney (2009) conducted a study in which participants had to categorize sounds and/or pictures of tools and instruments. These stimuli where either presented unimodal or bimodal and could be either intact, degraded (absolute threshold) or noisy (no detection possible). For intact bimodal stimuli they showed primarily subadditive multisensory interactions in the superior temporal sulci. For degraded stimuli the multisensory interactions became "additive (with a

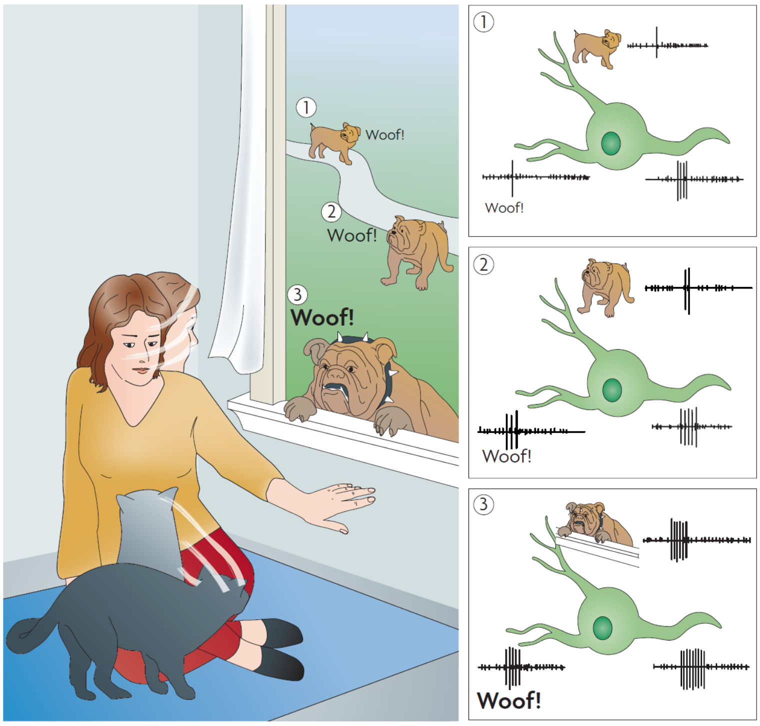

Figure 3: Different size of neuronal integration depending on the input signals. ① Weak input signals producing a superadditive output signal. ② Medium input signals producing as additive output signal. ③ Strong input signals producing a subadditive output signal. Picture taken from Stein and Stanford (2008).

trend to superadditive)" (Werner & Noppeney, 2009). Furthermore, for degraded stimuli participants showed significantly better categorization performances for bimodal stimuli compared to the best unisensory performance and most participants performed significantly better than what could be accounted for if stimuli were processed independently.

1.3 Experimental setup and hypothesis

To further investigate multisensory processes in working memory the difference working memory performances of participants at different levels of stimulus detection were examined.

More precisely at first a psychometric function (PF) was fitted for the detection of the visual and auditory stimuli that were covered with different amounts of white noise. From the PFs the signal to noise ratio (SNR) at absolute threshold (50% detection rate) and the SNR at 75% and 99% detection rate were extracted.

Then a working memory task (2-back task) was conducted with stimuli of the three different detection rates. In a 2-back task a participant is shown a stream of stimuli. At each trial (beginning from the third) they have to decide whether the shown stimulus (target stimulus) is the same as the stimulus that was shown two trials before (sample stimulus). This stream of stimuli was either unimodal (visual or auditory) or bimodal. For each modality stimuli of either one of the the three different detection rates were used. Thus, in total there were nine different conditions of the working memory task.

For stimuli at 99% detection rate the working memory task was similar to the one conducted by Hardiess et al. (2015) with the difference that stimuli were congruent.

Thus, for stimuli at 99% detection rate a better working memory performance in the bimodal condition compared to the unimodal conditions was expected.

For stimuli with lower detection rate (50% and 75%) a larger multisensory benefit than for stimuli at 99% detection rate was expected. Whether the multisensory benefit is larger for 50% or 75% detectable stimuli was subject of the experiments. There are arguments for either detection rate leading to a larger multisensory benefit (section 2).

The overall memory performance will almost certainly be worse for lower detectability of stimuli than for better detectable stimuli.

The multisensory benefit might be larger compared to Hardiess et al. (2015) because of the fact that congruent stimuli were used in the present task. As discussed earlier a congruent presentation does usually lead to better working memory performances.

To explore the potential different sizes of multisensory benefits in the 2-back task, I present a framework which allows (with some assumptions) to differentiate between multisensory benefits that are due to perception and multisensory benefits that are the due working memory processes. This makes it possible to compare the working memory performances for stimuli at different detection levels.

In the experiment by Hardiess et al. (2015) a performance increase showed, where participants performance increased abruptly for the last two of five blocks. This effect was most likely due to unbalanced blocks, rather than a learning effect. The first blocks contained considerably less go trials than the last two. If a learning effect is present, either of the stimulus material or of the task itself then this effects are counterbalanced for the 9 different working memory conditions between participants. The only learning effect that still might show is a learning effect between the two sessions. But this effect should be similar for all nine conditions of the working memory task and therefore it should have little influence of potential differences between conditions of the working memory task.

2 Separating perception benefits from working memory benefits in bimodal stimuli presentations - a signal detection theory approach

In bimodal conditions in the working memory task, in settings with stimuli that are not always detected, e.g. a stimulus at absolute threshold, a potential multisensory benefit should be attributed to perception as well as working memory processes.

First the issue is illustrated with a non-laboratory example. Then, to separate the perception and working memory performance, a set of assumptions are made and a theoretical measurement is introduced based on signal detection theory, that can help to interpret potential multisensory benefits for stimuli that are not always detectable.

A participant has to stand next to a street and 30 m across is a building with two balconies 10 m apart. A dog is standing on one of them. We ask the participant the question on which of the two balconies the dog stands. This might be an easy task if the sight is clear, but it gets harder if there is some mist in the air. We can ask the question at which density of mist our participants can still detect the person. To get the absolute threshold (defined as detecting a stimulus half of the time) we measure the density of the mist at which our participant can detect the dog at 50% of the time. The percent of correct answers corresponding to the absolute threshold varies depending on the experimental setting, i.e., how we ask the question and what a participant can answer. In our example the absolute threshold lies at 75% correct answers. By definition participants know the right answer 50% of the time, i.e., they can see the dog on one of the balconies one out of two times. In the cases they do not see the dog, they have to guess. The probability to guess the right balcony is 50%. Thus, they end up with 75% correct answers.

This example is a so called two alternative forced choice (2AFC) experiment. In such an experiment participants are shown two stimuli at once but only one of them contains the target and the other one only contains noise. The absolute threshold of an 2AFC experiment is defined, as stated above, at 75% correct answers. In a categorization experiment, where a participant has to correctly assign a stimulus to one of three categories, the absolute threshold lies at $\approx 67\%$ correct answers. In general the percent of correct answers (percent correct, abbreviated pc) that correspondents to the the absolute threshold can be calculated by $pc = 0.5 + \frac{0.5}{n}$, where n is the number of different answer possibilities (assuming answers are equiprobable).

But if now the dog barks while standing on the balcony, the setup changes. Participants are now presented with a stimulus at two different modalities simultaneously. If the mist is too dense to see anything, the participant could still answer the question correct, because they can hear the direction the dog is barking from. However, if there is noise from cars driving by, it gets more difficult again.

We now want to assess the absolute threshold in the case of two stimuli as in a bimodal stimuli presentation, i.e., hearing and seeing the dog simultaneously. If a participant can detect either of the targets alone with 50% probability, the question is with which probability can they detect the dog if both stimuli are present. This depends on which way they incorporate the information of the underlying one-dimensional probabilities.

For compound stimuli the one-dimensional criterion becomes a two dimensional decision boundary. There are some two-dimensional decision rules that can be analysed in one dimension.

The easiest but unlikely adequate decision rule is the so called decisional separability. Using this rule a participant simply ignores one of the two stimuli and the decision is based on the remaining stimulus alone (Figure 4 A).

If we consider both stimuli as independent from each other there are two possible rules a participant can rely their decision on. If the stimuli are processed independent from each other than a participant has two independent criteria, one criterion for each stimuli. One is called the maximum rule, the other one minimum rule.

The maximum rule states that a participant only responds with yes to the two shown stimuli, when both underlying single signals lie above each independent criterion, i.e., the participant sees and hears the dog (Figure 4 B).

The minimum rule, also called probability summation model states that a participant will respond with yes, if at least one of the two single signals lies above-criterion, i.e, the participant either sees or hears the dog (Figure 4 C).

There is another rule, the optimal rule, which will lead as the name suggests, to the optimal decision based on two compounds. But for this rule a participants need to integrate the two stimuli (Figure 4 B).

The optimal rule says an observer answers with yes if the sum of the two single signals lies above a single criterion. In this case if one signal is weak it can be compensated with the other stronger signal and the decision is made considering both signals together opposed to the rules before where the decision was made based on each signal considered alone. (Macmillan & Creelman, 2004; Treisman, 1998).

In our example participants probably use the probability summation model (PSM) if the stimuli are processed independently. They know every time they perceive a stimulus what the correct answer is, i.e., if they hear or see the dog they know which balcony the dog stands on. Thus it is only necessary to detect one stimulus in a trial to make the right decision. If participants do not process the stimuli independent, they integrate the stimuli and the optimal rule is used.

As is noted by Schwarz and Miller (2014) the use of the PSM on percent correct is not entirely accurate and leads to wrong assumption of gained redundancy. This is due to the fact that for multiple channels (stimuli) not only the hit rate increases but in the same way does the false alarm rate. To investigate a sensitivity increase of the bimodal detection, the false alarm rate has to be measured as well as the hit rate. For this purpose noise trials have to be included into an experimental task. This was not done in this study in the threshold estimation tasks, because it was noticed not before data collection was completed.

Despite this concern the PSM is stilled used in some computations later for an ideal observer to estimate an upper limit for independent processing of stimuli. This might lead to an overestimation of the computed bimodal performance for an ideal observer in this case. Thus, to rely on this measurement of an upper limit for independent processing, noise trials have to be included in future work. Therefore, this upper bound is not used for interpretations in this study.

Furthermore, as Treisman (1998) states there are two stages of decision making

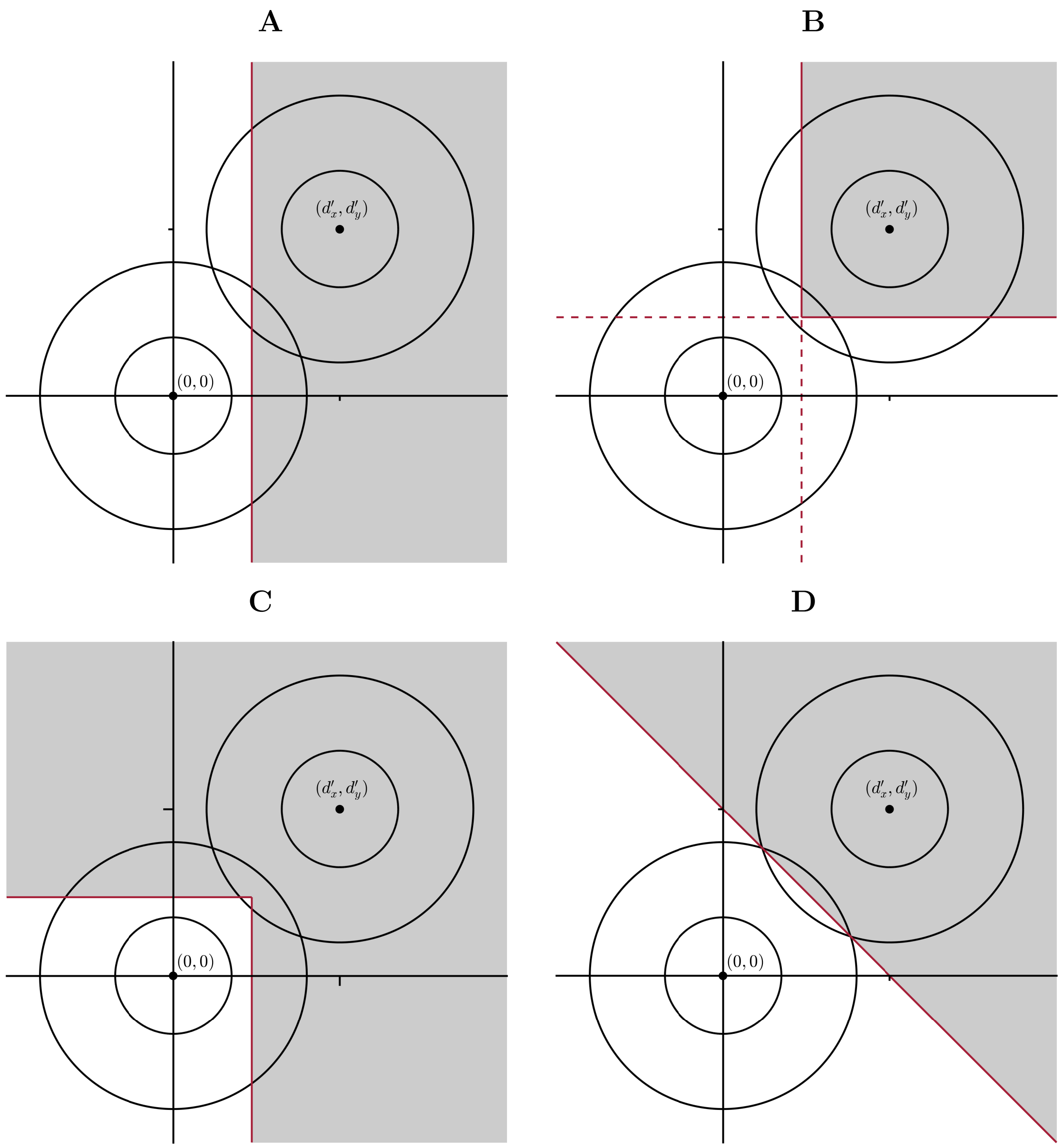

Figure 4: Each of the four panels show one decision rule for compound stimuli. **A**: Decisional separability. **B**: Maximum rule. **C**: Minimum rule. **D**: Optimal rule. The red lines are the decision boundaries. The grey area marks when a participant answers "yes" and the white area marks when a participant answers "no". Adapted from Macmillan and Creelman (2004).

(at least if both stimuli are independent from each other). In the first stage two (covert) independent decisions are made, whether the stimulus (signal) is present or not. In the second stage these two independent decisions are combined to one final decision as in the maximum or minimum rule. As presented above, both independent decisions are treated with the same weight, i.e., they were given the same amount of importance. This might not be the norm, especially if the two single stimuli are not equally easy to detect or when single signals can contain contradicting information. But the weight for both stimuli classically is seen as equal.

In the present case the assumption of equal weights can be assumed since the stimuli are controlled to have the same unimodal detection performance.

If the detection rate of each independent stimulus is 100%, i.e., there is no mist and no noise, then there is no benefit in hearing and seeing the dog at the same time. But for stimuli that cannot be detected 100% of the time a participant will very likely have a higher detection rate, a benefit, for the bimodal presentation. This potential benefit is called the multisensory benefit. It is defined as the performance increase in a bimodal setting compared to the best unimodal setting. If we consider *i*) the two stimuli as independent from each other, *ii*) the detection of one stimulus as sufficient to give the right response and *iii*) both stimuli with the same weight, than we can compute the detection rate of the bimodal presentation with the following formula for the PSM, given by Treisman (1998).

$$P_{VA} = P_V + P_A - P_V \cdot P_A \tag{1}$$

If the measured bimodal detection rate of a participant is significantly better than the one predicted by the PSM, than we can conclude, that the two stimuli cannot be processed independently. Then there has to be some sort of integration (Treisman, 1998). Note that the opposite does not hold. If the measured detection rate is worse than the predicted one we cannot argue that the stimuli are processed independently from each other, we simply do not know. An integration must not necessarily yield to a better performance than two independent processes. The PSM thus gives us a framework to check whether the performance of the participants in the bimodal detection task can be accounted for without integration or not. It can be seen as an upper limit for independent processing of two stimuli.

There is a large literature for the possible ways of integrating two stimuli and for modelling it (Macmillan & Creelman, 2004; Treisman, 1998) and the differences of the integration within and between senses (Ban, Preston, Meeson, & Welchman, 2012; Ernst & Banks, 2002; Ernst & Bülthoff, 2004; Hillis, Ernst, Banks, & Landy, 2002; Nandy & Tjan, 2008).

A potential multisensory benefit in the working memory task can be the result of a better perception of the bimodal stimuli (as illustrated above) and working memory processes. In the following paragraphs I try to separate the two processes from each other. With the help of signal detection theory I try to formalize the part of perception and the part of working memory that contribute to a multisensory benefit. This formalization can later help with the interpretation of the data.

According to the detectability of the stimuli a potential multisensory benefit in a working memory task has to be interpreted differently. For 100% detectability a multisensory benefit has to be attributed completely to a better working memory performance. For stimuli with non-perfect detectability the interpretation of a multisensory benefit gets more complicated. As discussed above the detection rate changes for bimodal stimuli. It is not trivial to measure the absolute threshold for a bimodal presentation of two stimuli because there are two stimuli to adjust and we want to have the same weight for both stimuli. But we can measure the percent correct of a bimodal stimulus presentation and compute the detection rate from it (see Equation (7)). Then, with the help of some assumptions, we can try to estimate to which extent a multisensory benefit leads to better perception and to which extent it is due to an increase in working memory performance.

Let us assume that participants have the same working memory performance (WMP) independent from the detection rate of the stimuli but dependent on the modality of the stimulus and the stimulus material itself. Then we can make use of the measured working memory performance ($\mathrm{WMP_m}$) of each participant for 100% detectable stimuli. Together with the measured bimodal and unimodal detection rates we can compute a theoretical working memory performance ($\mathrm{WMP_t}$) for each modality and stimulus detectability. If we compare the $\mathrm{WMP_m}$ with the $\mathrm{WMP_t}$

we can make the following propositions. If the $\mathrm{WMP_t}$ is better than the $\mathrm{WMP_m}$, the WMP was worse than it is for 100% detectable stimuli. If both WMPs are the same, we can assume that there is no difference between WMP for degraded and 100% detectable stimuli. If the $\mathrm{WMP_t}$ is worse than the $\mathrm{WMP_m}$, the WMP was better than for 100% detectable stimuli and a possible greater multisensory benefit cannot be attributed solely to a better bimodal perception performance. The probable interpretation would be a better WMP for degraded bimodal stimuli.

The details of the WMP computations are a little more complicated than stated above, depending on the task and the measured variable. In this work the working memory task used is is a 2-back task with 3 different stimuli. We measure the WMP using the sensitivity index d' which is computed from the hit rate and the false alarm rate. Therefore, we have to use them for our calculations. The details of the computations look as follows:

- Measure the detection rate DR for unimodal and bimodal stimuli (as we did in the threshold estimation task).
- Compute the detection rate in the working memory task $DR_{WM} = DR^k$. k depends on the task. In our case $k = 2$ because a participant only can make a correct response if they detect the stimulus in both trials, the sample trial and the target trial.
- Measure the hit rate H and the false alarm rate FA for each modality at 100% stimulus detectability.
- Compute the theoretical hit and false alarm rates H_t and FA_t. For this we assume that participants use the underlying distributions as guessing rate, or rather as the probability to say yes (P_{yes}) in trials they do not perceive a stimulus. This is used to compute the hit rate of the guessed trials H_{guess} and the false alarm rate of the guessed trials FA_{guess}. It should be noted that this is not the optimal guessing strategy.
  The optimal guessing strategy depends on the distribution of go and no-go trials. If there were more no-go trials present then the highest d' will be achieved for always answering with no. If there were more go trials present then the highest

d' will be achieved for always answering with yes.

This is due to the fact that the hit and false alarm rate do not influence d' in a linear way. The sensitivity index d' is computed with the inverse of the cumulative distribution function of the Gaussian distribution which grows much stronger from approximately 0 to 0.1 and 0.9 to 1 than for intermediate values.

$$H_t = DR_{WM} \cdot H + H_{guess} \cdot (1 - DR_{WM}) \quad \text{with } H_{guess} = P_{yes} \cdot P_{go} \tag{2}$$

$$FA_t = DR_{WM} \cdot FA + FA_{guess} \cdot (1 - DR_{WM}) \quad \text{with } FA_{guess} = P_{yes} \cdot P_{no-go} \tag{3}$$

- Adjust hit and false alarm rates if necessary. For a hit rate of 1, the hit rate gets adjusted to $H_t = \frac{\#go\ trials - 0.5}{\#go\ trials}$ and for a hit rate rate of 0 $H_t = \frac{0.5}{\#go\ trials}$. For false alarm rates the same adjustment with #no-go trials respectively (Stanislaw & Todorov, 1999).
- At last compute the theoretical WMP with the inverse of the cumulative distribution function of the Gaussian distribution.

$$d'_t = Z(H_t) - Z(FA_t) \tag{4}$$

With the formula for $\mathrm{WMP_t}$ we can also calculate a theoretical upper limit for each working memory task. This upper limit is given by an ideal observer (a participant with a perfect memory). They would be able to always answer correct as long as they are able to detect the stimulus at a given trial and the stimulus two trials before.

In present setup we had three different stimuli. The trials consisted out of 54 signal trials and 102 noise trials. We assume that an ideal observer uses the underlying distribution of go and no-go trials as their guessing strategy. We use the PSM to compute the detection rate of bimodal stimuli. As stated earlier this is not entirely accurate.

For 100% detectable unimodal stimuli this yields to a sensitivity index of $d' \approx 4.94$ for unimodal as well as for bimodal conditions. For unimodal stimuli at absolute threshold this yields to a sensitivity index of $d' \approx 0.56$ for unimodal stimuli and a sensitivity index of $d' \approx 1.60$ for bimodal stimuli. Changing the unimodal detection rate to 75% we get a sensitivity index $d' \approx 1.60$ for unimodal stimuli and $d' \approx 3.18$ for bimodal stimuli.

An unintuitive but import fact is that it makes a difference at which point a participant guesses (following the ideas of Treisman (1998)). A participant can either already make a guess at the *detection step* or later at the *decision step*.

The *decision step* is the step at which a participant has to make the decision which answer they give. If they do not detect a stimulus they still have to make a decision if this stimulus is the same as the sample stimuli was and take a guess. But n steps further when the target stimulus becomes the sample stimulus they still do not know which stimulus it was and they have to guess again.

In this case the participant can either guess both times correctly, both times wrongly or guess correctly one of the two times and the other time wrongly. If the participant guesses according to the underlying distribution of go and no-go trials $P_{go} = 1 - P_{no-go}$ then the probability of guessing both correct is $P_{go} \cdot P_{go}$, guessing one correct is $P_{go} \cdot P_{no-go}$ and guessing both wrong is $P_{no-go} \cdot P_{no-go}$.

The probability to make a correct decision on any trial the participant did not detect either sample or target stimulus, based on this guessing strategy, is

$$P_{go}^2 + P_{no-go}^2 \tag{5}$$

The *detection step* is directly after perception. If a participant did not detect the shown stimulus they can guess right away which stimuli it was. This means that they chose randomly between one of the stimuli and stick with it and memorize it as if they did perceive it. Thus, if a participant does not detect a stimulus he only has to guess once, when the stimulus is the target stimulus. When the stimulus becomes the sample stimulus they have already decided which stimulus it should be.

In this case the probability that they guessed right is dependent on the number of different stimuli used in the task and the distribution of go trials P_{go}. Assuming each stimuli is presented the same amount of times and a participant has no bias they will use the underlying stimuli distribution as their guessing strategy $P_{guess} = \frac{1}{n}$, where n is the number of different stimuli. Depending on if the trial is a go or a no-go trial the probability can be computed. If the trial is a go trial the possibility to make a right decision is $P_{go} \cdot P_{guess}$. If the trial is a no-go trial the possibility to make a right decision is $P_{no-go} \cdot (1 - P_{guess})$.

Hence, the probability to make a right decision in trials they did not detect a stimulus is

$$P_{go} \cdot P_{guess} + P_{no-go} \cdot (1 - P_{guess}) \tag{6}$$

As can be seen right away the two strategies will lead to the same outcome if and only if the number of stimuli is is the same as number of go trials ($P_{go} = P_{guess} = \frac{1}{n}$).

It is not easy to identify which strategy participants use and even harder to determine what their guessing rate is.

But with $P_{go} = \frac{1}{n}$ we can at least be sure that if a participant uses the underlying distributions as their guessing rate then the the choice of strategy does not matter.

In this case though one has to compute the perception rate of the bimodal stimuli from the measured percent correct (pc). Fortunately we can do that. The same way we compute the pc for a given perception rate in a specific experimental setting, we can do the opposite. In a 2AFC task, 50% perception leads to 75% correct answers. With the probability of guessing the correct answer (P_{guess}) and the detection rate (DR) we can computed the pc with the formula $pc = DR + (P_{guess}) \cdot (1 - DR)$. We can solve this equation for the detection rate and get:

$$DR = \frac{pc - (P_{guess})}{1 - (P_{guess})} \tag{7}$$

It is important to note, that the bimodal detection rate is the joint probability of three different events. It is the sum of the detection of both stimuli, the detection of only the visual stimulus and the detection of only the auditory stimulus.

For example, when a participant can detect a visual and a auditory stimulus at detection threshold (50%) the probability given by the PSM is $P(VA) = P(V) + P(A) - P(V) \cdot P(A) = 0.5 + 0.5 - 0.5 \cdot 0.5 = 0.75$. The multisensory benefit, defined as the difference between the bimodal performance and the best unimodal performance, is $P(VA) - P(A) = 0.25$. However the probability that a participants detects both stimuli at once is given by $P(V) \cdot P(A)$. The probability that they detect one stimuli exclusively, either the visual or the auditory one, is given by $P(V) \cdot \neg P(A) + \neg P(V) \cdot P(A)$. Thus, if both stimuli are detected separately at absolute threshold, the simultaneously detection of both stimuli is 25%. The detection of either stimulus exclusive is 50%. This changes if both stimuli have a unimodal detection rate of 75%. The PSM predicts a bimodal detection rate of ≈ 0.94. That results in a multisensory benefit of approximately 11%. The probability to detect both stimuli simultaneously is $\approx 56\%$ and to detect either one exclusive is $\approx 38\%$.

This again has an influence of possible different multisensory benefits in the working memory task for different degraded stimuli and the interpretations. Because a potential working memory benefit has to be attributed due to both stimuli being detected at once. Thus for higher unimodal detection rates we get a higher share of simultaneous detection compared to detecting either stimulus exclusive.

3 Materials and Method

3.1 Participants

The participants of the experiment were 21 students of the University of Tübingen. Their participation was voluntary and as allowance they either received 20€ (16 participants) or a certification they needed to fulfil the requirements of a course (5 participants). Fifteen of the 21 participants were female, six male. Sixteen participants were right-handed. The mean age was 23 years (18-28 years old).

Due to problems with threshold measurements and psychometric function fitting six of the 21 participants had to be excluded from the analysis (Section 4.1). The remaining 15 participants were 13 females and two males. Thirteen of them were right-handed. The age of the remaining participants reached from 19 to 28 years, with a mean age of 22.4 years.

All participants performed all tasks in two session. Both session roughly lasted 60 *min* and were conducted at two (consecutive) days.

3.2 Stimulus generation

To generate a set of stimuli Gabor patches for the visual stimuli and a major chord for the auditory stimuli were used. Different amounts of white noise were added to achieve different levels of detection.

Three different stimuli were created for each of the two modalities. These three stimuli had an orientation classification. They could either be *left*, *middle* or *right*. The auditory *middle* stimulus was the chord played as sounding from the front, or as can be perceived with headphones, "inside" the head. The visual *middle* stimulus was a vertical Gabor patch. The *left* and *right* auditory stimuli were perceived as sounding from the corresponding side. The *left* and *right* visual stimuli were the Gabor patch tilted to the according side.

Therefore, the visual and auditory stimuli had an intrinsic similarity. This similarity was always present if stimuli were presented bimodal. Therefore, bimodal stimulus presentations were always congruent.

Even though the auditory *middle* stimulus can be perceived as sounding from the front or inside the head and the visual *middle* stimulus as vertical or not tilted, participants most likely make the connection between the stimuli of different modalities and the three different orientations. The instructions made it clear that these three orientations exist and stimuli are always paired according to their orientation. (Figure 48)

Since the literature shows that absolute thresholds vary between people, the thresholds were measured for every participant. For visual stimuli - in particular the contrast sensitivity function (CSF) - differences can have various causes. Elderly people experience a natural decline in sensitivity to higher spatial frequencies (Owsley, Sekuler, & Siemsen, 1983; Schieber, 1992) as can be seen in Figure 5. Also various different impairments and diseases of the eye can lead to a decrease of the CSF (Coren, 2003). Li, Polat, Makous, and Bavelier (2009) discovered that the CSF can be enhanced via neural plasticity. They showed that playing action video games enhance contrast detection of a single low-contrast Gabor patch. Despite all those possible causes healthy participants in the same age group should have a very similar visual threshold.

Sensitivity of auditory detection, though, varies in a severe way. As much as 20 dB above or below the mean threshold at a specific frequency is still considered as "normal hearing" (Moore, 2012).

3.2.1 Visual stimuli

The visual stimuli used in the experiment were Gabor patches of elliptical shape of different orientation. All Gabor patches had a black centre. For the three different orientations angles of $-45°$, $0°$ and $+45°$ were chosen. In Figure 6 one can see the three different Gabor patches without any noise. In the experiment these Gabor patches were masked by different intensities of additive white Gaussian noise (AWGN).

The stimuli are created on the fly during the experiment using a script developed for this use. Thus, every aspect of the Gabor patches can be controlled and customized. The most important parameters outside of the direction (the angel) and the noise in dB, are the wavelength of the underlying sinus, set to 22 $pixels$, and the variances of the Gaussian of 10 $pixel$ in x direction and 30 $pixel$ in y

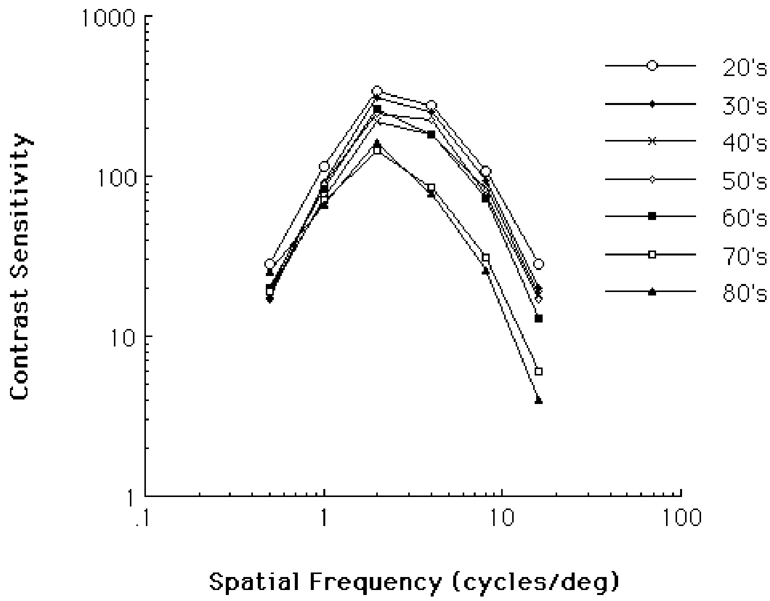

Figure 5: Contrast sensitivity functions of different age groups adapted from Schieber (1992).

direction. Furthermore the contrast was set to 1 and the phase of the sine wave was set to 0.75 to get a minimum at the centre of the picture, i.e., a black centre. The computed matrix of the Gabor patch is than transformed into 8 bit unsigned integers, hence containing intensity values from 0 to 255. With a fixed distance from the eyes to the screen approximately 60 cm and the above stated values of the Gabor patch the spatial frequency is approximately 2 $\frac{cycles}{degree}$, at which the contrast sensitivity is at its highest (max CSF between $2-5$ $\frac{cycles}{degree}$ (Campbell & Robson, 1968) (see also Figure 5).

The stimuli are normalized, after adding AWGN, to the dynamic range, so no clipping can occur. This means the extreme values are mapped to the maximal and minimal value of the digital representation. Therefore, for visual stimuli that are represented in 256 grey steps, the minimal value correspondents to 0 and the maximum correspondents to 255.

Figure 6: The three Gabor patches used in the visual experiment without AWGN.

3.2.2 Auditory Stimuli

For the auditory experiment a C major chord played on a guitar with nylon strings was chosen. Corresponding to the directions of the visual stimuli, the tone could either sound from the left, from the middle or from the right.

The perception of localization of sound stems from several different mechanisms. A sound played on the right side of the head needs longer to reach the right ear than it needs to reach the left ear. This difference is called interaural time difference (ITD) and is most prominent in lower frequencies. There also exists an intensity difference between sounds from different directions. This interaural intensity difference (IID) stems from the fact, that the the head is an obstacle for the sound wave and thus the intensity of the sound is less on the far side from the sound source compared to the facing side. The IID is most prominent for frequencies below 500 Hz in normal circumstances (Blauert, 1997; Feddersen, Sandel, Teas, & Jeffress, 1957; Moore, 2012). But for sound sources that are very close to the head of the listener the situation changes and IID persist even at low frequencies (Brungart & Rabinowitz, 1999).

Another very important mechanism is the so called head-related transfer function (HRTF). The HRTF describes how the shape of the head, the pinnae, ear canal and the nasal cavity all transform the sound depending on its localization and its frequencies. Some frequencies get boosted, some others get attenuated (Blauert, 1997; Moore, 2012). The HRTF varies significantly between persons and is the main reason sound via headphones is perceived in a different quality.

In the present experimental setup participants are wearing headphones. There are several disadvantages when using headphones, e.g., missing head shadows, room acoustics and pinnae reflections to name a few (Moore, 2012). But there are advantages as well. Headphones isolate the participants from disturbing noise, participants can move their body and head without changing their localization to the sound source and all parameters of the sound can be controlled way easier.

For the perception of localization in the present experimental setup only the ITD was used. The IID was not used because the sound mostly contained frequencies below 500 Hz. The HRTF was also not used because of the excessive effort and the lack of equipment and time needed to measure it for every subject.

The perceived direction of the sound thus only depends on the delay between the left and right ear piece of the headphones. To achieve an effect of approximately $\pm 70°$ an time difference of 0.54 ms was used based on the work of Feddersen et al. (1957) (Figure 7).

The stimulus used in the experiment was a c major chord. The basic C major chord is a recording of the chord played on a guitar with nylon strings. This mono sound with 44100 Hz was taken from the freesound database[3]. For the present

[3] https://freesound.org/people/SpeedY/sounds/8612/

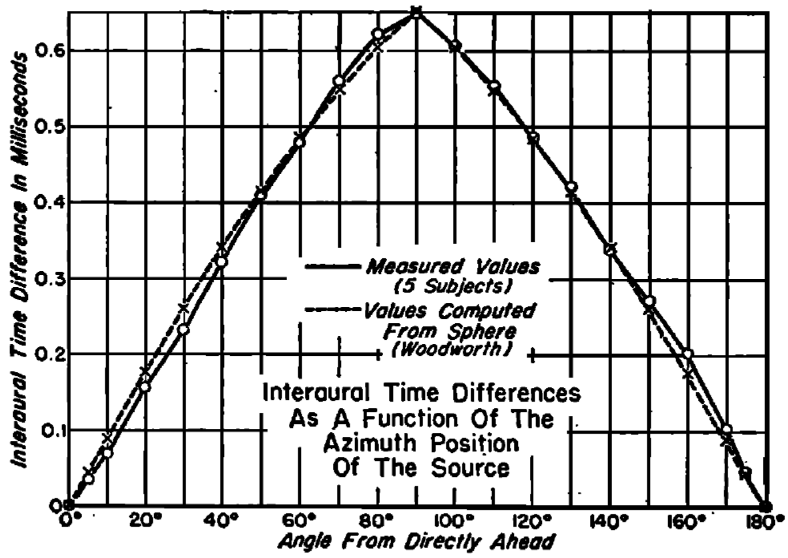

Figure 7: Interaural time difference as function of the azimuth (Feddersen et al., 1957).

purpose the sound trace was copied to create a stereo sound. Then the sound was cut to a length of 1000 ms. This was the basic sound that was manipulated in each trial of the experiments. To create the desired localization the necessary sound trace was shifted and zeros were added (no sound) in front. Then additive white Gaussian noise was added to both sound traces independently. A fade in and a fade out were added to make the sounds ending and beginning smooth without any prominent glitches. At last the stimuli was normalized to the interval of $[-1, 1]$ to prevent clipping and to get a loudness of approximately 55 dB_{SPL}.

3.2.3 Additive white Gaussian noise (AWGN)

The noise of the visual and auditory stimuli was created using the awgn() function of the *Communications System toolbox in MATLAB*.

This function creates AWGN based on the given signal to noise ratio (SNR) and adds it to the input signal in the below-mentioned way. For each value in the given, discrete input signal, a noise value gets added to it. Each noise value is drawn from a Gaussian distribution. The noise is independent and identically distributed (i.i.d.).

For visual stimuli a random value from the Gaussian distribution was added to each pixel. For the auditory stimuli a random value from the Gaussian distribution was added to each value of the digital representation of the audio signal. Independent noise was created for the left and the right audio channel.

Computation of AWGN:

i) Measure the power of the signal x, which is given in the discrete domain with n points:

$$P_{signal} = \frac{1}{n}\sum_{1}^{n}|x(n)|^2$$

ii) Compute the signal power in *db*:

$$P_{signal,dB} = 10 \cdot log_{10}(P_{signal})$$

iii) Compute the required noise power in *dB* for a given SNR in *dB*:

$$P_{noise,dB} = P_{signal,dB} - SNR_{dB}$$

iv) Add noise to the signal following the Gaussian distribution ($\sigma^2 = 1,\ \mu = 0$):

$$g(x) = \frac{P_{noise,dB}}{2\pi} \cdot e^{-\frac{1}{2}x^2}$$

3.2.4 Threshold estimation

To estimate the thresholds for three different detection rates of 50%, 75% and 99% four psychometric functions (PFs) for each participant were fitted for each modality.

PFs were fitted on the sampled data points, one for each stimulus. Then the data from the three stimuli (*left, middle, right*) was pooled. On this pooled data another PF was fitted. This PF then was used to determine the SNR for the three detection rates. These SNRs were then used for all stimuli in the bimodal threshold estimation task and in the working memory task. See Figure 8 for a schematic representation. In some cases this procedure had to be adjusted as described in Section 3.4.

For the threshold estimation tasks a categorization task was used in which a participant was presented with one stimulus per trial and the participant had to chose which category the stimulus belonged to. In fact a categorization task is not the preferable task to measure thresholds and a nAFC task should be favoured. The advantage of a nAFC task is that a possibly bias of participants will not show as likely in contrast to a categorization task.

Despite the disadvantage of the categorization task it was chosen because of its easier realization with three different stimuli.

Data was sampled according to the psi-marginal method for each stimulus (*left, middle, right*) separately for the visual and auditory modality. The psi-marginal method (N. Prins, 2013) is an adaptive, Bayesian method based on the the widely used psi method (Kontsevich & Tyler, 1999).

The psi-marginal method computes the stimulus for the next trial on the basis of the already shown stimuli in the trials before. After each stimulus presentation the method gets updated. More precisely, after each trial the Bayesian posterior distribution is calculated for all possible values (priors) of the slope β and of the threshold γ, that are treated as free parameters. The lapse rate λ is used as marginalized parameter, which means that it is variable but does not contribute directly to the entropy, which only gets minimize over the parameters γ and β. The stimulus level that is shown in the next trial is computed through minimizing the entropy for a set of possible stimulus values and their possible outcome of a positive or negative response. The value that leads to a minimal posterior entropy is chosen for the next trial.

We used three different psi-methods updater interwoven in one task. For each stimulus (left, middle, right) a separate PF was fitted and its entropy minimized. Thus, eight different PFs were fitted in total. One for the three different visual and auditory stimuli and one over the data of all three stimuli of each modality pooled together. This was done to detect possible different detection thresholds for the three different stimuli (*left, middle, right*), which are distinct particularly for the auditory stimuli.

3.3 Experimental Design

All experiments were running on a computer operating under windows 7 with 4 GHz of ram. The experiments were programmed using *Matlab R2015b* with the *Psychophysics Toolbox Version 3* (Brainard & Vision, 1997; Kleiner et al., 2007; Pelli, 1997), the *Palamedes Toolbox* (F. Prins N. & Kingdom, 2015; N. Prins, 2009) and the *Psignifit 4 Toolbox* (Schütt, Harmeling, Macke, & Wichmann, 2016).

To answer participants used an old *Logitech OEM* mouse with three buttons in the threshold estimation tasks. In the working memory task participants had to answer with a *HP office mouse* with 2 buttons.

In all experiments visual stimuli were presented on a *EIZO FlexScan L768 Slim Edge*. The brightness was set to 100 and the contrast to 80. Auditory stimuli were presented via *Audio-Technica ATH-M50* monitor headphones at approximately 55 dB_{SPL}. Participants were sitting approximately 60 cm in front of the screen wearing the headphones at all times. Participants were encouraged to make small breaks between blocks and experiments.

3.3.1 Threshold estimation tasks

The experimental setting was the same for visual and auditory threshold estimation task. Before the experiment started the screen showed instructions and the three different stimuli, without noise, were shown. The experiment then started after a press of the space key. At the beginning of each trial a fixation cross was shown for 500 ms. The participants were told to look at the fixation cross the whole time during experiments. In the visual task a stimulus got presented in the middle of the screen for 1000 ms followed by a visual mask for 150 ms to inhibit afterimages. In the auditory task a auditory stimulus was played over the headphones for 1000 ms and no mask was shown.

Participants had to answer with the three button mouse. The left mouse button had to be pressed when a *left* stimulus was shown, the middle mouse button when a *middle* stimulus was shown and the right mouse button when a *right* stimulus was shown. The participants were able to answer beginning with the stimulus onset. The next trial started only if an answer was given, but not before the end of the mask time (in the auditory task no mask was shown but the the same time was added before a trial could start). A schematic representation of a trial of the visual threshold estimation task can bee seen in Figure 9.

The visual and auditory threshold estimation task each consisted out of 210 trials in total which were divided among two blocks containing 105 trials each. Every stimulus was presented 70 times. The stimulus order was randomly determined at the start of the experiment. The participants were free to decide when to start each block and could take a break between the two blocks.

The setting of the bimodal detection task was analogue to the visual and auditory detection task. But in each trial an auditory and a visual stimulus were presented together. In total the experiment consisted of 90 trials divided into two blocks of 45 trails each. Presented were the three different stimuli level at 50%, 75% and 99% unimodal detection rate for each stimulus. These thresholds were the ones estimated in the visual and auditory threshold estimation experiments before. For each detection rate bimodal stimuli were shown 30 times (10 *left*, 10 *middle*, 10 *right*). In each trial a fixation cross was shown for 500 ms followed by the simultaneous presentation of the visual and auditory stimulus for 1000 ms. A visual mask was shown for 150 ms after stimulus presentation. The next trial started only if the participant had given an answer.

3.3.2 Working memory task

As working memory task a 2-back task was used. There were nine different conditions. For all three modalities (visual, auditory, bimodal) the three different detection rates of 50%, 75% and 99% were used. Each participant took part in all 9 conditions, resulting in 3×3 within subject design.

Each condition of the 2-back task consisted of 3 blocks with 54 trials each

(162 trials in total). They differ only in the modality presented (visual, auditory, bimodal) and the detection rate (50%, 75%, 99%). The 3 blocks of each condition were split among the two experimental sessions. In the first session each participant conducted one block of each condition in different orders (Section 3.3.3). In the second session each participant did the remaining two block of each condition. The third block of one condition followed directly after the second one. Participants were encouraged to take short breaks between blocks and conditions. They were able to start the next block on their own behalf.

Each block started with a short notification which stimuli would be used in the upcoming block(s). Once a block was started the fixation cross was shown for 2 s. Each trial started with the fixation cross shown for 500 ms. After that the stimulus (or stimuli if bimodal) were presented for 1000 ms. In the visual and bimodal conditions a mask was shown after stimulus presentation for 150 ms. A participant had 2 s after stimuli onset to answer. Regardless if an answer was given or not, 2 s after stimulus onset the next trial starts with the fixation cross being shown for 500 ms. To answer participants had to press the left mouse button in go trials (sample and target stimulus the same) and the right mouse button in no-go trials (sample and target stimulus different). A schematic representation of the working memory task can be seen in Figure 10. Participants were instructed to answer even if they were not sure whether the target stimulus was the same as the sample stimulus. Furthermore they were told that in bimodal conditions, again, a visual stimulus that was tilted to the left was always paired with an auditory stimulus sounding from the left side and so forth. (The complete stimulus instruction given in written form to the participants can be seen in Appendix C in Figure 49).

3.3.3 Task order

To counteract possible order and learning effects the task order was changed for each participant and divided into two sessions.

In the first session a participant conducted the visual and auditory threshold estimation task first for the three detection rates that are needed in the other tasks. Every second participant started with the auditory task followed by the visual and the other participants did it the other way around. After the detection rates were estimated all participants did the bimodal threshold estimation task. After that each participant did one block each of the nine different working memory tasks.

Session two consisted of the remaining 2 blocks for each condition of the working memory task. The sequence in which the nine different conditions had to be done was the same as in the first session for each participant.

The order in which a participant had to conduct the memory tasks were arranged according to the Williams design (Williams, 1949). The Williams designs is a (generalized) Latin square that is also balanced for first order carryover effects (see Appendix C, Table 3). This leads to 18 different task orders each used for one participant.

To counteract a possible learning effect of the go trial distribution in the working memory task, 27 different blocks were computed, three blocks being taken together for one specific trial-order. Each of these blocks has a different permutations of 54 the trials. Within a block each stimulus is presented 18 times and each stimulus is a go trial six times leading to 18 go trials and 34 no-go trials (the first two trials in each block are dismissed). A go trial is a trial were the target stimulus is the same as the sample stimulus. A no-go trial is a trial in which target and sample trial are different. For each participant each trial-order was used once. The trial-orders were counterbalanced over all participants, such that each trial-order was used twice in total for each condition.

3.4 Experimental adjustments

After the 4th participant attended the experiment one flaw of the setup became apparent. For three of the four tested participants the auditory detection threshold for the three different stimuli, left, middle and right, were too different leading to a bad fit when pooling the data of all stimuli together.

To fix this issue the experiment was changed for the following participants. After the visual and auditory threshold estimation tasks, the experiment supervisor had a look at the PFs. If they met the expectations, i.e., if the thresholds were similar for all three stimuli and the PF for the fit over the pooled data was decent, the experiment was continued as before. If the PFs did not meet the expectations, some further threshold estimations were done. For the modality the PFs were not good enough, the participant had to redo the threshold estimation task. It was the same task as before but only one PF was fitted over all stimuli. Thus, only one psi-method was used and updated for all stimuli. This second threshold estimation task had 105 trials divided between two blocks. With the second fit a better PF for all stimuli is achieved and the problem of the different detection thresholds for different stimuli is mostly avoided.

Even though the threshold differences still exist, at least the main hypotheses can be tested with it. In general a usable PF is obtained, leading to more sensible SNR values at 50%, 75% and 99% detection rate.

One downside of the method is that those PFs have a larger width as the one fitted for the single stimuli in most cases. This behaviour can be seen in the PFs in Appendix A. This is not the desired outcome, but it was the best available solution. Since the working memory task is completely balanced, the same amount of left, middle and right stimuli are assured as go trials and as no-go trials in each condition (but not in each block). Therefore, a better detection of one particular stimulus should be compensated for by the worse detection of another stimulus.

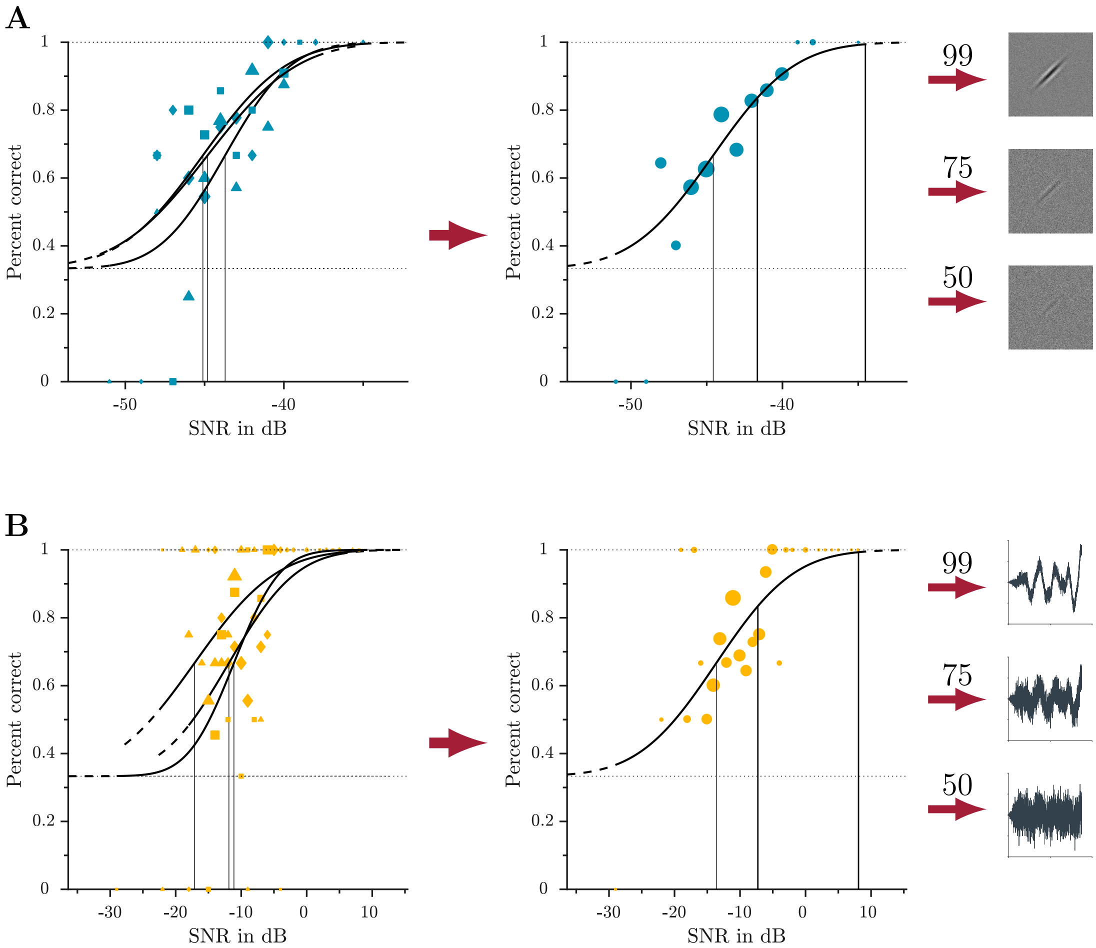

Figure 8: **A**: Left graph: Data points from the threshold estimation task for the three different visual stimuli and the fitted PFs over them. Diamonds are data points of the *left* stimulus, triangles are data points of the *middle* stimulus and squares are data points of the *right* stimulus. The size of data points indicate the amount of times they were presented. Middle graph: Pooled data points from the three different stimuli and the PF fitted over the pooled data. Again the size of the data points indicate the amount of time they were presented. Right graph: Stimuli with AWGN for the three detection levels. **B**: Equivalent representation for auditory stimuli.

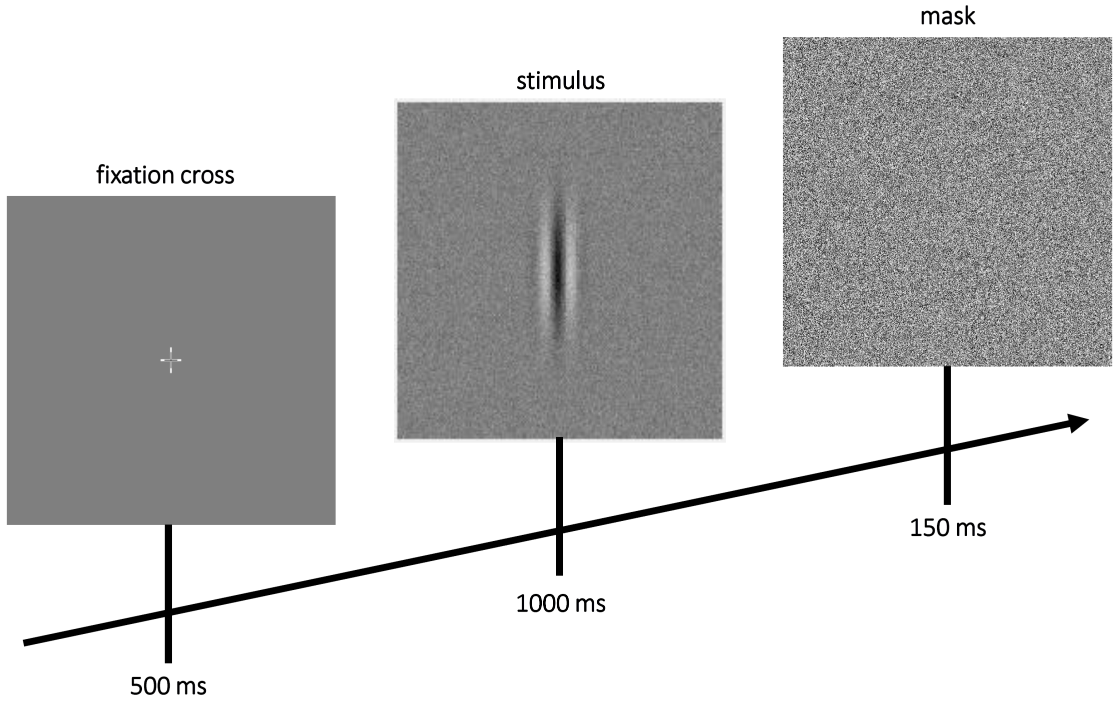

Figure 9: Schematic representation of one trial of the visual threshold estimation task. Each trial begins with a fixation cross shown for 500 ms followed by a stimulus shown for 1000 ms. After the stimulus a mask is shown for 150 ms.

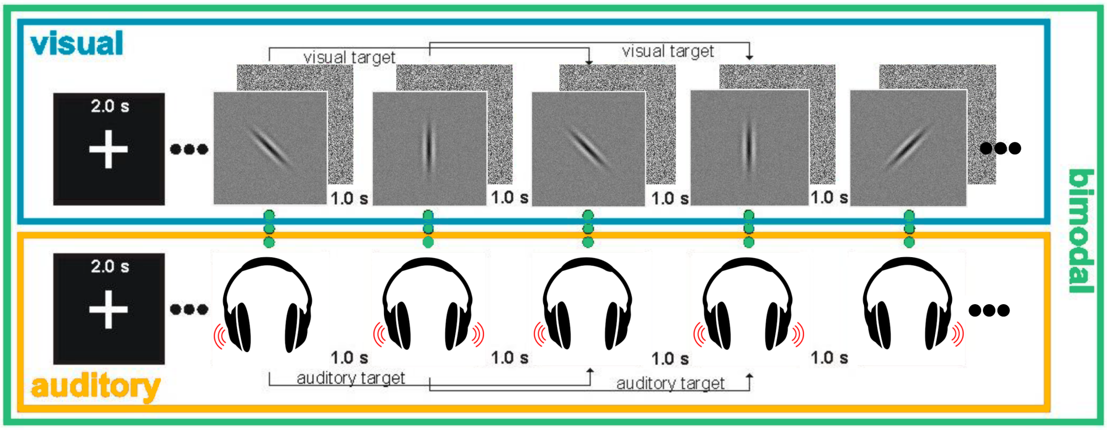

Figure 10: Schematic representation of five trials of the working memory task. In each trial a visual, auditory or bimodal stimulus is shown for 1000 ms. After a visual or bimodal stimulus presentation a mask is shown for 150 ms. Beginning with the third trial a participant has to decide whether the presented stimulus is the same or different as two trials before. In the bimodal condition the direction of the visual stimulus is always paired with an auditory stimulus sounding from the same direction.

4 Results

4.1 Excluded participants

As mentioned in Section 3.4 the procedure was slightly changed after the fourth participant. To decide whether to exclude those participants or not the PFs were assessed. For three of those four participants the fits did not match the requirements and those participants were excluded from further analysis. To maintain a balanced design those three participants were replaced and the new participants did the experiments in the same order as the participants they replaced.

Furthermore another three participants were excluded from further analysis because of an erroneous usage of the *psignifit 4 toolbox*. We used this toolbox to fit the psychometric functions. Using Bayesian methods one has to set priors, which were sometimes badly chosen. Therefore, the priors were not sensible in the case of three participants yielding to improper fits. This led to a lower SNR for 99% detection than for 50% detection. This mistake did not get noticed until the data collection was concluded and data evaluation started. Therefore there are no replacements for those 3 participants. Examples of both mentioned problems can be seen in Figure 11.

The data of the excluded participants was dismissed. Thus, the further analyses was conducted on the remaining 15 participants. This unfortunately spoiled the balance of the experimental design.

4.2 Threshold estimation task

4.2.1 Visual and auditory

The visual and auditory threshold estimation task was used to fit psychometric functions and to get the individual thresholds of the participant at 50%, 75% and 99% detection rate. The PFs were computed automatically after the experiment was done. The fits were evaluated by the experiment supervisor. This evaluation was no extensive analysis but a short look at the fits to check if they were usable and whether the thresholds lay approximately in the expected range.

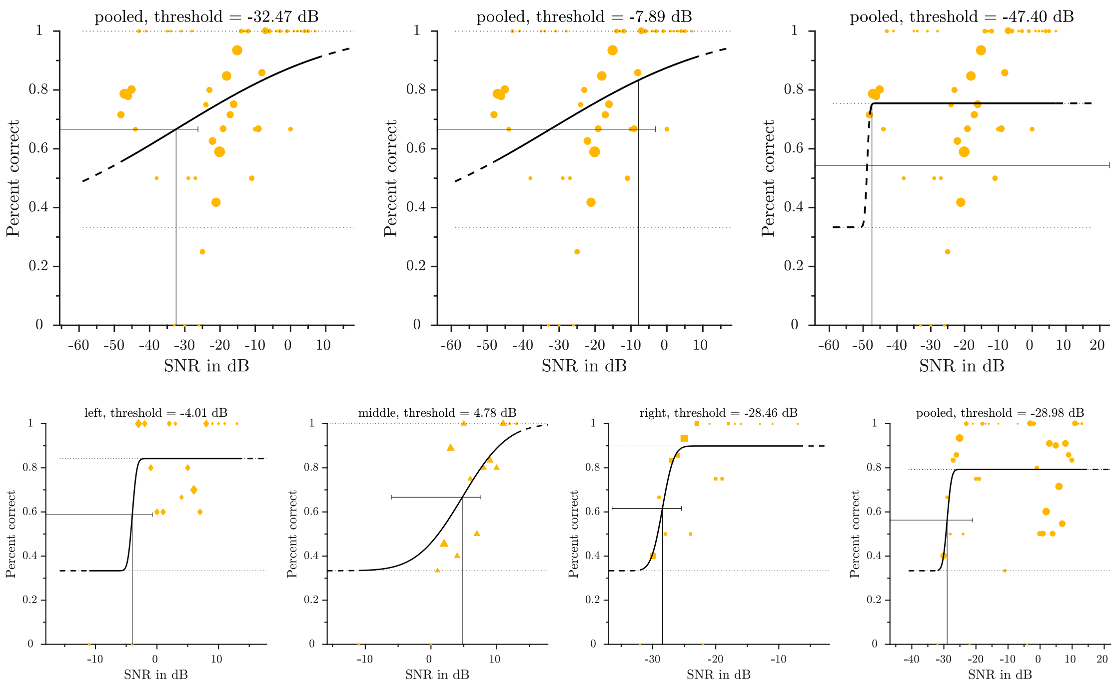

Figure 11: Examples for bad fitted psychometric functions. **First row**: From left to right: PF for 50%, 75% and 99% threshold. The prior was not sensible leading to an lower SNR for stimuli at 99% detection rate than for 50% detection rate. **Second row**: Large differences between *left*, *middle* and *right* detection thresholds and PFs, leading to a bad fit for the pooled data of all three stimuli.

All PFs can be seen in Appendix A. The visual PFs can be seen in Figure 30-Figure 36 and the auditory PFs can be seen in Figure 37-Figure 43. The values of every SNR value of all participants (except the excluded ones) are shown in Table 1.

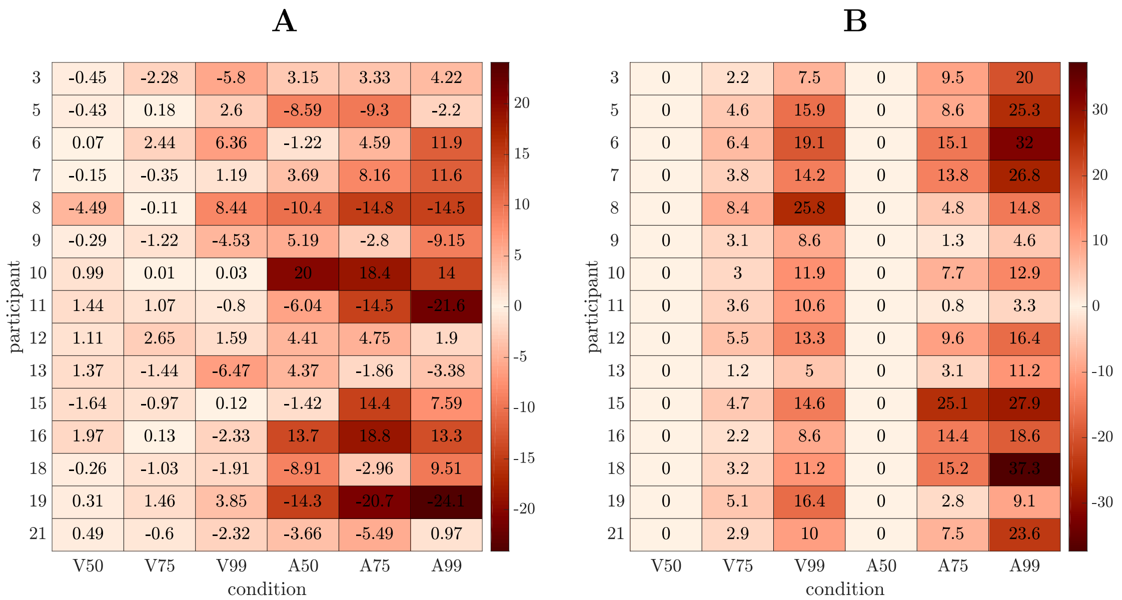

Figure 12: Heat maps of SNRs. Each row represents all SNRs of one participant. Columns present all SNRs of a modality and detection rate. **A**: Columns are centred by their mean. **B**: Each value is the difference from the measured value and the value of the SNR for the 50% detection rate of the modality.

The variations of signal to noise ratios can be seen in two different heat maps. In Figure 12 A the SNR of each participant is shown for both modalities at each detection rate. Each row is centralized around its mean to show the differences for each modality and detection rate. In Figure 12 B the SNR of each participant is shown for each modality and detection rate again. But this time the SNR of the 50% detectable stimuli is subtracted from the SNR of each detection rate, separately done for both modalities. The heat map shows us the size of the differences of the SNR within one participant and modality. Large differences correspondent to a flat raising PF. Furthermore all SNR can be seen in Table 1.

4.2.2 Bimodal

In the bimodal detection task the percent correct of the simultaneous presentation of the visual and auditory stimuli was measured for all three detection levels. We compared these values with the limit for independently processed stimuli provided by the PSM. As noted in Section 2 the PSM is only used on percent of correct responses and not on hits and false alarms as it should be. The performance of bimodal stimuli of all participants can be seen for stimuli with 50% and 75% detection rate for unimodal stimuli in Figure 13. The green diamonds mark the percent correct of each participant. The dashed black line marks the percent correct for unimodal stimuli (66% correct and 83% correct). The red line marks the mean over all participants and the solid black line marks the performance given by the PSM.

For stimuli with a unimodal detection rate of 99% the multisensory benefit is almost not existent due to ceiling performance of detection of unimodal stimuli. Therefore, the plot is rather uninformative and can be seen in the Appendix C in Figure 47.

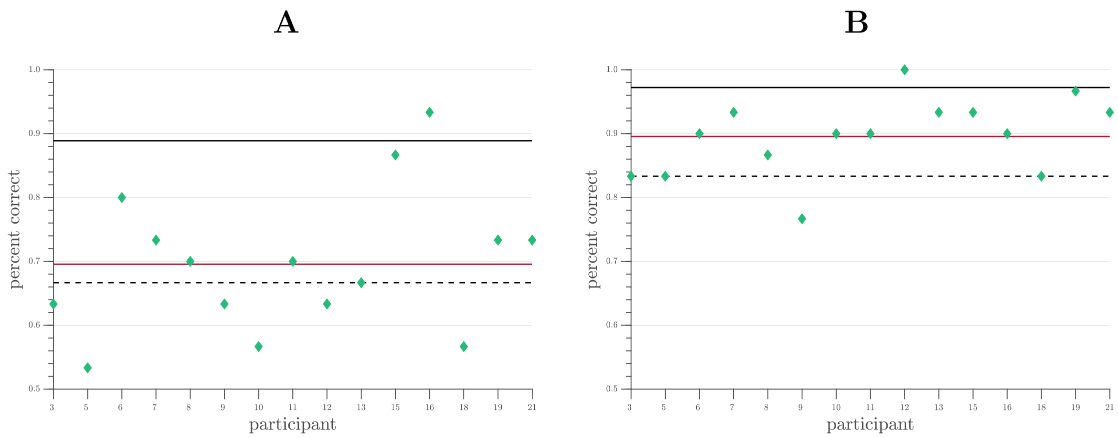

Figure 13: A: Performance of bimodal detection task for each participant for unimodal detection rates of 50% (66% correct). Each diamond shows the performance of a participant. The dashed black line marks the percent correct for unimodal stimuli, the solid black line the performance computed with the PSM and the red line marks the mean over all participants. **B**: Same plot for unimodal detection rate of 75% (83% correct).

participant	detection rate	visual				auditory				
		left	middle	right	pooled	left	middle	right	pooled	2. est.
	50	-45.70	-44.60	-44.64	-45.51	-6.11	-1.84	-11.72	-7.32	
3	75	-43.18	-41.09	-41.61	-43.34	-4.31	6.10	-7.40	2.16	
	99	-37.49	-32.54	-38.52	-38.01	0.17	12.99	3.20	12.67	
	50	-42.11	-46.59	-46.38	-45.49	-24.70	-50.94	-19.10	-50.88	-19.06
5	75	-41.07	-44.50	-41.26	-40.88	-23.36	-50.53	-14.03	-50.60	-10.48
	99	-38.55	-39.39	-28.72	-29.61	-20.08	-49.64	-4.70	-49.66	6.25
	50	-42.64	-49.63	-41.17	-44.94	-19.54	-2.66	-14.48	-11.94	
6	75	-39.94	-44.12	-38.26	-38.41	-14.69	2.15	-14.00	4.10	
	99	-34.16	-30.59	-31.24	-25.12	-2.70	13.37	-12.84	18.53	
	50	-43.63	-45.99	-45.60	-45.41	-17.89	14.86	-27.86	-30.54	-6.15
7	75	-41.91	-44.64	-39.43	-41.43	-13.64	28.17	-21.91	-30.15	7.30
	99	-37.68	-41.34	-26.62	-31.68	-3.25	30.63	-7.29	-29.21	18.24
	50	-46.45	-55.18	-44.19	-49.24	-28.71	-44.93	-12.05	-48.43	-20.78
8	75	-45.23	-48.63	-42.95	-40.97	-28.33	-29.65	-11.13	-36.93	-16.47
	99	-42.23	-34.07	-39.89	-24.10	-27.37	5.77	-8.86	4.74	-6.04
	50	-46.39	-44.43	-43.87	-45.42	-16.77	1.86	-8.97	-18.98	-4.93
9	75	-42.51	-41.35	-42.99	-42.62	-14.42	8.21	-2.56	-18.07	-3.71
	99	-32.94	-35.85	-40.81	-35.73	-8.67	16.43	8.37	-15.81	-0.72
	50	-42.70	-47.34	-42.00	-44.07	10.15	11.52	8.44	9.87	
10	75	-40.78	-45.10	-40.59	-40.73	16.60	18.68	12.37	16.59	
	99	-36.10	-39.58	-37.33	-32.60	24.75	19.07	19.31	19.61	
	50	-43.04	-45.22	-43.04	-43.66	-25.65	-41.23	-4.51	-42.06	-16.53
11	75	-41.14	-41.25	-39.40	-40.25	-18.06	-40.83	3.48	-28.82	-16.14
	99	-36.50	-35.03	-30.50	-31.89	0.77	-39.85	11.77	3.72	-15.18
	50	-39.29	-47.99	-43.19	-44.02	-1.01	25.19	4.74	-8.39	-6.44
12	75	-38.00	-46.02	-40.99	-38.26	4.86	31.80	10.77	-8.01	2.75
	99	-34.89	-41.25	-35.56	-31.02	-7.01	26.40	20.02	-7.05	11.36
	50	-43.08	-43.19	-43.31	-43.62	10.66	-0.66	-24.47	-27.21	-6.01
13	75	-41.96	-40.36	-41.58	-42.27	16.76	-0.01	-16.20	-25.58	-2.74
	99	-39.21	-35.59	-37.34	-38.96	23.96	2.45	4.06	-21.61	5.17
	50	-44.10	-50.72	-44.41	-46.79	-17.71	5.28	-25.50	-11.60	
15	75	-42.38	-48.47	-42.25	-42.18	-16.71	10.55	-16.77	12.01	
	99	-38.17	-42.97	-37.00	-32.42	-14.24	20.31	-6.25	15.96	
	50	-42.47	-43.57	-43.32	-43.14	0.59	-1.25	8.78	2.96	
16	75	-40.31	-41.20	-40.36	-40.60	15.61	9.56	12.97	18.37	
	99	-35.03	-36.56	-33.14	-34.38	18.54	-11.03	20.74	22.73	
	50	-45.22	-45.82	-44.46	-45.15	-43.27	-10.08	-2.68	-48.46	-19.62
18	75	-43.86	-42.75	-41.27	-41.97	-32.54	-9.68	5.85	-48.05	-3.32
	99	-40.53	-36.68	-33.44	-34.17	-9.83	-8.69	10.43	-47.05	17.87
	50	-43.59	-48.98	-42.10	-44.87	-30.00	-10.34	-23.14	-30.76	-24.46
19	75	-42.04	-42.15	-40.09	-39.99	-27.69	-3.67	-16.69	-27.75	-21.51
	99	-38.24	-26.97	-35.11	-28.02	-22.05	5.40	-1.12	-20.38	-14.29
	50	-45.17	-43.61	-44.97	-44.56	-11.87	-17.12	-11.11	-13.62	
21	75	-42.32	-41.38	-42.01	-41.76	-6.40	-10.47	-7.38	-7.28	
	99	-35.34	-36.76	-36.26	-34.94	5.73	5.82	1.82	8.08	

Table 1: Table of all SNRs for all detection rates (50%, 75%, 99%) for both modalities and each participant (except excluded ones).

4.3 Working memory task

4.3.1 Sensitivity index (d')

The sensitivity index d' was computed using the hit and false alarm rates that were computed for each block. In each block the first two trials were dismissed. Trials in which no answer was registered, i.e., the participant did not give an answer in time or not at all, were dismissed. The number of go trials and no-go trials did get adjusted accordingly. Hit and false alarm rates of 0 or 1 were corrected according to the adjustments of Stanislaw and Todorov (1999).

Fur further analysis which do not account for blocks as a factor the sensitivity index is computed after the hit and false alarm rates were computed over all three blocks for each condition of a participant. Thus, there are not as many corrections for hit and false alarm rates of 0 and 1 needed to be computed.

In the following, data was tested to be approximately normal distributed with the help of Q-Q Plots and the Shapiro-Wilk test before conducting ANOVAS. Sphericity was tested with Mauchly's sphericity test and in case of sphericity violations, the degrees of freedom got adjusted according to the Greenhouse-Geisser correction. All post hoc tests are Bonferroni corrected. For all statistical test the significance level was set to $\alpha = .05$.

4.4 Main Analyses

4.4.1 The effect of blocks, modality and detection rate on d' and RT

A full model repeated measure ANOVA was conducted with the factors of *modality*,*detection rate* and *block* on d'. The data can be seen as line plots in Figure 14. The Results of the ANOVA show no significant three way interaction $[F(8,\ 112) = .95,\ p = .48,\ \eta_p^2 = .06]$. The two way interaction of *modality* $\times$ *detection rate* reached significance with $[F(2.69,\ 37.64) = 14.42,\ p < .001,\ \eta_p^2 = .51]$, the two way interaction of *modality* $\times$ *block* reached significance with $[F(4,\ 56) = 4.77,\ p < .01,\ \eta_p^2 = .26]$. The two way interaction of *detection rate* $\times$ *block* $[F(2.77,\ 38.79) = 2.87,\ p = .52,\ \eta_p^2 = .17]$ did not reach significance.

All three main effect were significant, *modality* with $[F(1.15,\ 16.12) = 23.40,\ p < .001,\ \eta_p^2 = .63]$, *detection rate* with $[F(2,\ 28) = 117.64,\ p < .001,\ \eta_p^2 = .89]$ and *block* with $[F(2,\ 18) = 16.52,\ p < .001,\ \eta_p^2 = .54]$.

Pairwise comparisons of the factor *block* show a difference between block 1 and block 2 ($p < .01$) and a difference between block 1 and block 3 ($p < .01$). No difference between block 2 and block 3 was existent ($p = 0.99$).

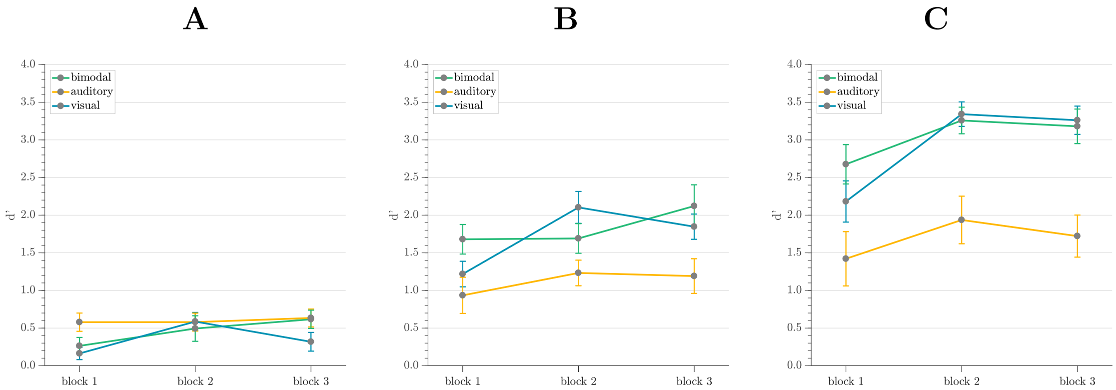

Figure 14: Line plots showing d' (mean ± SEM) of each modality and for each block. (**A**): 50% detection rate(**B**): 75% detection rate. **C**: 99% detection rate.

A full model repeated measure ANOVA was conducted with the factors of *modality*,*detection rate* and *block* on RT. The data can be seen as line plots in Figure 15. The Results of the ANOVA show no significant three way interaction $[F(8,\ 112) = 1.77,\ p = .91,\ \eta_p^2 = .11]$. The two way interaction of *modality* $\times$ *detection rate* reached significance with $[F(4,\ 56) = 2.87,\ p < .05,\ \eta_p^2 = .17]$. The two way interactions of *modality* $\times$ *block* $[F(1.70,\ 23.74) = .05,\ p = .93,\ \eta_p^2 = .00]$ and the two way interaction of *detection rate* $\times$ *block* $[F(1.84,\ 25.81) = 1.10,\ p = .34,\ \eta_p^2 = .07]$ did not reach significance. The main effect of *modality* with $[F(2,\ 28) = 11.76,\ p < .001,\ \eta_p^2 = .46]$ was significant, as well as the main effect of *detection rate* with $[F(2,\ 28) = 6.28,\ p < .01,\ \eta_p^2 = .31]$. The main effect of *block* did not reach significance $[F(1.06,\ 14.83) = 3.99,\ p = .06,\ \eta_p^2 = .22]$.

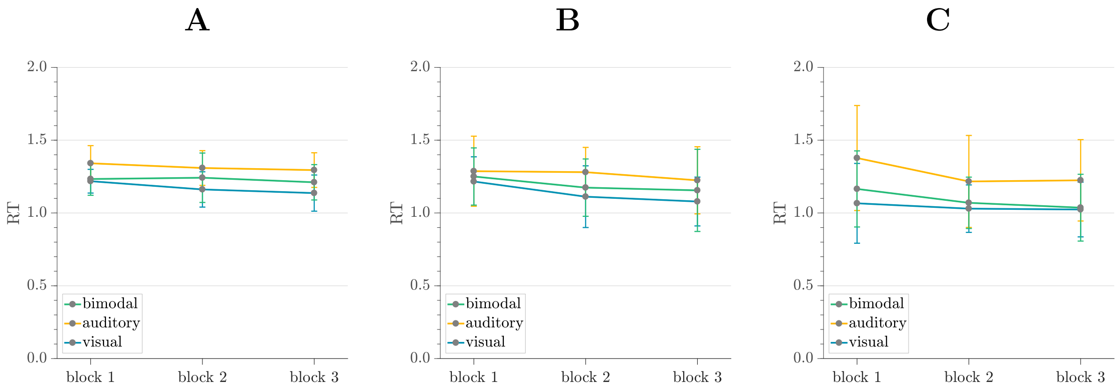

Figure 15: Line plots showing RTs (mean ± SEM) of each modality and for each block. (**A**): 50% detection rate(**B**): 75% detection rate. **C**: 99% detection rate.

4.4.2 Blocks taken together - the effect of modality and detection rate on d' and RT

The effect of modality (visual, auditory, bimodal) and detection rate (50%, 75% and 99%) on the sensitivity index d' were tested in a full model repeated measure ANOVA.

The ANOVA with showed a significant interaction effect of *modality* × *detection rate* on d' $[F(2.75,\ 38.53) = 13.88),\ p < .001,\ \eta_p^2 = .50]$. Furthermore the main effect of *modality* reached significance $[F(1.27,\ 17.74) = 22.22,\ p < .001,\ \eta_p^2 = .61]$ as did the main effect of *detection rate* $[F(2,\ 28) = 82.64,\ p < .001,\ \eta_p^2 = .86]$.

To get a better understanding which modalities are different at each detection rate three separate repeated measure ANOVAs were carried out, one at each detection rate, with *modality* as the only between subject factor. The results of these 3 ANOVAs can be seen in Figure 17.

The ANOVA for stimuli at 50% detection rate showed a significant effect of *modality* $[F(2,\ 28) = 5.28,\ p < .05,\ \eta_p^2 = .27]$. Bonferroni-corrected post hoc tests revealed that the conditions across subjects (mean ± standard error of the mean (SEM)). The auditory condition ($d' = 0.58 \pm 0.09$) was significantly better than the the visual ($d' = 0.35 \pm 0.06$) condition. There was no significant difference between the ($d' = 0.44 \pm 0.08$) condition and either of the unimodal (visual and auditory) conditions.

The ANOVA for stimuli at 75% detection rate showed a significant effect of

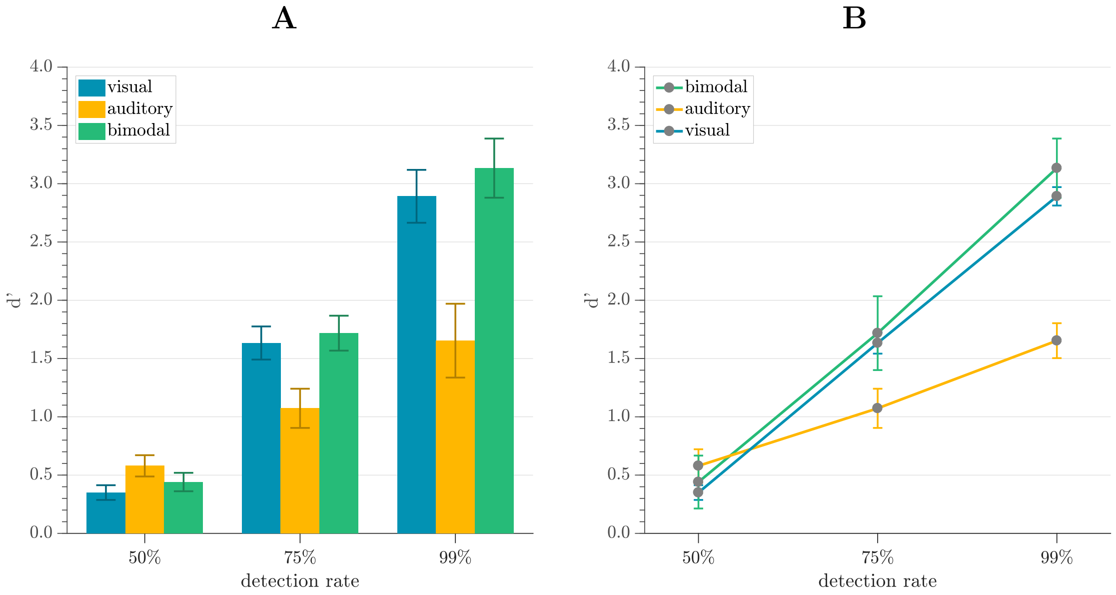

Figure 16: **A**: Bar plot showing mean d' of each condition. Error bars are shown are standard error of the mean (SEM). **B**: Line plot showing the mean d' value of each condition. Error bars represent the SEM.

modality $[F(1.45,\ 20.31) = 7.77,\ p < .01,\ \eta_p^2 = .36]$. There was a significant difference between the auditory condition ($d' = 1.07 \pm 0.17$) being worse than the bimodal condition ($d' = 1.72 \pm 0.15$). The visual condition ($d' = 1.63 \pm 0.14$) did not differ significantly from either the auditory nor the bimodal condition.

The ANOVA for stimuli at 99% detection rate reached a significant effect of *modality* $[F(2,\ 28) = 23.36,\ p < .001,\ \eta_p^2 = .63]$. The auditory condition ($d' = 1.65 \pm 0.32$) was significantly worse than visual ($d' = 2.89 \pm 0.23$) and the bimodal ($d' = 3.13 \pm 0.25$) condition. Between the visual and bimodal conditions no significant difference could be shown.

The effect of *modality* and *detection rate* on reaction time (RT), given in seconds, was tested with a full model repeated measure ANOVA.

The ANOVA showed a significant interaction of *modality* and *detection rate* on RT $[F(4,\ 56) = 2.83),\ p < .05,\ \eta_p^2 = .17]$. Furthermore the main effect of *modality* reached significance $[F(2,\ 28) = 11.74,\ p < .001,\ \eta_p^2 = .46]$ as did the main effect of *detection rate* $[F(2,\ 28) = 6.44,\ p < .01,\ \eta_p^2 = .32]$.

Again three separate ANOVAs were conducted for each detection rate to get a

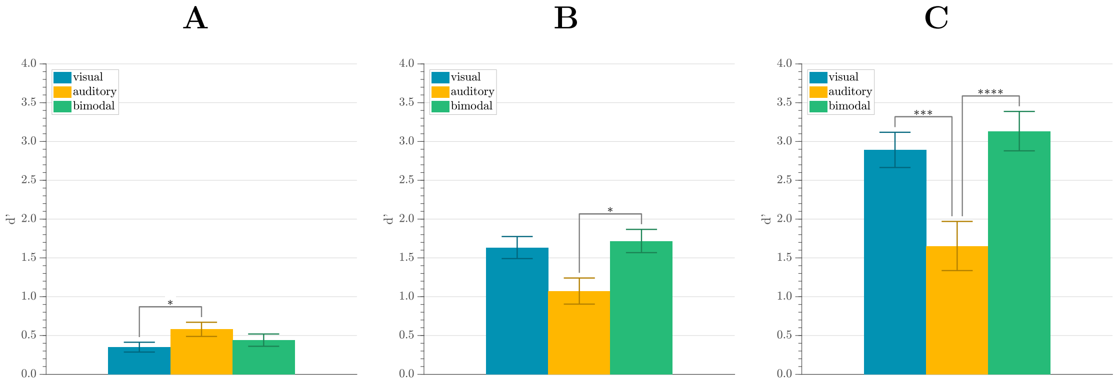

Figure 17: Bar plots showing d' (mean ± SEM) of each modality at 50% detection rate (**A**), 75% detection rate (**B**) and 99% detection rate (**C**). Also the significances for the ANOVAs are shown in the plot

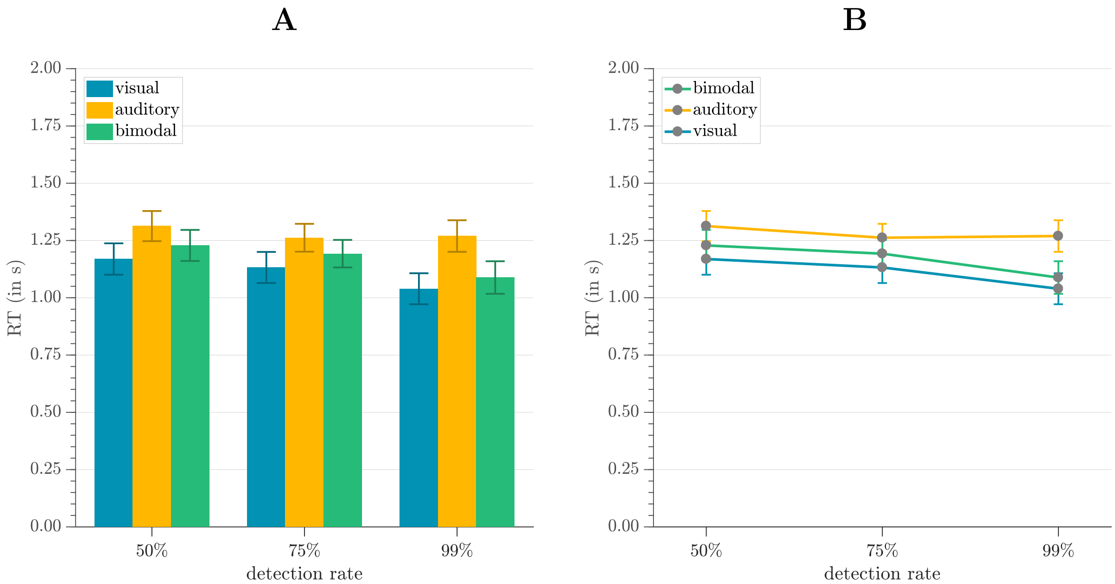

Figure 18: **A**: Bar and line plots showing the mean RT of each condition. The error bars shown are standard error of the mean. **B**: Line plot showing the mean RT value of each condition. Error bars represent the SEM.

better understanding of the data. The results of these ANOVAs can bee seen in Figure 19.

The ANOVA at 50% detection rate showed a significant effect of *modality* $[F(2,\ 28) = 5.57,\ p < .01,\ \eta_p^2 = .28]$. The RT of the auditory modality $RT = 1.31 \pm 0.07$ was significant slower than visual $(RT = 1.17 \pm 0.07)$. The bimodal condition with intermediate RT $(RT = 1.23 \pm 0.07)$ does not differ from either of the two others.

The ANOVA at 75% detection rate showed a significant effect of *modality* $[F(1.30,\ 18.21) = 7.14,\ p < .05,\ \eta_p^2 = .34]$. The RT of the auditory condition $(RT = 1.26\ s \pm 0.06\ s)$ being significantly slower than the RT of the visual condition $(RT = 1.13\ s \pm 0.07\ s)$ and significantly slower than the RT of the bimodal condition $(RT = 1.19\ s \pm 0.06\ s)$. There is no significant difference between the visual and bimodal condition.

The ANOVA at 99% detection rate showed a significant effect of *modality* $[F(2,\ 28) = 13.10,\ p < .001,\ \eta_p^2 = .48]$. The RT of the auditory condition $(RT = 1.27\ s \pm 0.07\ s)$ again being significantly slower than the RT of the visual condition $(RT = 1.04\ s \pm 0.07\ s)$ and significantly slower than the RT of the bimodal condition $(RT = 1.09\ s \pm 0.07\ s)$. There is no significant difference between the visual and bimodal condition.

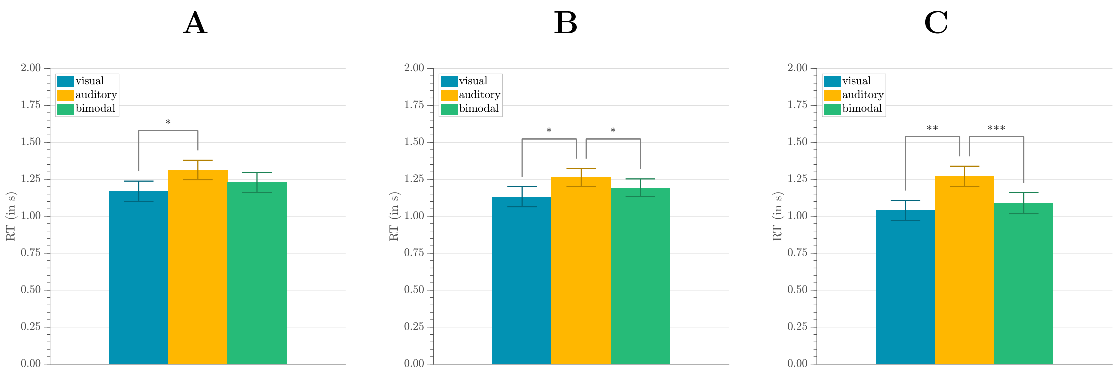

Figure 19: Bar plots showing RT (mean ± SEM) of each modality at 50% detection rate (**A**), 75% detection rate (**B**) and 99% detection rate **C**). Also the significances for the ANOVAs are shown in the plot.

4.5 Exploratory data analyses

4.5.1 Between factor: Strategies

In the questionnaire approximately the half of the participants (7 of 15) stated, that they used a (external) strategy (see Section 4.6). Therefore, a new between subject factor *strategy* was added and a mixed model (3×3×2) ANOVA was performed with the two within factors *modality* and *detection rate* as well as the newly created between factor *strategy* and all interactions. The three way interaction of *modality* × *detection rate* × *strategy* did not reach significance [$F(4,\ 52) = 1.29,\ p = .29,\ \eta_p^2 = .09$]. The interaction of *modality* × *detection rate* [$F(4,\ 52) = 13.96,\ p < .001,\ \eta_p^2 = .52$] reached significance. The two way interactions of *modality* × *strategy* [$F(2,\ 26) = 1.03,\ p = .37,\ \eta_p^2 = .07$] and *detection rate* × *strategy* [$F(2,\ 26) = .40,\ p = .68,\ \eta_p^2 = .03$] did not reach significant. The main effects of *modality* [$F(1.29,\ 16.78) = 22.65,\ p < .001,\ \eta_p^2 = .64$] and *detection rate* [$F(2,\ 26) = 78.42,\ p < .001,\ \eta_p^2 = .86$] were significant. The main effect of the between factor *strategy* [$F(1,\ 13) = .02,\ p = .91,\ \eta_p^2 = .00$] did not reach significance.

4.5.2 Between factor: Multisensory benefit

Another way to split up the participants into two groups is via their multisensory performance in the bimodal threshold estimation task. Participants with a higher multisensory integration in the threshold estimation task may benefit more from the bimodal stimuli presentation in the working memory experiment.

To assess a potential difference participants were split in two groups. The split was done via mean split of the performance of bimodal presented stimuli for 75% detection rate (see Figure 13 B).

In Figure 21 the mean sensitivity index of the two groups is shown. A mixed model ANOVA revealed no significant three way interaction of *modality* × *detection rate* × *multisensory benefit* [$F(2.70,\ 35.05) = 0.48,\ p = .68,\ \eta_p^2 = .04$]. The two way interaction between *modality* × *detection rate* reached significance [$F(2.70,\ 35.05) = 12.63,\ p < .001,\ \eta_p^2 = .49$]. The interaction of *modality* × *multisensory benefit* [$F(1.27,\ 16.50) = .10,\ p = .81,\ \eta_p^2 = .01$] and *detection rate* ×

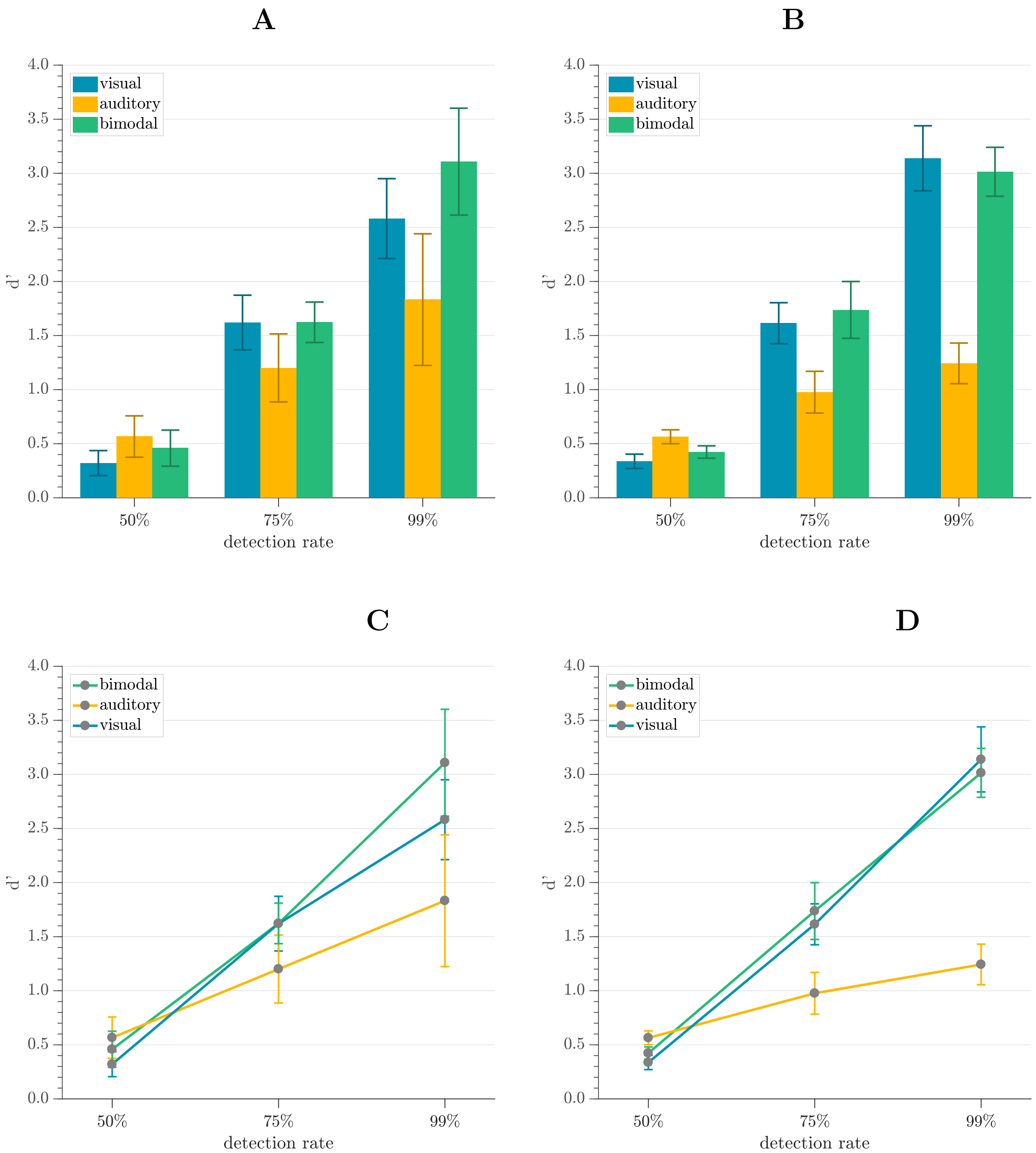

Figure 20: A: Bar plot showing the mean d' of each condition for participants that did not state using a strategy. Error bars show the SEM. **B**: Bar plot showing the mean d' of each condition for participants that used an external strategy. Error bars show the SEM. **C**: Same data as in A but as line plot. **D**: Same data as in B but as line plot.

multisensory benefit $[F(1.43,\ 18.57) = .60,\ p = .50,\ \eta_p^2 = .04]$ did not reach significance. The main effects of *modality* $[F(1.27,\ 16.50) = 20.40,\ p < .001,\ \eta_p^2 = .61]$ and *detection rate* $[F(1.43,\ 18.57) = 76.76,\ p < .001,\ \eta_p^2 = .86]$ reached significance. The main effect of the between factor *multisensory benefit* $[F(1,\ 13) = .34,\ p = .57,\ \eta_p^2 = .03]$ did not reach significance.

4.5.3 Between factor: Trials with no answers

Participants were instructed to answer in each trial of the 2-back task, even if they had to guess. But most participants did not answer at least in some trials. Figure 22 A shows the total number of all trials in which participants did not give an answer separate for each condition. Figure 22 B shows a heat map of the same data. In this heat map each row correspondents to one participant and each column corresponds to one condition.

A mixed model ANOVA was conducted with the two within factors *modality* and *detection rate* as well as the between factor *no answers*. This between factor was introduced as a mean split of participants that had more or less trials with no answer registered. This lead to two groups of 6 participants with a higher number than the mean and 9 participants with a lower number than the mean.

The interaction of *modality* × *detection rate* × *no answers* did not reach significance $[F(4,\ 52) = .88,\ p = .49,\ \eta_p^2 = .06]$ as did the interaction of *modality* × *no answers* $[F(1.29,\ 16.73) = .88,\ p = .39,\ \eta_p^2 = .06]$ and *detection rate* × *no answers* $[F(1.35,\ 17.49) = 1.20,\ p = .31,\ \eta_p^2 = .09]$. The interaction of *modality* × *detection rate* reached significance $[F(4,\ 52) = 12.68,\ p < .001,\ \eta_p^2 = .49]$. The main effect of *modality* $[F(1.29,\ 16.73) = 22.50,\ p < .001,\ \eta_p^2 = .63]$ and *detection rate* $[F(1.35,\ 17.49) = 80.25,\ p < .001,\ \eta_p^2 = .86]$ did reach significance. The main effect of *no answers* did not reach significance $[F(1,\ 13) = .17,\ p = .67,\ \eta_p^2 = .01]$.

4.5.4 Correlations: d'-d'

Correlations of the performance of one modality with the performance of another modality in the working memory task, at each detection rate, can be seen in Figure 23.

In each chart, the points show the correlations of the 15 participants between

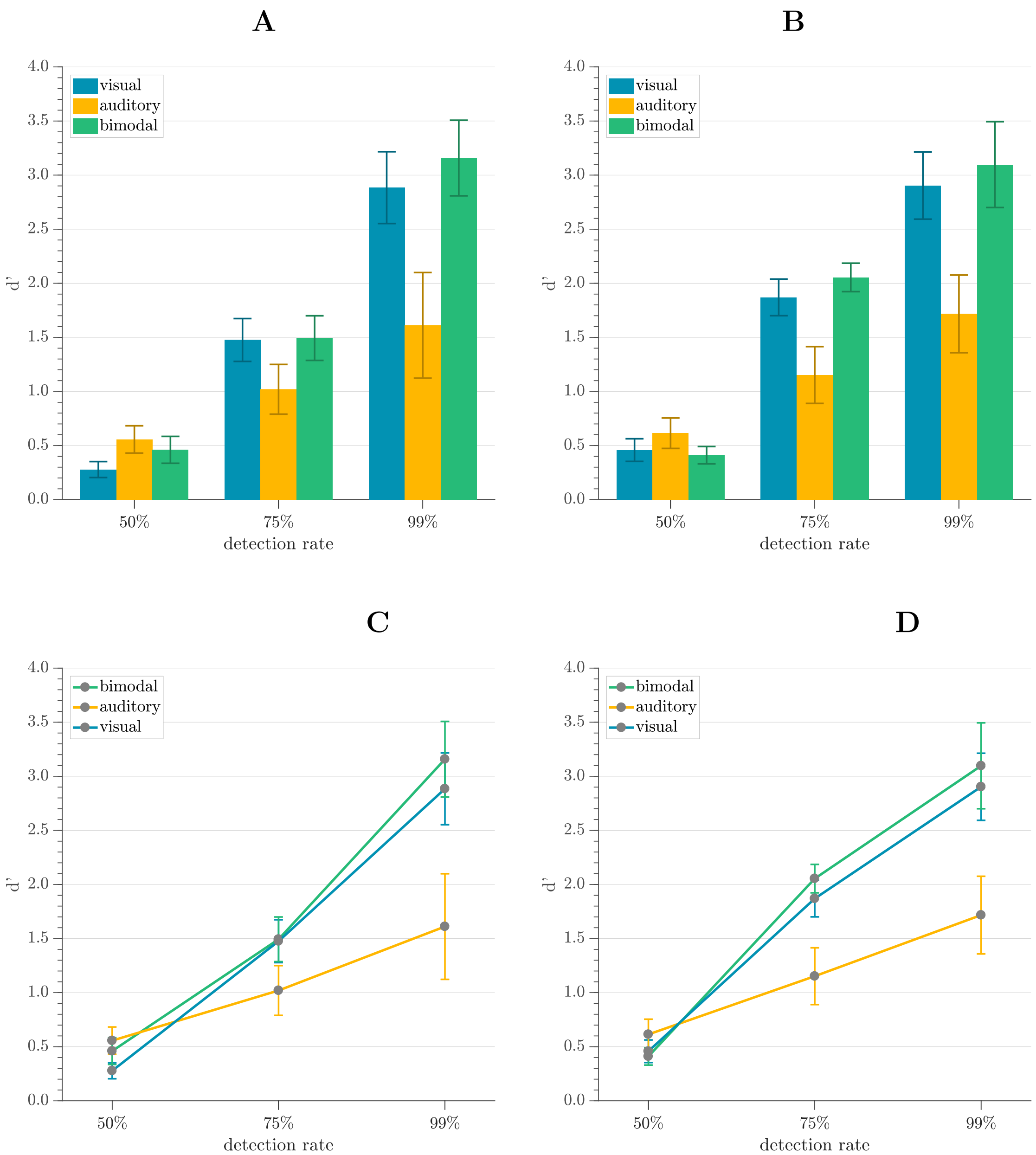

Figure 21: A: Bar plot showing the mean d' of each condition for participants with low multisensory benefit at 75% detection rate. Error bars show the SEM. **B**: Bar plot showing the mean d' of each condition for participants with high multisensory benefit at 75% detection rate. Error bars show the SEM. **C**: Same data as in A but as line plot. **D**: Same data as in B but as line plot.

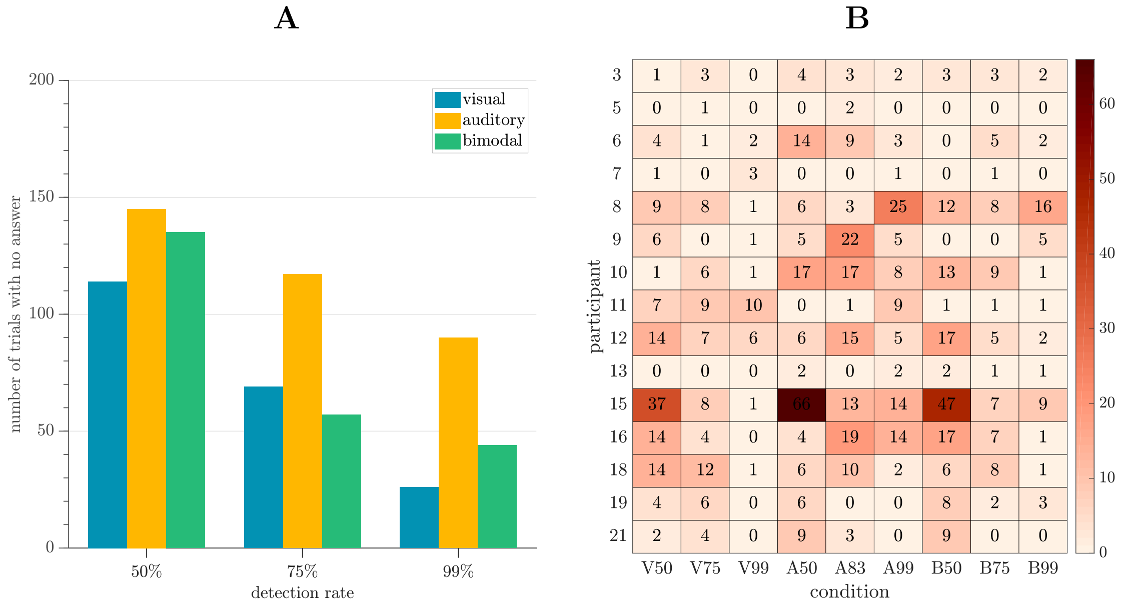

Figure 22: **A**: Bar plot showing the total number of trials with no registered answer over all participants for each condition. **B**: Heat map showing the distribution of no answers for each participant (rows) and detection rate (columns).

two conditions of the working memory experiment. The dashed red line is the line of the Theil-Sen robust linear regression. Spearman's rank correlation coefficient (ρ) and its significance (p) are shown within the charts.

4.5.5 Correlations: d'-RT

Correlations of the performance with the RT in the working memory task for each modalities at each detection rate can be seen in Figure 24.

In each chart, the points show the correlations of the 15 participants between two conditions of the working memory experiment. The dashed red line is the line of the Theil-Sen robust linear regression. Spearman's rank correlation coefficient (ρ) and its significance (p) are shown within the charts.

4.5.6 Correlations: d' computed - d' measured

Correlations of the computed performance with the measured performance in the working memory task, for each modality have been computed. For stimuli with

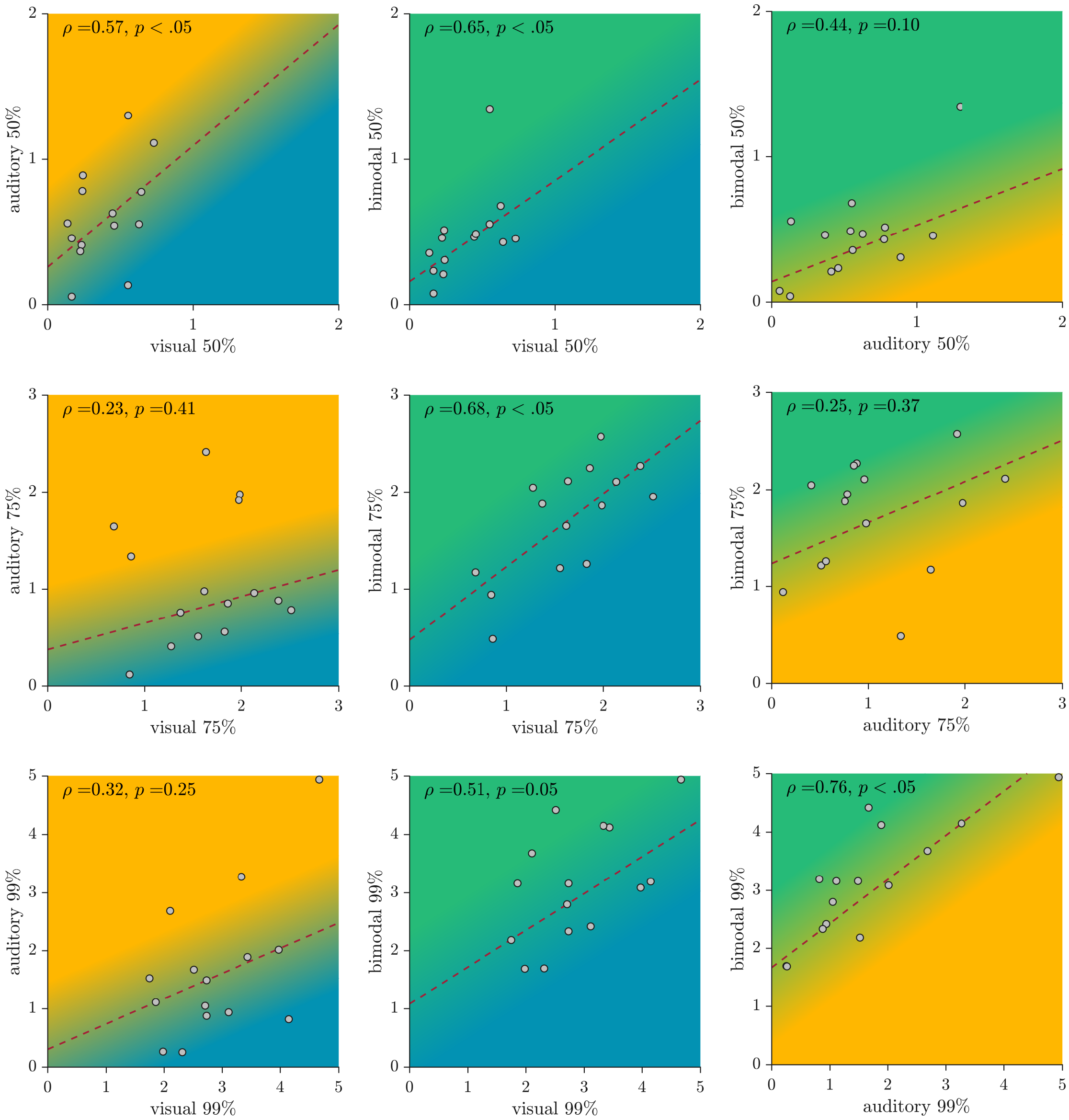

Figure 23: Correlation of individual performances between different modalities. Each point represents one participant. The red dashed line represents the Theil-Sen robust linear regression. Spearman's rank correlation coefficient (ρ) and its significance (p) are shown within each diagram. Each row contains correlations of the same detection rate and each column contains correlations of the same conditions. In each

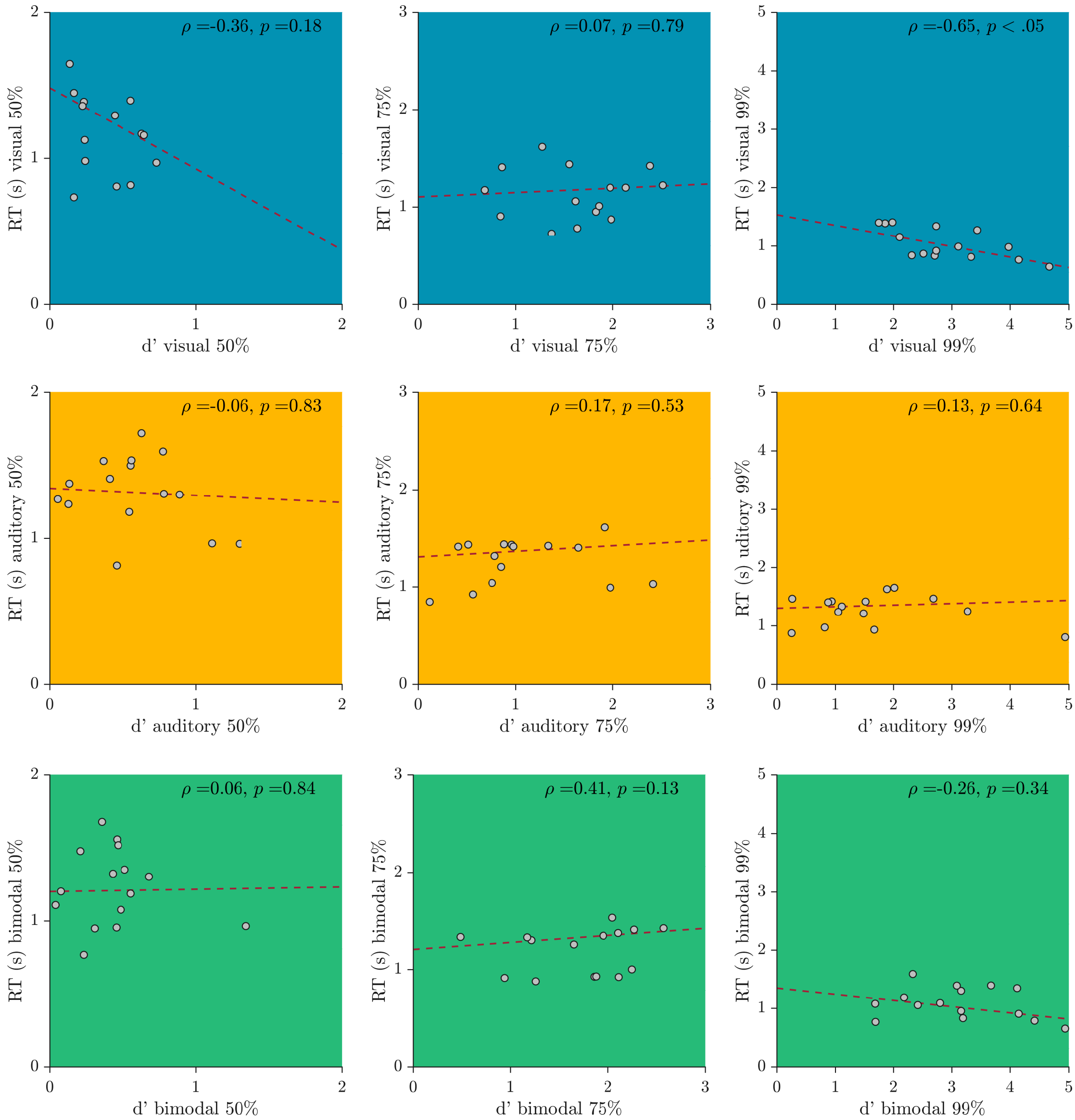

Figure 24: Correlation between individual performances of d' and the RT of the same modality. Each point represents one participant. The red dashed line represents the of the Theil-Sen robust linear regression. Spearman's rank correlation coefficient (ρ) and its significance (p) are shown within each diagram. Each row contains correlations of the same detection rate and each column contains correlations of the same conditions.

unimodal detection rate of 50% the correlations can be seen in Figure 25. For stimuli with a unimodal detection rate of 75% the correlations can be seen in Figure 26. The computed d' values have been calculated according to the approach mentioned in Section 2.

In each chart, the points show the correlations of the 15 participants between two conditions of the working memory experiment. The dashed red line is the line of the Theil-Sen robust linear regression. The dotted black line marks the angle bisector. Spearman's rank correlation coefficient (ρ) and its significance (p) are shown within the charts.

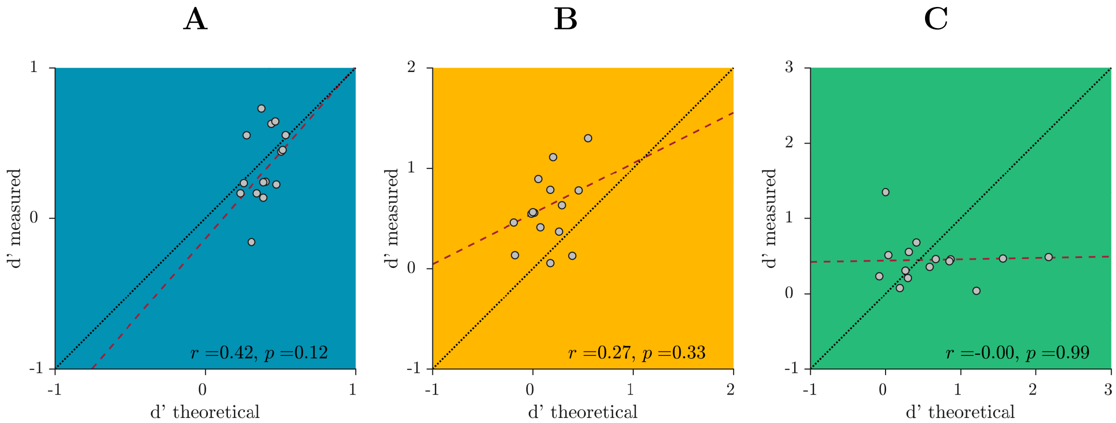

Figure 25: Correlation between individual performances of the theoretical d' and the measured d' measured at 50% detection rate. **A** visual **B** auditory and **C** bimodal. Each points represents a participant. The dotted black line is the angle bisector and the dashed red line represents the Theil-Sen robust linear regression. Spearman's rank correlation coefficient (ρ) and its significance (p) are shown within each diagram.

4.6 Questionnaire evaluation

The evaluation of the questionnaire participants had to answer at the end of the second session can be seen in the following.

The complete German questionnaire as used in the experiment and an English version can be seen in Appendix A in Figure 50. The questionnaire consisted out of 4 questions. Participants were told to answer the questions in the given order

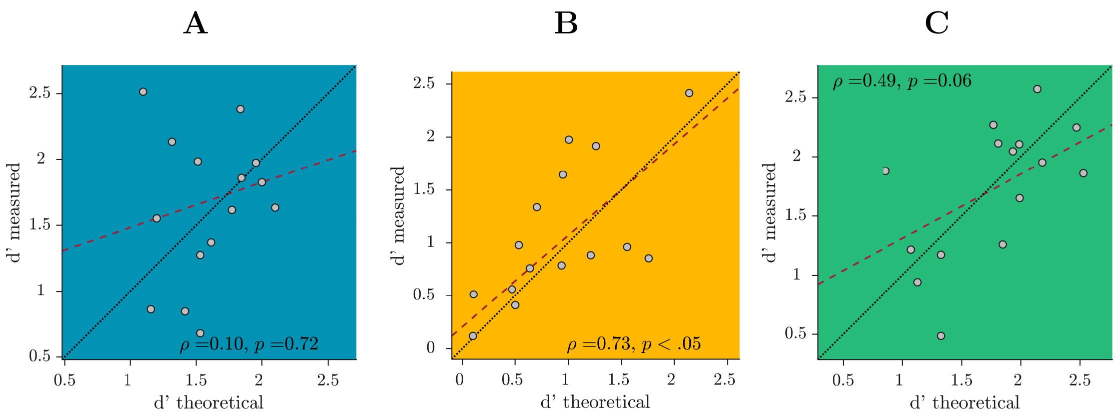

Figure 26: Correlation between individual performances of the theoretical d' and the measured d' at 75% detection rate. **A** visual **B** auditory and **C** bimodal. Each points represents a participant. The dotted black line is the angle bisector and the dashed red line represents the Theil-Sen robust linear regression. Spearman's rank correlation coefficient (ρ) and its significance (p) are shown within each diagram.

and not to jump back to an earlier question.

The first question was *"How would you rate the difficulty of the working memory task today?"* To answer participants had to make a mark on a scale ranging from *easy* to *hard*. A histogram of the answers can be seen in Figure 27 with the x-axis representing the answer scale as was present in the questionnaire. The black vertical line marks the mean over all participants. This mean was computed transforming the answers into numbers, ranging from 1 (*easy*) to 5 (*hard*). The marks were binned into steps of 0.5. Using this transformation the mean was calculated $\mu = 3.5$.

The second question was *"Did you use a specific strategy? If so, please describe it in a few words."* This question could be answered by checking one of two boxes, either *"No"* or *"Yes"*. Nine of the fifteen participants (60%) stated that they used a special strategy. The largest group (six participants) either repeated the last two stimuli in their mind or in a low voice to themselves. One participant reported that they used their fingers to remember the direction of the last two stimuli. One participant reported that they concentrated on the visual stimuli in the bimodal

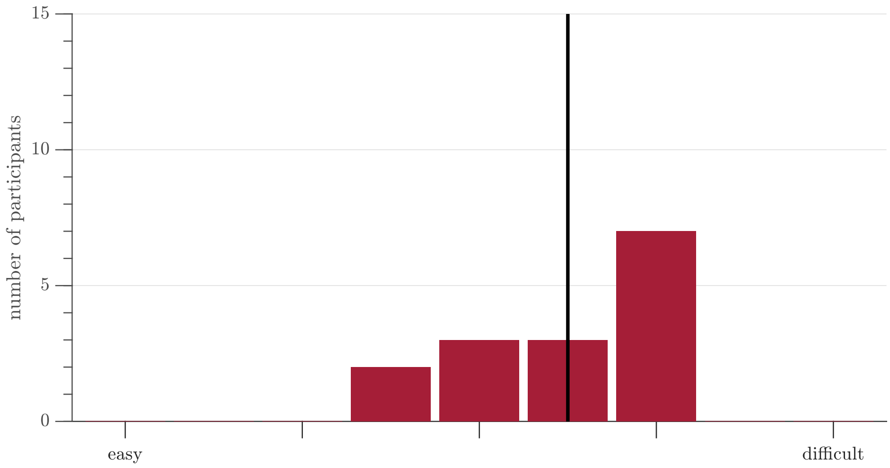

Figure 27: Histogram of answers of question 1 of the questionnaire *"How would you rate the difficulty of the working memory task today?"*. The x-axis represents the scale participants used two answer. The black vertical line shows the mean over all participants.

condition and the remaining participant reported concentrating on the auditory stimuli in the bimodal condition.

For the between factor *strategy* as used in the ANOVA above the participants that either repeated the last two stimuli in their mind or in a low voice to themselves and the participant that used their fingers were taken together into a group using an external strategy.

The third question was *"Did you think that one (or two) of the below mentioned types were easier than the others? You can pick more than one answer."* Participants could mark one (or more) of the following possibilities: *"all the same difficulty", "visual (only pictures) easier", "auditory (only tones) easier"* and *"bimodal (pictures + tones) easier"*. Thirteen of the 15 participants (87%) stated that they thought the visual condition was the easiest and 6 (40%) stated, that the bimodal was the easiest. Of those 6 participants only 2 did not check both, visual and bimodal. None of the participants checked either "auditory (only tones) easier" or "all the same difficulty". A histogram of the answers can be seen in Figure 28.

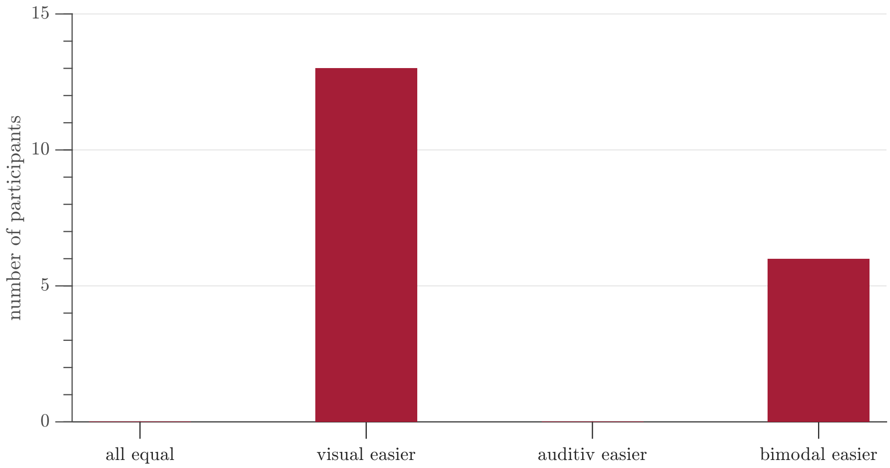

Figure 28: Histogram of answers of question 3 of the questionnaire *"Did you think that one (or two) of the below mentioned types were more easy than the others? You can pick more than one answer."* The x-axis shows the possible answers. Participants answered with marking the correspondent boxes.

The fourth question was *"In the bimodal task (pictures + tones), did you concentrate on pictures or tones?"*. A histogram of the answers can be seen in Figure 29. The x-axis represents the answer scale that was shown to the participants and the black vertical line marks the mean over all participants. The mean was calculated analogous as in question 1, resulting in $\mu = 4.1$.

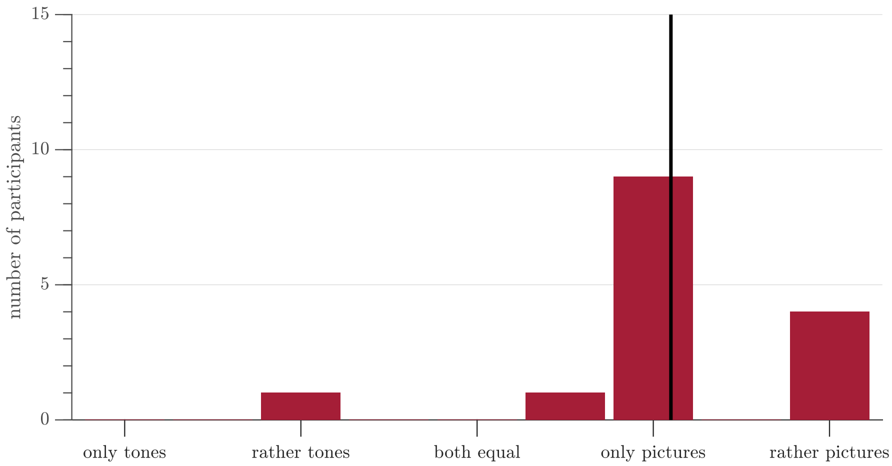

Figure 29: Histogram of answers of question 4 of the questionnaire *"In the bimodal task (pictures + tones), did you concentrate on pictures or tones?"* The x-axis represents the scale participants used two answer. The black vertical line shows the mean over all answers.

5 Discussion

5.1 Threshold estimation task

All fitted psychometric functions for every participant can be seen in Appendix A. As noted before there is one fit for each stimulus, i.e., for the Gabor patch turned left, middle and right and the c major chord sounding from either the left, middle or right. The data of the single fits than are pooled together and another psychometric function is fitted over all data points for one modality.

Some participants had drastically different thresholds for the different stimuli (left, middle, right), i.e., the different directions, in the auditory domain. For example one participant (Figure 11) had an absolute threshold of 4 dB for the left stimulus, 5 dB for the middle one and -28 dB for the right one. This huge detection difference is problematic. The fit over the pooled data of all three stimuli looks rather underwhelming and has an absolute threshold of $-29dB$, which is lower than for anyone of the thresholds of the three stimuli itself. A possible fix would be to take the three different thresholds of each stimulus for further tasks. But this would almost certainly lead to a unwanted learning effect. Participants would be able to differentiate the three auditory stimuli due to their different noise levels. Thus, one signal to noise ratio for all three directions was used.

Unfortunately, another problem occurred. In some cases the PFs had a very steep slope. This happened most likely due to bad fits of the pooled data. This leads to almost identical SNR values for the different detection rates. For example, participant 11 had a SNR of -17 dB for a detection rate of 50%. At 75% detection rate they had a SNR of -16 dB and at 99% detection rate they had a SNR of -15 dB (see Table 1 and Figure 40). Consequently the auditory stimuli has almost the same amount of noise for all three detection rates. Those problems are only seen in auditory fits of four participants (participant 2, 4, 9 and 11).

As mentioned earlier those problems were noticed only after the data collection of the forth participant and the experiment then was adjusted. The threshold estimation task was done a second time if this problem occurred. In the repetition of threshold estimation task the psi-method was updated for all stimuli together.

This leads to only one PF, usually with a larger width. This second PF leads to a better estimation of the threshold even if participants have different thresholds for the three stimuli.

As can be seen in Appendix A for 7 of the 15 participants the auditory threshold estimation task had to be repeated. These participants might not have as good of an estimation of the threshold as participants with similar threshold values for all three stimuli. Of the participants who needed to do the auditory threshold estimation task a second time, two of them (participants 9 and 11) still had a really steep PF but their data was not excluded from the further analysis, since the already low number of participants.

Furthermore, they did not show different behaviour in the working memory task compared to other participants. For example participant 10 with a steep PF and SNR of $-5.28\ dB$ at 50% detection rate, $-3.97\ dB$ at 75% detection rate and -0.69 at 99% detection rate shows a bad auditory working memory performance for every detection rate. That is somewhat expected since the SNR values are pretty similar. But participant 11 shows the same pattern of auditory working memory performance. In contrast their PF is rather flat and the SNR values differ much more ($-20.84\ dB$ at 50% detection rate, $-16.00\ dB$ at 75% detection rate and $-6.07\ dB$ at 99% detection rate), as can be seen in Table 1 and Figure 45.

The large variations of signal to noise ratios can be seen in two different heat maps. In Figure 12 A the SNR of each participant is shown for both modalities at each detection rate. Each row is centralized around its mean to show the differences for each modality and detection rate. In Figure 12 B the SNR of each participant is shown for each modality and detection rate again. But this time the SNR of the 50% detectable stimuli is subtracted from the SNR of each detection rate, separately done for both modalities. The heat map shows us the size of the differences of the SNR within one participant and modality. Large differences correspondent to a flat raising PF. Furthermore all SNR can be seen in Table 1.

5.2 Working memory task

First a possible learning effect was examined. An repeated measure $2 \times 3 \times 3$ ANOVA with the factors *modality*, *detection rate* and *block* was conducted to examine their effect on d'. Most important the factor *block* did not reach significance in the three way interaction and neither in both two way interactions. But it reached significance as a main effect. Post hoc tests reveal a significant difference between block 1 and block 2, block 1 and 3, but no difference between block 2 and 3. Thus a learning effect occurred between block 1 and block 2, or rather between the two sessions, as can be seen in Figure 14. An analogues $2 \times 3 \times 3$ ANOVA was done to examine the effect on RTs. There was no significant effect of the factor *block* neither in any interaction nor as main effect (Figure 15).

As mentioned in the introduction this learning effect is not surprising. Since there is no interaction of the learning effect with different modalities or detection rates the data of the three blocks was pooled together. This was done for the hit rates and false alarm rates to appease the necessity of using the correction for values of 0 and 1 when computing d'.

The most apparent and most important result is the fact that no multisensory benefit could be shown at either detection rate (see Figure 16). For all detection rates the visual condition does not differ significantly from the bimodal condition and thus there exists no multisensory benefit.

A full model 3×3 repeated measure ANOVA showed a significant interaction effect between the two factors *modality* and *detection rate*. This interaction is most likely due to the fact, that the working memory performance in the auditory condition does not increase as much for better detectable stimuli as it does for the visual and bimodal conditions.

Furthermore, for stimuli with 50% detection rate, the WMP in the auditory condition was significantly better than in the visual condition (see Figure 17). This changed for 75% and 99% detectable stimuli. There the WMP of the auditory condition was significantly worse than the WMP of the visual and bimodal condition.

This somewhat strange interaction effect is most likely due to a worse detection threshold estimation. As mentioned the confidence intervals of the auditory thresholds were much wider than for visual thresholds. In particular a lower SNR is not that unlikely for auditory stimuli at detection threshold. Thus the better working memory performance of the auditory condition for 50% detectable stimuli.

The correlations of the performance of the different conditions in the working memory task show that participants that performed good at a certain modality tended to perform good at the other modalities as well. This trend existed for all detection rates (see Figure 23).

The slopes of the Theil-Sen regression lines all show a positive correlation. Spearman's rho is positive for all nine correlations with values between .23 and .76. Four out of the nine correlation coefficients reached significance.

The correlations of the performance with the RT for each condition show no general tendency. For visual stimuli there is a significant correlation for stimuli with 99% detection rate. Participants that had better performance tended to have faster RTs in this condition (see Figure 24).

For auditory and bimodal stimuli there are no significant correlations and the slopes of the Theil-Sen regression are near zero.

To asses potential different multisensory benefits, a measure was introduced in Section 2. The correlation of the computed and the measured working memory performance can be seen in Figure 25 and Figure 26.

For 50% detectable stimuli the correlations show, that the computed and measured d' values do correlate to some extent for the visual and auditory conditions. In the visual condition the number of participants with a measured d' better than predicted by the computed d' are almost the same as the other way around. In the auditory condition all participants but 2 have a higher measured d' than is predicted. This might be another sign for the bad psychometric fits for auditory stimuli. In the bimodal condition most participants measured performance match good with the predicted WMP. Four participants have substantially different measured and predicted WMP (no statistical test was conducted). One participant has a noticeable better measured than predicted WMP, three participants have a noticeable worse measured than predicted WMP.

For stimuli with 75% detection rate the differences between predicted and measured WMP are larger. For all modalities there are approximately as many participants with better measured than predicted WMP as participants with worse measured than predicted WMP.

For the auditory and bimodal condition the correlation is stronger compared to the correlation for stimuli at 50% detection rate, but weaker for the visual one.

Overall the measurement might be a useful tool to investigate potential multisensory benefits in future experiment but needs to be evaluated first. In theory if participants use a guessing strategy similar to the one used in the computations of the $\mathrm{WMP_t}$ and the assumption, that the WMP is the same or very similar for the same stimuli at different detection rates, then a strong correlation should show for unimodal stimuli. Points should lie on or near the angle bisector. Though, at this point with the low amount of data and the suboptimal PFs no real conclusion can be drawn.

The fact that no multisensory benefit is existent, even for stimuli with 99% detection rate and despite the results from Hardiess et al. (2015), can be the results of several different reasons that will be be discussed in the following paragraphs. One should always keep in mind that there might be no multisensory benefit existent in working memory or in particular in the n-back paradigm. However, the huge body of work done on this topic, as mentioned in the introduction, makes this rather unlikely.

A major point is, that the best unimodal performance of the participants for 99% detectable stimuli was far better in the present study than in the earlier study ($d' = 2.89$ vs $d' \approx 2.6$). This might be due to the fact that only three different stimuli were used, compared to the six different stimuli used in the study done by Hardiess et al. (2015). In the present study many participants had ceiling or near ceiling working memory performance, i.e., hit rates larger than 0.95 or false alarm rates smaller than 0.05. In particular this means that participants had at most two misses out of 54 go trials or at most five false alarms in 102 no-go trials (dismissed trials with no answers are ignored here). This should not happen because hit rates of 1 and false alarm rates of 0 (or near those extrema) are seen as statistical sampling error and should only occur occasionally (Macmillan & Creelman, 2004).

Assuming that participants can do the task perfectly and errors are solely the results of lapses, with a lapse rate $\leq 5\%$, a d' can be computed which should not be exceeded. Thus, 5% lapses, or rather 95% correct trials can be seen as a lower limit for a perfect memory performance. In the present setting five errors lead to a lapse rate of 0.49. Two misses and three false alarms yield to a sensitivity index of $d' \approx 3.68$. For the 99% detectable stimuli out of the 15 participants, three exceeded this value in the visual condition, one in the auditory and four in the bimodal condition. In total eight out of 45 ($> 17\%$) d' values were at a critical high value. This is clearly more than what would be expected as a sampling error. If this analysis is done on single blocks, the percentage gets even higher. With a existing ceiling effect, even at unimodal conditions, a multisensory benefit is implausible to occur.

In contrast to the results from Hardiess et al. (2015) the performance in the auditory condition is significantly worse than the performance in the visual condition for stimuli at 75% and 99% detection rate. This is exactly the other way around than what they measured in their study.

This result coincides with the perception of participants. All participants stated that the auditory working memory task was more difficult than the visual and/or bimodal working memory task (see Figure 28). Some even reported that the auditory stimuli were annoying or distracting them.

One possible cause might be the difficulty to distinguish the auditory stimuli. The measurement of the auditory threshold was the same as the measurement of the threshold for the visual stimuli, but participants reported that they perceived the visual stimuli easier to distinguish. This can be seen in the PFs. For visual stimuli the thresholds for the three stimuli (left, middle, right) have low variance between and within subjects. For auditory stimuli the thresholds have a way larger variance between subjects and far more important within subjects (see Figure 12). Furthermore, the 95% confidence intervals of the threshold are huge for the auditory PFs. This behaviour is most likely due to the different thresholds for the three different stimuli, most probably the result of differences in sensitivity of the two ears. Also, the modulation of the sound was only done with the interaural time difference ignoring other influences as the interaural intensity difference and the head-related transfer function. Thus, the direction of the auditory stimuli sounded rather artificial.

Another hypothesis that can explain the absence of a multisensory benefit in the working memory task is the stated effect of the the reversed performance in the unimodal condition. In general the visual perception is the dominant one if people are confronted with inconsistent information gained from vision and hearing (McGurk & MacDonald, 1976; Warren, Welch, & McCarthy, 1981) or for vision and touch (Rock & Victor, 1964) but only when the variance of the visual estimation is lower than the variance of haptic estimation (Ernst & Banks, 2002). Thus, if no integration takes place, the multisensory benefit might only occur when a participant is not as certain about the memory of the visual stimulus compared to the auditory. In such a setting the auditory stimulus helps to compensate for the dominant but unreliable visual stimulus. If integration is existent, a dominant but unreliable stimulus might get the same weight as a subordinate but reliable stimulus in the integration process and therefore lead to a better memorization.

On the other hand, if the memory of the visual stimulus is more reliable and no integration is existent, the memory of the less reliable and subordinate auditory stimulus might not be taken into account or at least have not much influence. If both stimuli get integrated the uncertainty of the subordinated stimulus may lead to the mitigation or the complete disregard of the stimulus.

But because there are many other factors that differ in the two experiments, no statement can be made whether this hypothesis holds. In a recent experiment a similiar effect showed (Brunetti et al., 2017). In their study participants attended a 2-back task. Participants had to concentrate either on the visual or auditory modality. In the task irrelevant modality the stimulus presented could be congruent or incongruent. Furthermore the stimulus material where either 1-4 black disks and 1-4 sinus bursts (quantities) or the written numbers 1,2,3,4 paired with the same numbers spoken (digits). The better RT performance occurred for auditory targets only for quantities and for visual targets only for digits. So the same argument would hold with the assumption, that the sinus bursts are a worse precept than the dots, i.e. that quantities are easier to remember in the visual-spatial domain. The digits in contrast are presumably easier to remember in the phonological domain.

It is also possible that participants used different strategies in the working memory task.

In the questionnaire participants were asked if they used a special strategy. As can be seen in Section 4.6 nine of 15 participants did state doing so. In particular seven of these nine participants stated that they used some sort of external strategy. Those participants used the help of an external - non working memory - process. Six participants reported, that they either repeated the direction of the last two stimuli in their mind (phonological loop) or in a low voice to themselves. One participant stated that they used their fingers to note the direction of the last two stimuli.

It is likely that the participants that used an external strategy used the concepts *left*, *middle* and *right*, because of the necessity to have a concept for the mentioned external strategies. The concept of direction was already introduced in the instructions for the threshold estimation task (see Figure 48) and participants most likely adapted and learned it.

Therefore, this behaviour is only possible if the stimulus material can be conceptualized. If on the other hand the stimuli are in some abstract form, as in Hardiess et al. (2015), it is way harder to have some concepts and use an external strategy.

To investigate a potential different working memory performance using an external strategy a $2 \times 3 \times 3$ mixed model ANOVA was conducted with the between factor *strategy* and the within factors *modality* and *detection rate*. The results of this ANOVA can be seen in Figure 20. The results show that there is no difference between the two groups and the overall effects are the same as without the factor *strategy*.

It has to be noted, that participants that did not note using an external strategy still might have used one. They might just not have noticed it themselves or might not have rated it as a 'special strategy' as asked for in the questionnaire. Hence, a difference might still exist and should be controlled for with appropriate stimuli.

Another problem is the question how to treat trials in which no answer was registered, i.e. the participant did not give an answer in time. Since the distribution of go trials and no-go trials is not even, the sensitivity index is different whether these trials are treated as go or no-go trials. Thus, in the analysis such trials were simply dismissed and the number of go and no-go trials was adjusted accordingly. However, this leads to less total trials that can be evaluated. Furthermore, participants probably did not answer when they did not know which the correct answer was. Dismissing those trials therefore leads to higher values of d'. As can be seen in Figure 22 A for conditions with 50% detectable stimuli the number of trials with no registered answer is the highest, getting lower for conditions with 75% detectable stimuli and being minimal for trials with 99% detectable stimuli. In general the auditory conditions have the most trials with no registered answer.

In Figure 22 B one can see the distribution of these trials for all participants and conditions. There was one participant with a noticeable high number of trials with no registered answer. But there are others that have many trials with no registered answer as well. No exclusion criterion was defined for trials with no registered answer that a participant should maximal reach and doing it post hoc is not ideal either. Therefore, a mean split was done to examine a possible effect. A $2 \times 3 \times 3$ mixed modal ANOVA with the within factors *modality*, *detection rate* and the between factor *no answers* did reveal that there is no significant effect of the factor *no answers* neither as interaction nor as main effect. The significance of other effects did also not differ from the 3×3 repeated measures ANOVA without the factor *no answer*.

The RTs show no unexpected behaviour. A full model repeated measures ANOVA with the factors modality and detection rate showed a significant interaction of the two factors. Both main effects reached significance, too. As can be seen in Figure 18 the reaction time of the auditory condition did barely change for the different detection rates.

In Figure 19 the differences between the modalities for each detection rate are displayed. As can be seen the auditory modality was slower at all three detection rates followed by the bimodal and visual modalities. This difference is significant for auditory and visual modality at all detection rates and between auditory and bimodal for detection rates of 75% and 99%.

The bimodal and visual conditions showed lower RTs for stimuli with a higher detection rate. This is most likely due to the fact, that the auditory stimuli is time dependent and a participant needs some time to detect the stimulus, even at 99% detection rate. The visual stimuli in contrast are time constant and thus a participant can detect the stimulus much faster compared to the auditory ones and faster the easier the stimulus can be detected.

In the bimodal condition participants stated in the questionnaire (Figure 29), that they concentrated more on the visual stimuli in the bimodal condition. Thus, if the visual stimulus is detected in the bimodal condition a participant already can make the decision whether the target stimulus was the same as the sample stimulus or not and therefore the bimodal condition can be faster than the auditory one.

5.3 General discussion

In summary there is to say that there is potential for experimental improvements. There are several possible confounding variables that need to be accounted for. They are in no particular order the possible strategies participants can use, the different unimodal working memory performance and the potential different multisensory benefit gained from that. The fact that the auditory stimuli had bad psychometric fits and might even have been disruptive and the usage of only three different stimuli that lead to the ceiling effect that occurred in this task.

In summary for future research there are a few variables to improve or tweak. Most importantly, the auditory stimuli should be improved. One way to improve auditory stimuli is to use sound and pictures of instruments or tools, like Werner and Noppeney (2009) did. Or if the congruence of visual and auditory stimuli is not wanted, one could use sounds of different frequencies, like different chords, or sinus tones and not use different directions of the tones at all.

A interesting modification of the task would be to use stimuli that are not congruent or have any relationship. It is possible that the task is more difficult if the stimuli have no inherent connection or are even incongruent. Integration might be different with congruent stimuli compared to neutral or incongruent stimuli.

To really answer the question, whether or not it makes a difference if one can merge both stimuli to one concept further experiments are necessary with a suited experimental design. The analysis of the data revealed that there was no significant difference. But the experiment was not designed to examine this question and participants might have used the same strategies in both groups.

The visual modality often is the dominant one, if compared with the auditory or other modalities. This can lead to the assumption, only a multisensory benefit occurs if the non dominant domain really helps. In other words if the auditory stimuli are way harder to perceive and distinguish than people might actually not benefit from the information of them at all. On the other hand, if the auditory information is of high quality and the visual information is worse, than the visual information may still be regarded as an important factor of the overall information. This proportion could be tweaked in further experiments to see if a effect is existent.

Furthermore, some methodical improvements should be done. The threshold estimation task should be a nAFC task to circumvent the problems of the usage of a categorization task.

Also, there should be an exclusion criterion for participants with too many trials with no registered answer in the working memory task or the task could be changed such that the next trial only starts after an answer was registered as was done in the threshold estimation task.

Another parameter to tune is the perceptual load. In the bimodal condition a participant has more perceptual input than in unimodal conditions, since there are two stimuli presented simultaneously. To balance the load, stimuli only containing noise can be presented in the other modality in unimodal conditions, as Werner and Noppeney (2009) have done. This has some advantages. First there is a signal in both modalities in each condition. In the unimodal conditions the other modality though does not contain any information. This leads to the same perceptual load in every condition and the important information has to be extracted from an similar input. If the noise in one modality is disruptive, as was reported by some participants, then this effect occurs in every condition.

Ideally auditory and visual stimuli should be tested for identical working memory performance. Unfortunately, there is probably no way to do this without investing a large amount of work and time. But with a set of such stimuli one could test the mentioned possible influences on the multisensory working memory performance.

To conclude the most important step to further investigate is to find a set of auditory stimuli that do not possess the problems that were existent in the present work.

References

Angelaki, D. E., Gu, Y., & DeAngelis, G. C. (2009). Multisensory integration: psychophysics, neurophysiology, and computation. *Current opinion in neurobiology*, *19*(4), 452–458.

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in cognitive sciences*, *4*(11), 417–423.

Baddeley, A. D. (2012). Working memory: theories, models, and controversies. *Annual review of psychology*, *63*, 1–29.

Baddeley, A. D., Allen, R. J., & Hitch, G. J. (2011). Binding in visual working memory: The role of the episodic buffer. *Neuropsychologia*, *49*(6), 1393–1400.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of learning and motivation*, *8*, 47–89.

Ban, H., Preston, T. J., Meeson, A., & Welchman, A. E. (2012). The integration of motion and disparity cues to depth in dorsal visual cortex. *Nature neuroscience*, *15*(4), 636.

Blauert, J. (1997). *Spatial hearing: the psychophysics of human sound localization.* MIT press.

Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. *Spatial vision*, *10*, 433–436.

Brunetti, R., Indraccolo, A., Mastroberardino, S., Spence, C., & Santangelo, V. (2017). The impact of cross-modal correspondences on working memory performance. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(4), 819.

Brungart, D. S., & Rabinowitz, W. M. (1999). Auditory localization of nearby sources. head-related transfer functions. *The Journal of the Acoustical Society of America*, *106*(3), 1465–1479.

Campbell, F. W., & Robson, J. (1968). Application of fourier analysis to the visibility of gratings. *The Journal of physiology*, *197*(3), 551–566.

Coren, S. (2003). *Sensation and perception.* Wiley Online Library.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological bulletin*, *104*(2), 163.

Cowan, N. (1999). An embedded-processes model of working memory. *Models of working memory: Mechanisms of active maintenance and executive control*, *20*, 506.

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current directions in psychological science*, *19*(1), 51–57.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429.

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in cognitive sciences*, *8*(4), 162–169.

Feddersen, W., Sandel, T., Teas, D., & Jeffress, L. (1957). Localization of high-frequency tones. *the Journal of the Acoustical Society of America*, *29*(9), 988–991.

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of cognitive neuroscience*, *11*(5), 473–490.

Goolkasian, P., & Foos, P. W. (2005). Bimodal format effects in working memory. *The American journal of psychology*, 61–78.

Hardiess, G., Erhardt, S., & Mallot, H. (2015). Bimodal perceptual integration facilitates representation in working memory. In *Perception* (Vol. 44, pp. 341–341).

Heikkilä, J., Alho, K., Hyvönen, H., & Tiippana, K. (2015). Audiovisual semantic congruency during encoding enhances memory performance. *Experimental psychology*, *62*(2), 123.

Hillis, J. M., Ernst, M. O., Banks, M. S., & Landy, M. S. (2002). Combining sensory information: mandatory fusion within, but not between, senses. *Science*, *298*(5598), 1627–1630.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., Broussard, C., et al. (2007). What's new in psychtoolbox-3. *Perception*, *36*(14), 1.

Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision research*, *39*(16), 2729–2737.

Li, R., Polat, U., Makous, W., & Bavelier, D. (2009). Enhancing the contrast sensitivity function through action video game training. *Nature neuroscience*, *12*(5), 549–551.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide.* Psychology press.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746.

Moore, B. C. (2012). *An introduction to the psychology of hearing.* Brill.

Morey, C. C., & Cowan, N. (2004). When visual and verbal memories compete: Evidence of cross-domain limits in working memory. *Psychonomic bulletin & review*, *11*(2), 296–301.

Nandy, A. S., & Tjan, B. S. (2008). Efficient integration across spatial frequencies for letter identification in foveal and peripheral vision. *Journal of Vision*, *8*(13), 3–3.

Owsley, C., Sekuler, R., & Siemsen, D. (1983). Contrast sensitivity throughout adulthood. *Vision research*, *23*(7), 689–699.

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, *10*(4), 437–442.

Prins, F., N. & Kingdom. (2015). *Palamedes: Matlab routines for analyzing psychophysical data.* www.palamedestoolbox.org.

Prins, N. (2009). Faa kingdom. *Palamedes: Matlab routines for analyzing psychophysical data*.

Prins, N. (2013). The psi-marginal adaptive method: How to give nuisance parameters the attention they deserve (no more, no less). *Journal of Vision*, *13*(7), 3–3.

Quak, M., London, R. E., & Talsma, D. (2015). A multisensory perspective of working memory. *Frontiers in human neuroscience*, *9*.

Rock, I., & Victor, J. (1964). Vision and touch: An experimentally created conflict between the two senses. *Science*, *143*(3606), 594–596.

Schieber, F. (1992). Aging and the senses. *Handbook of mental health and aging*, *2*, 251–306.

Schütt, H. H., Harmeling, S., Macke, J. H., & Wichmann, F. A. (2016). Painfree and accurate bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vision research*, *122*, 105–123.

Schwarz, W., & Miller, J. O. (2014). When less equals more: Probability summation without sensitivity improvement. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(5), 2091.

Stanford, T. R., & Stein, B. E. (2007). Superadditivity in multisensory integration: putting the computation in context. *Neuroreport*, *18*(8), 787–792.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior research methods, instruments, & computers*, *31*(1), 137–149.

Stein, B. E., Burr, D., Constantinidis, C., Laurienti, P. J., Alex Meredith, M., Perrault, T. J., ... others (2010). Semantic confusion regarding the development of multisensory integration: a practical solution. *European Journal of Neuroscience*, *31*(10), 1713–1720.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, *9*(4), 255–266.

Thompson, V. A., & Paivio, A. (1994). Memory for pictures and sounds: Independence of auditory and visual codes. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *48*(3), 380.

Treisman, M. (1998). Combining information: probability summation and probability averaging in detection and discrimination. *Psychological Methods*, *3*, 252–265.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of memory and language*, *28*(2), 127–154.

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*(6984), 748.

Wahn, B., & König, P. (2017). Is attentional resource allocation across sensory modalities task-dependent? *Advances in cognitive psychology*, *13*(1), 83.

Warren, D. H., Welch, R. B., & McCarthy, T. J. (1981). The role of visual-auditory “compellingness” in the ventriloquism effect: Implications for transitivity among the spatial senses. *Perception & Psychophysics*, *30*(6), 557–564.

Werner, S., & Noppeney, U. (2009). Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. *Cerebral cortex*, *20*(8), 1829–1842.

Williams, E. (1949). Experimental designs balanced for the estimation of residual effects of treatments. *Australian Journal of Chemistry*, *2*(2), 149–168.

Xie, Y., Xu, Y., Bian, C., & Li, M. (2017). Semantic congruent audiovisual integration during the encoding stage of working memory: an erp and sloreta study. *Scientific reports*, *7*(1), 5112.

Appendices

A Psychometric Fits

Psychometric functions fitted in the visual and auditory threshold estimation task. Each page contains the auditory or visual fits of three participants.

Each row contains the data of one participant. The psychometric functions in one row from left to right: Pf for the left stimulus, PF for the middle stimulus, PF for the right stimulus and PF for all stimuli pooled together. If the last plot in one row has a red curve, then a second fit was done as explained in Section 3.4 and it is shown instead of the original fit. Transparent plots are PFs of excluded participants.

Each plot shows the fitted psychometric function as the black (or red) curve. The dotted line marks the guessing rate if (33%). The vertical black line marks the detection threshold (50% detection rate) and the horizontal black line shows the 95% confidence interval. The blue dots show the percent correct for the individual signal levels. The size of the marker represents the number of times a stimulus level got shown (the larger the marker the more times it got shown).

The PFs for the visual stimuli can be seen for participant 1-3 in Figure 30, for participant 4-6 in Figure 31, for participant 7-9 in Figure 32, for participant 10-12 in Figure 33, for participant 13-15 in Figure 34, for participant 16-18 in Figure 35, for participant 19-21 in Figure 36.

The PFs for the auditory stimuli can be seen for participant 1-3 in Figure 37, for participant 4-6 in Figure 38, for participant 7-9 in Figure 39, for participant 10-12 in Figure 40, for participant 13-15 in Figure 41, for participant 16-18 in Figure 42, for participant 19-21 in Figure 43.

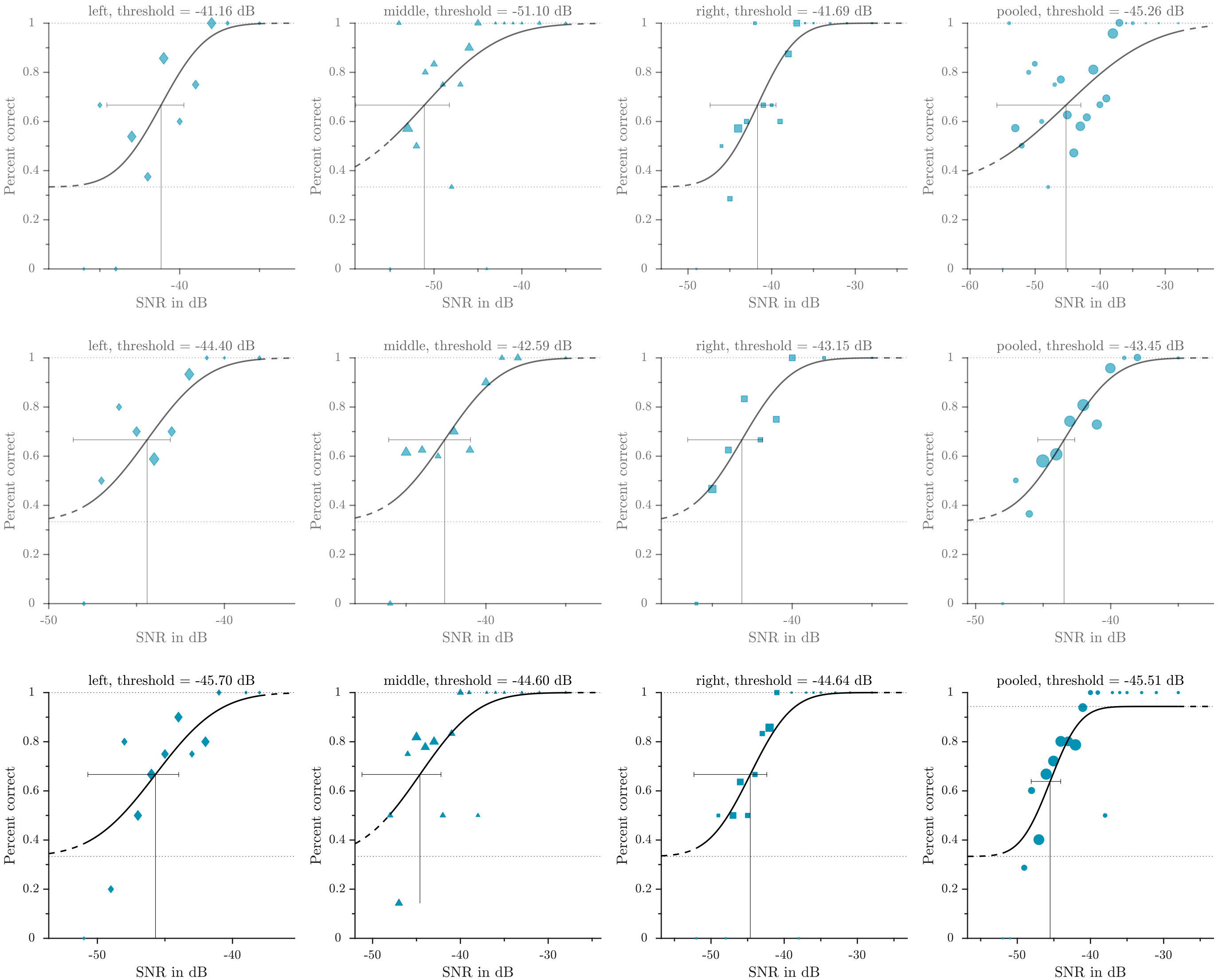

Figure 30: Psychometric functions of visual stimuli. Participants 1-3.

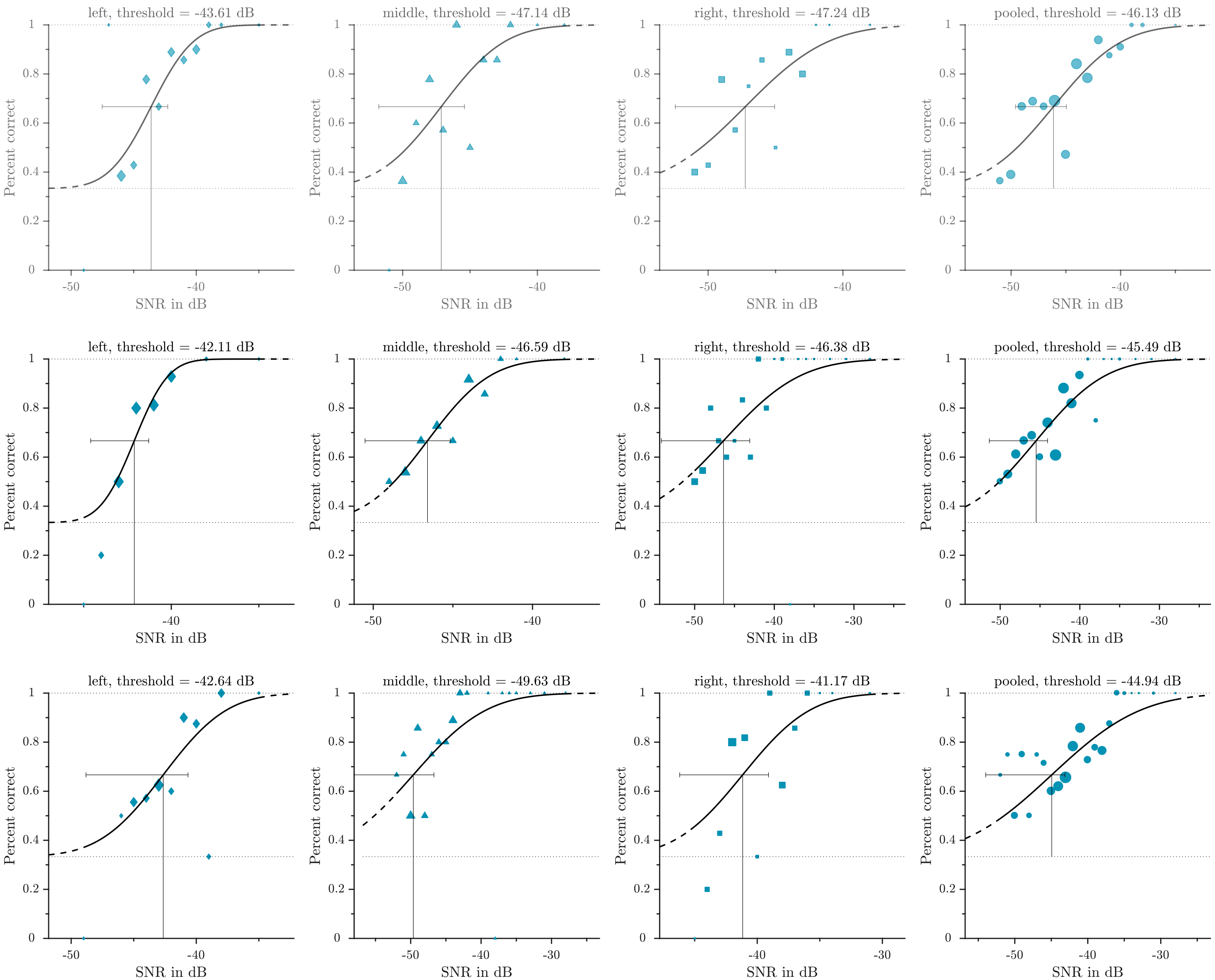

Figure 31: Psychometric functions of visual stimuli. Participants 4-6.

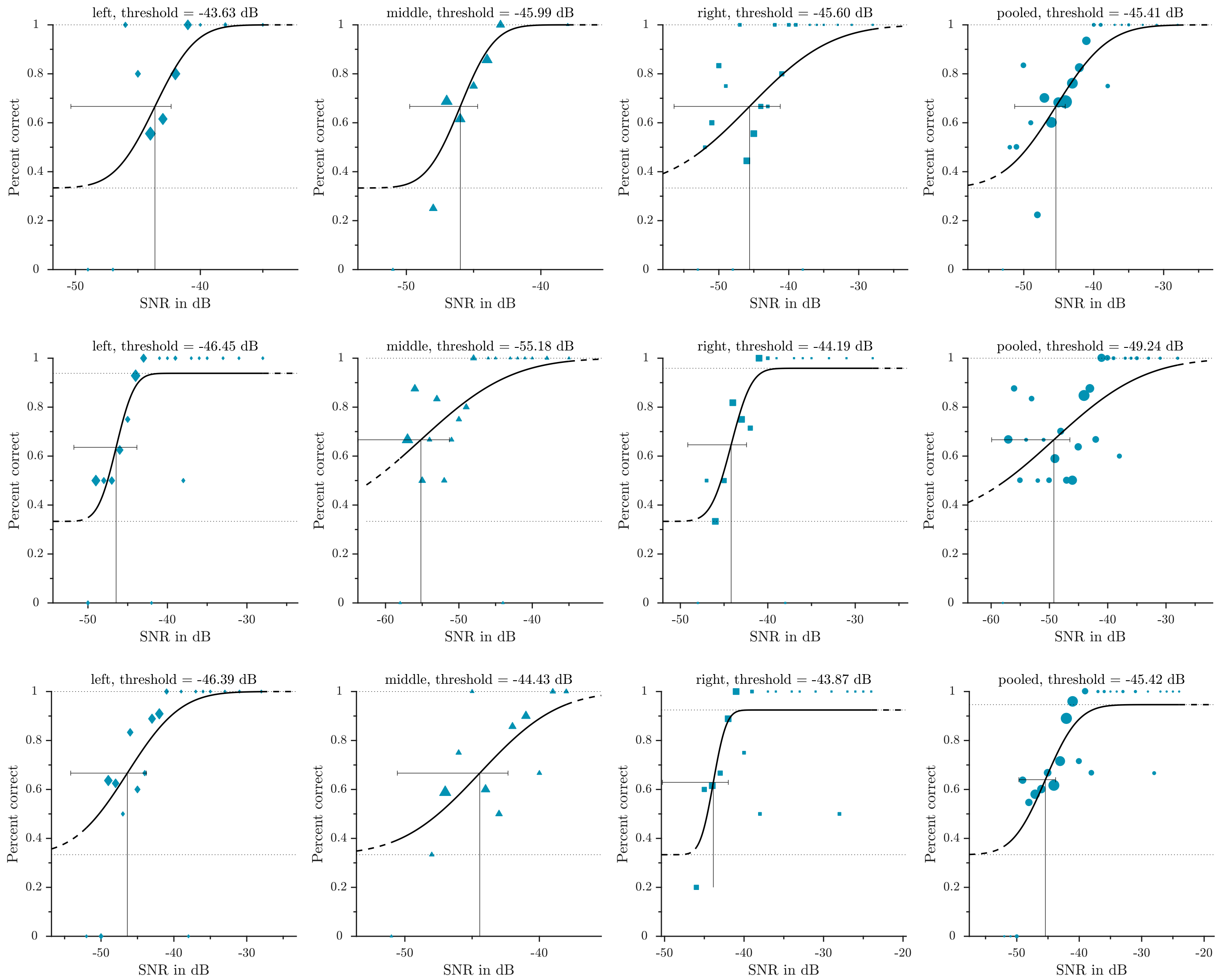

Figure 32: Psychometric functions of visual stimuli. Participants 7-9.

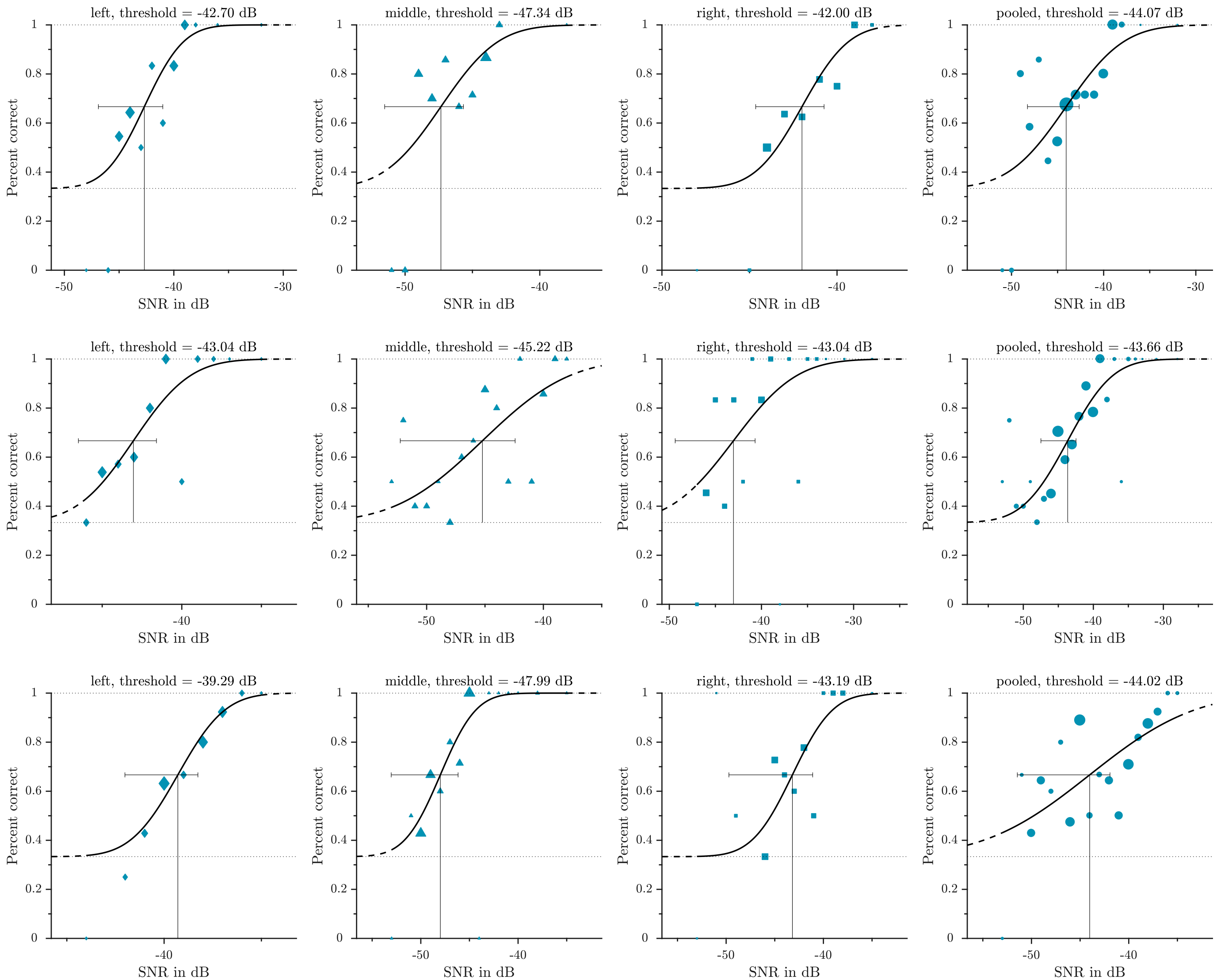

Figure 33: Psychometric functions of visual stimuli. Participants 10-12.

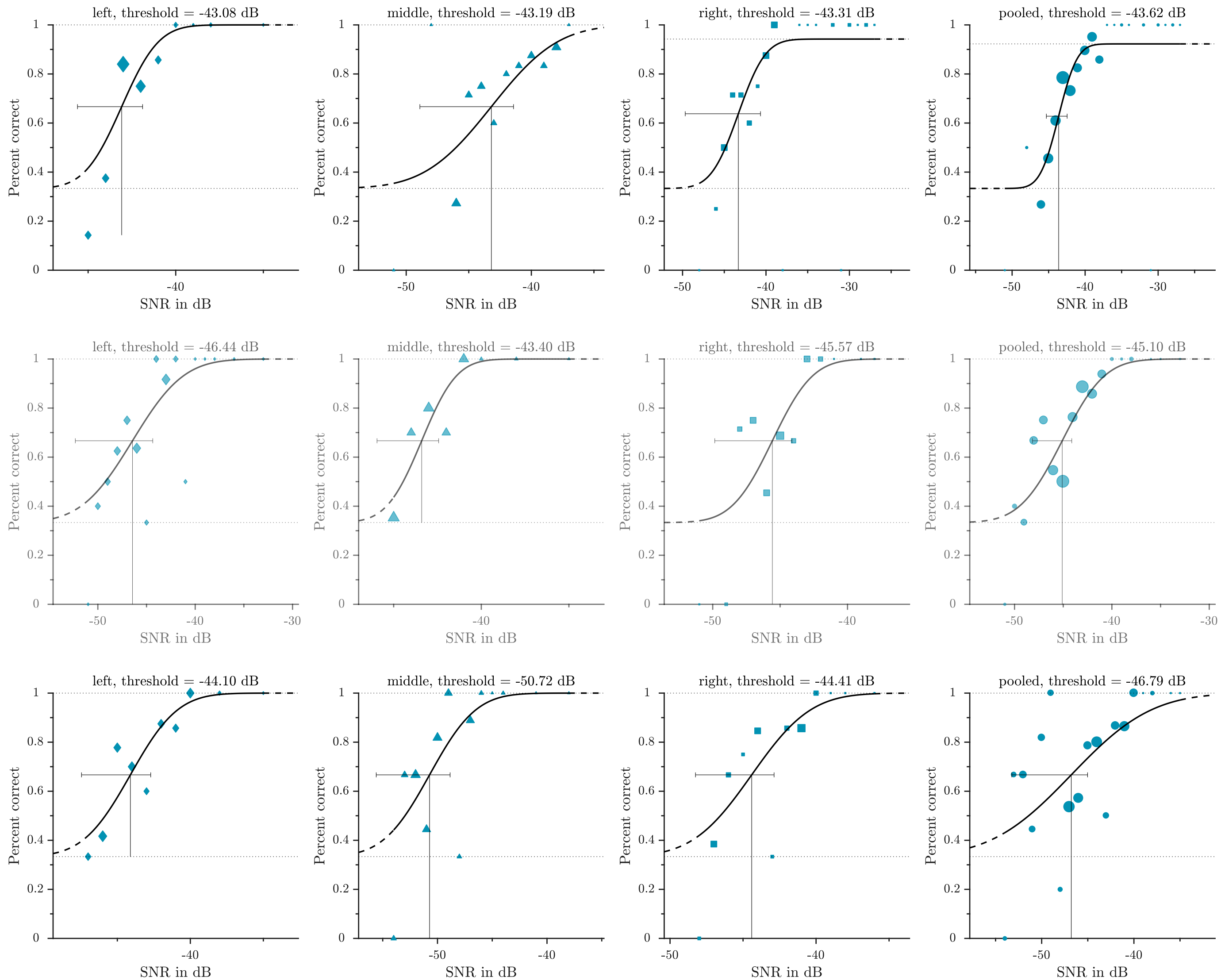

Figure 34: Psychometric functions of visual stimuli. Participants 13-15.

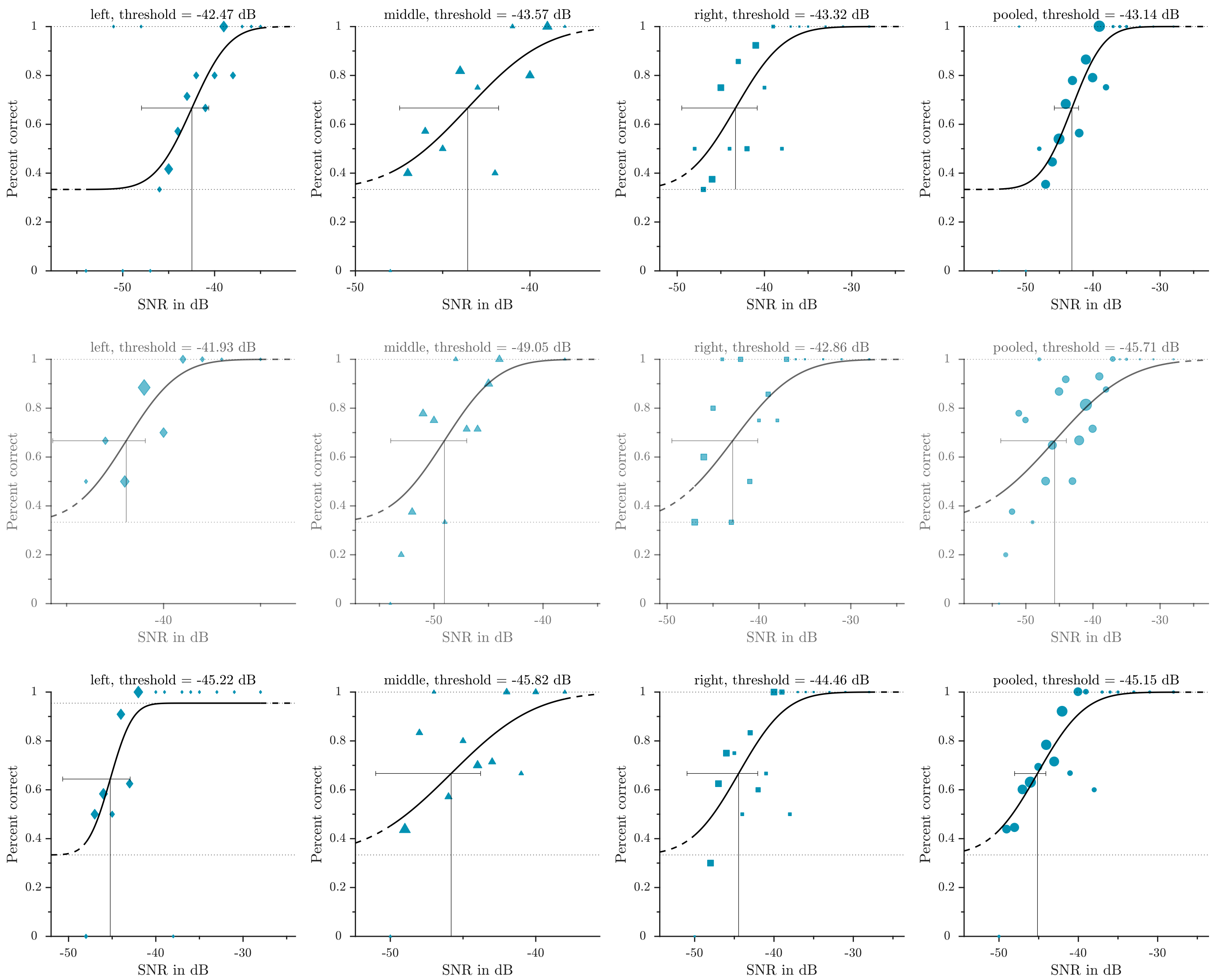

Figure 35: Psychometric functions of visual stimuli. Participants 16-18.

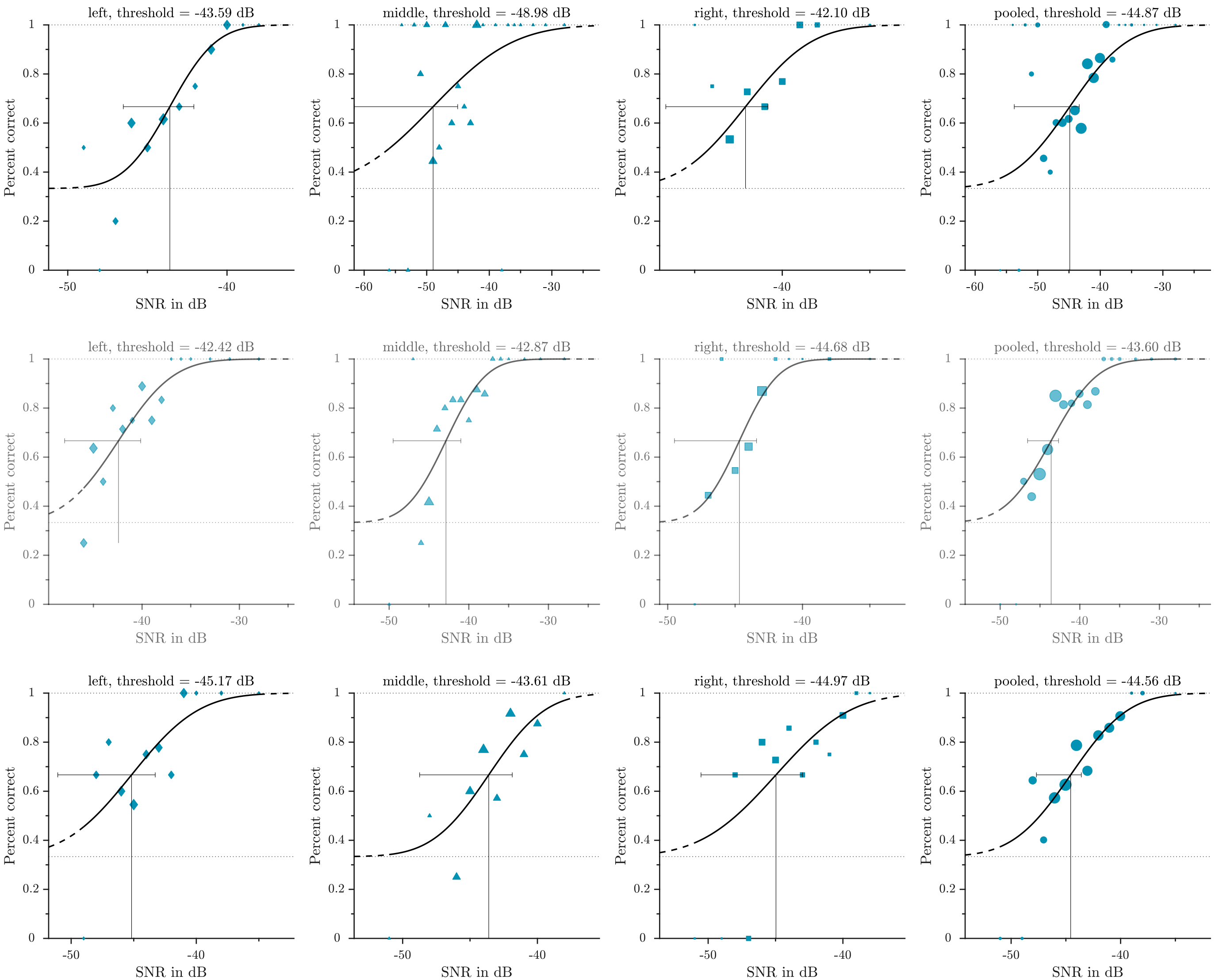

Figure 36: Psychometric functions of visual stimuli. Participants 19-21.

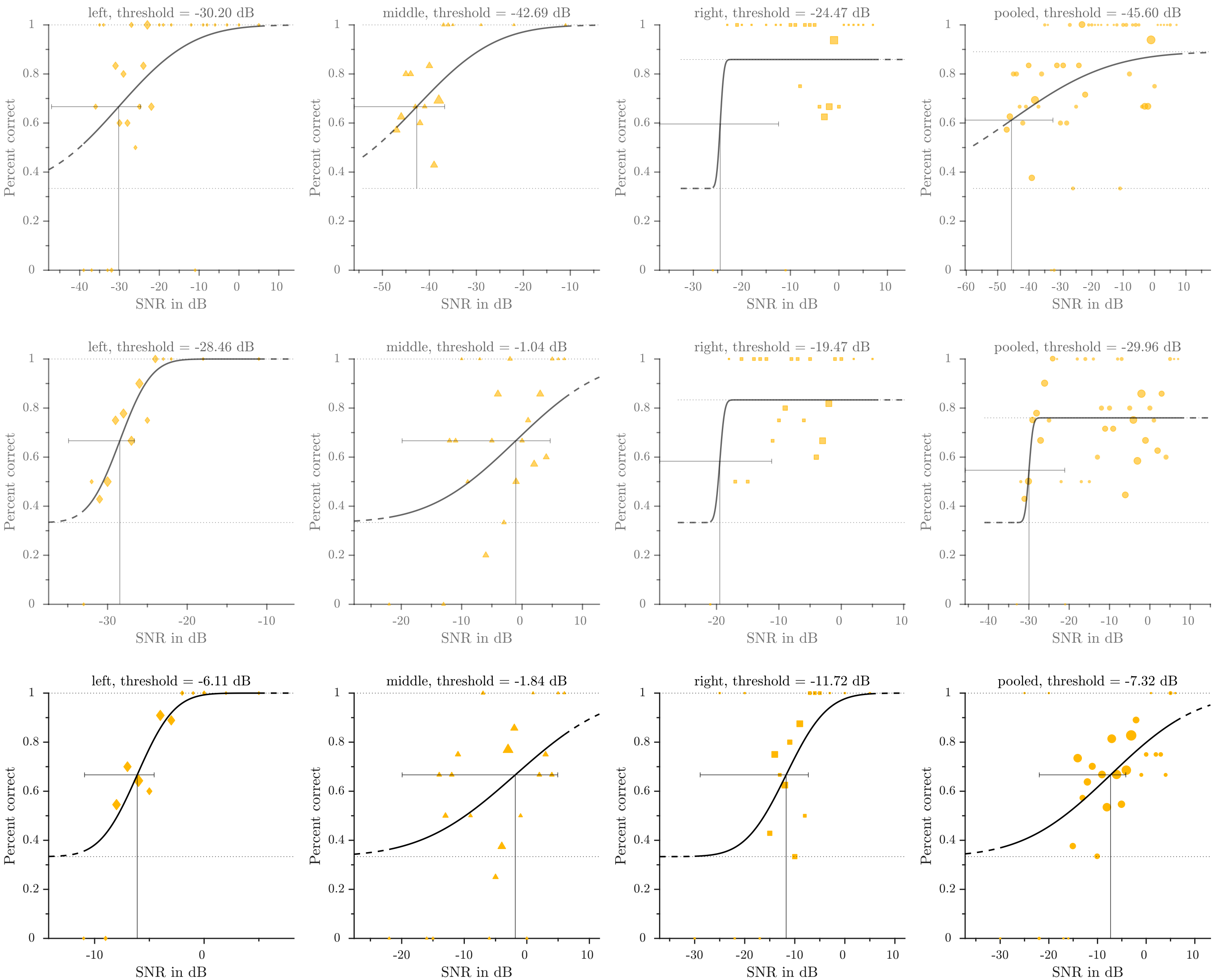

Figure 37: Psychometric functions of auditory stimuli. Participants 1-3.

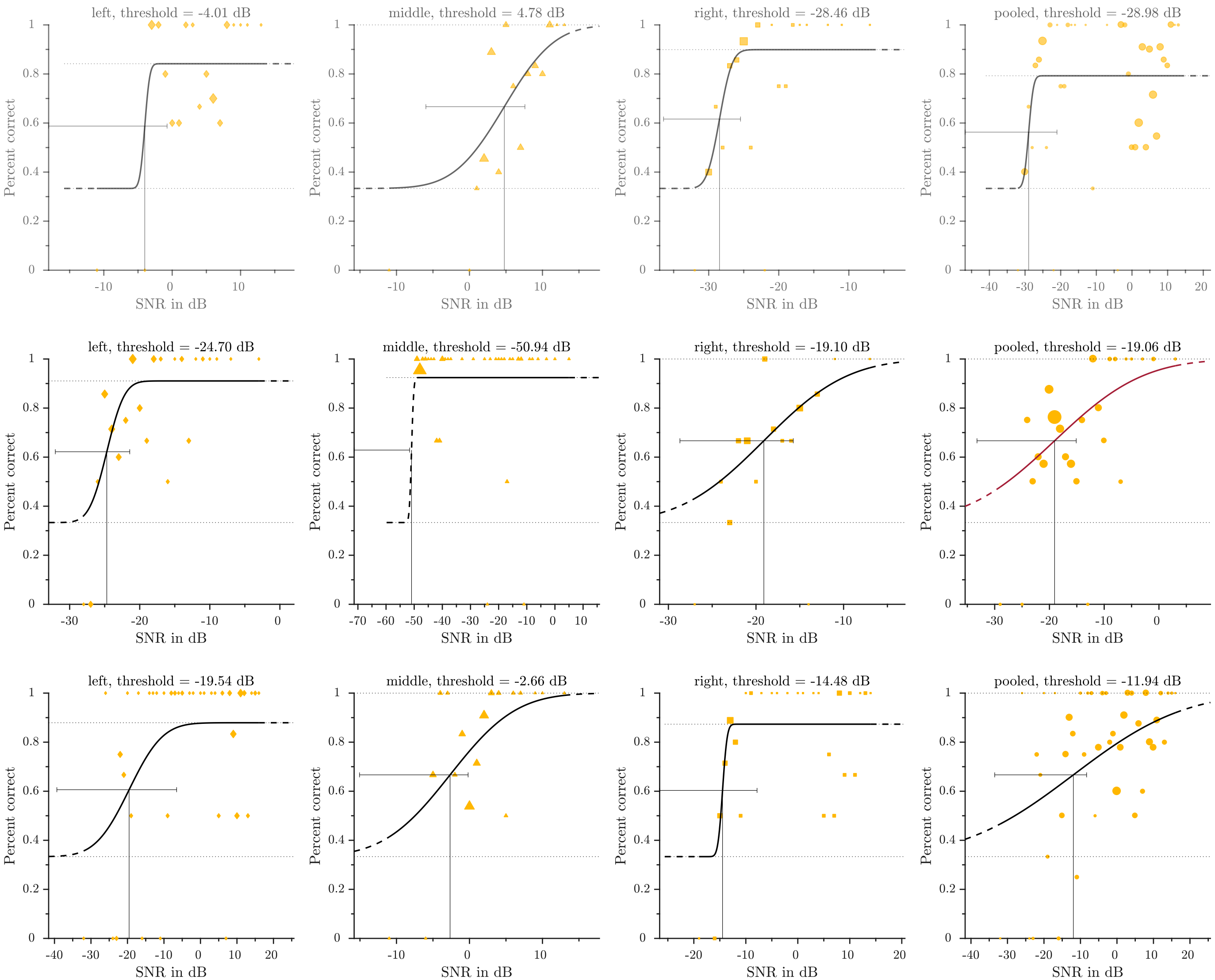

Figure 38: Psychometric functions of auditory stimuli. Participants 4-6.

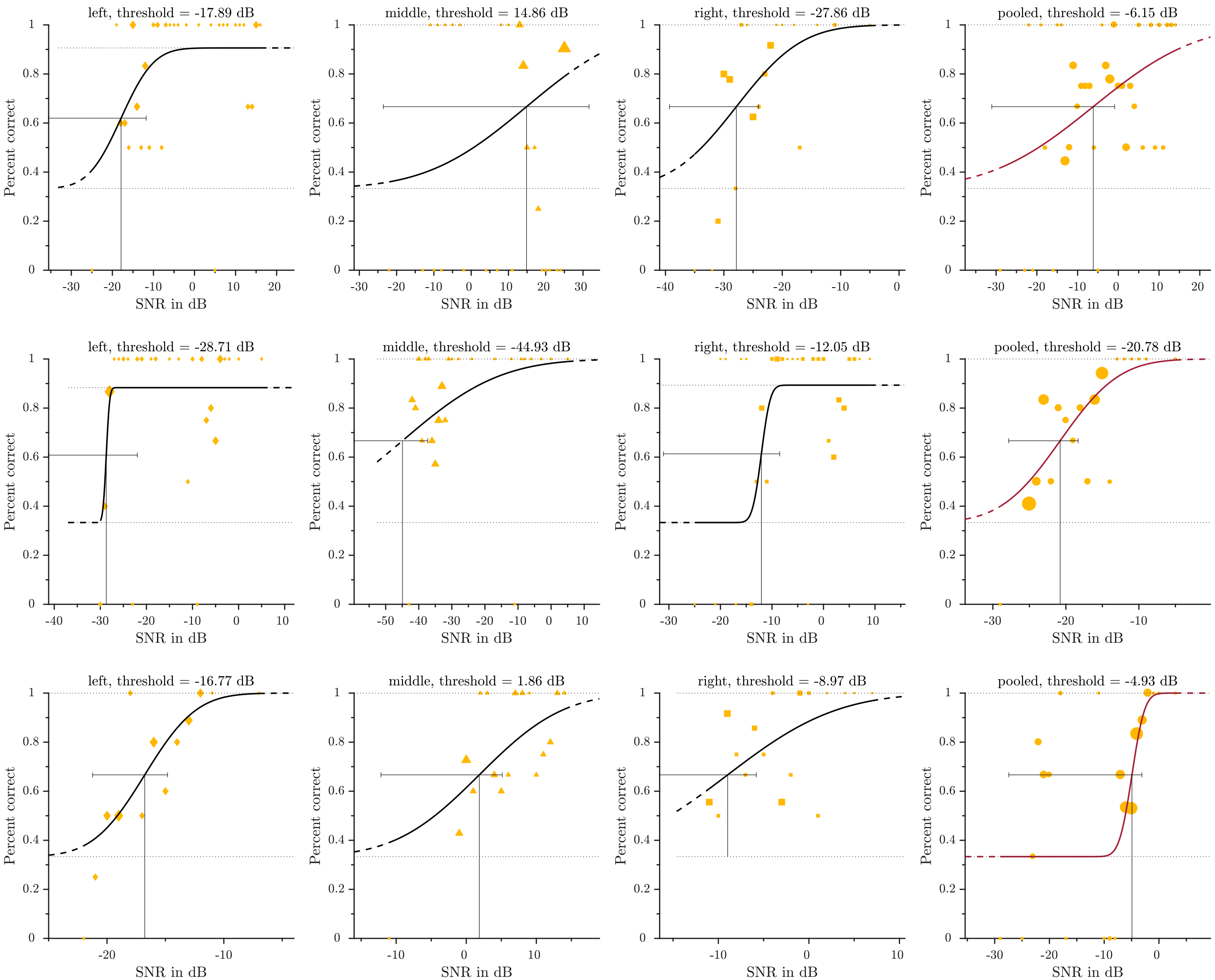

Figure 39: Psychometric functions of auditory stimuli. Participants 7-9.

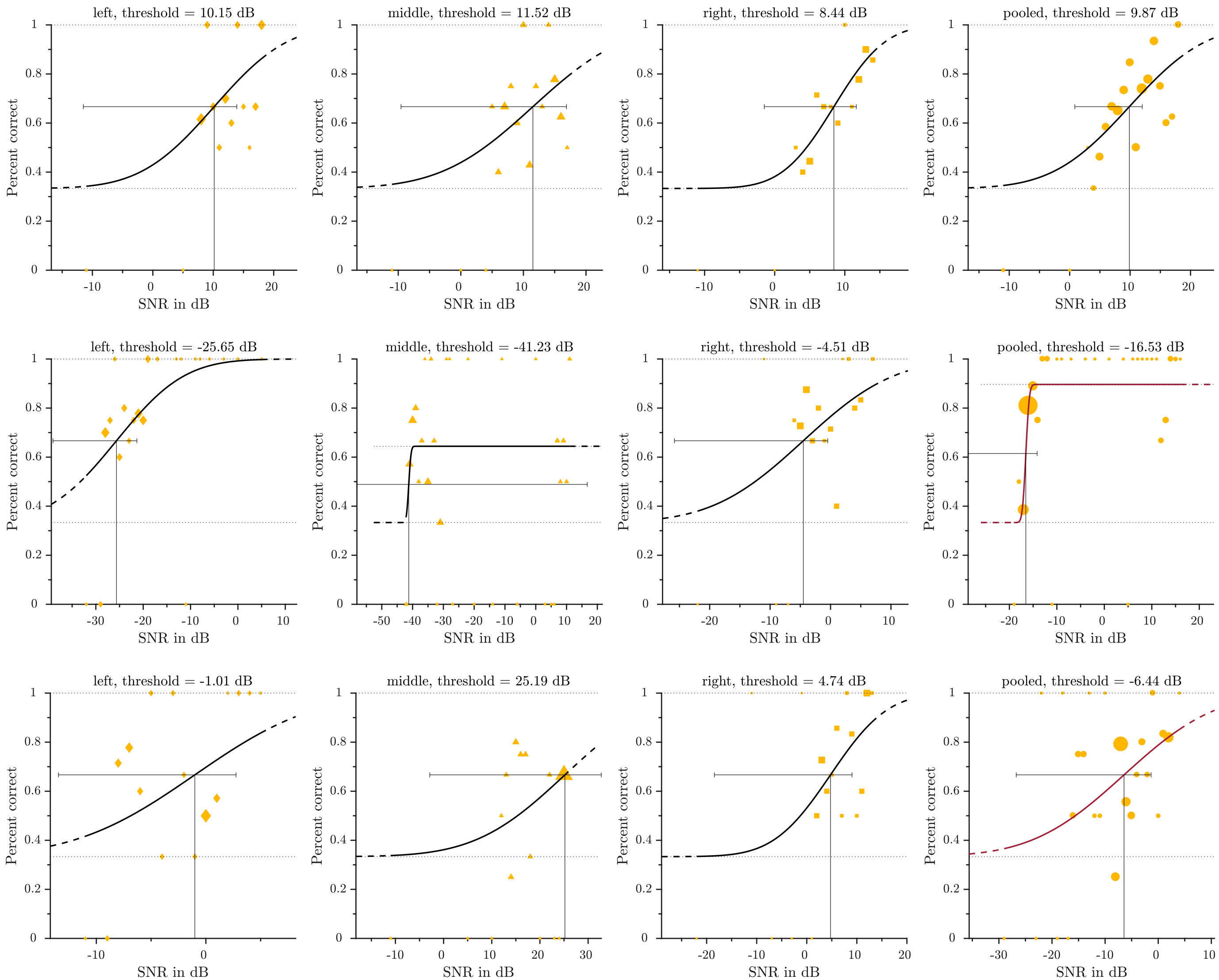

Figure 40: Psychometric functions of auditory stimuli. Participants 10-12.

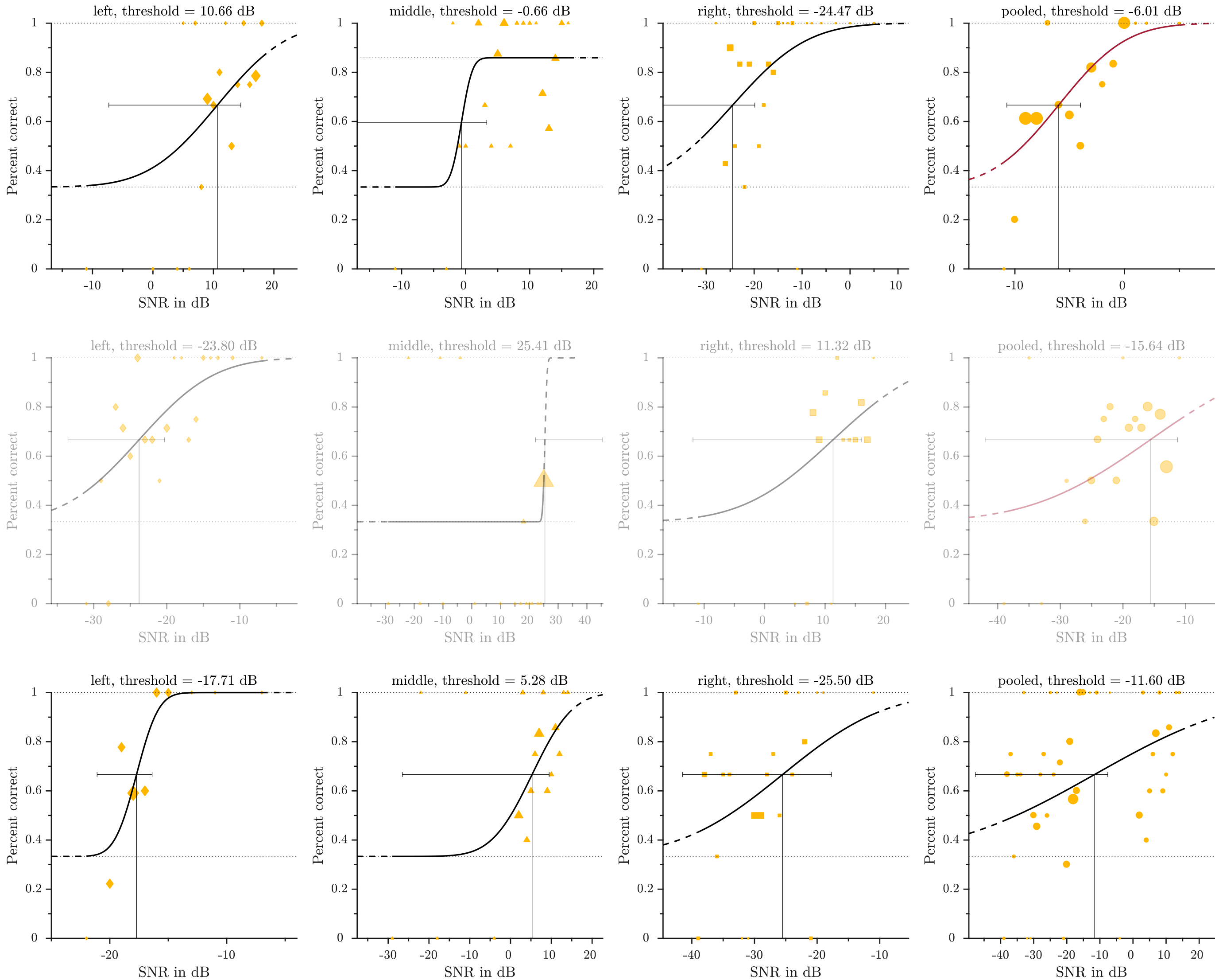

Figure 41: Psychometric functions of auditory stimuli. Participants 13-15.

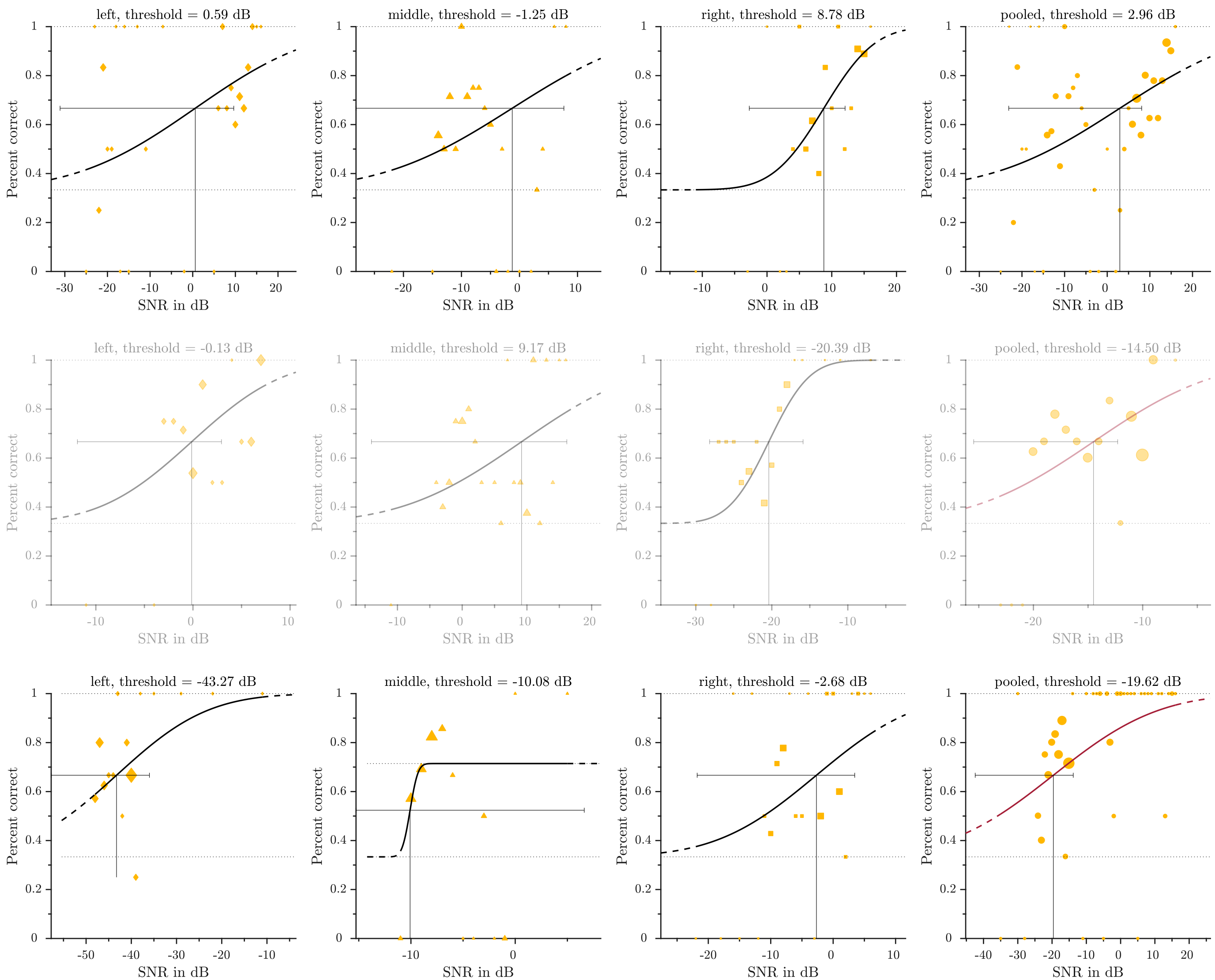

Figure 42: Psychometric functions of auditory stimuli. Participants 16-18.

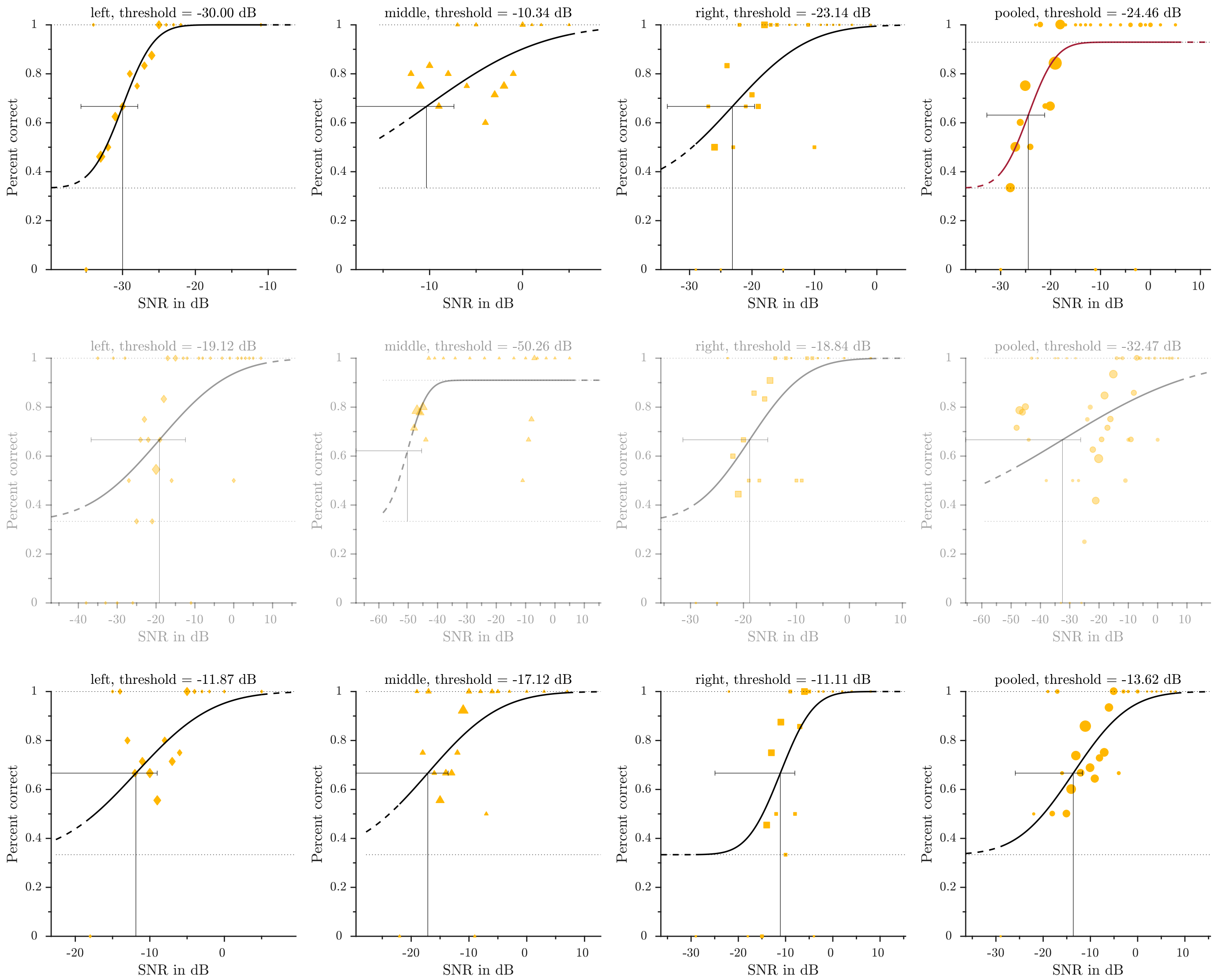

Figure 43: Psychometric functions of auditory stimuli. Participants 19-21.

B Individual working memory performances (d')

The performance of the working memory task for each participant (except of excluded ones) can be seen in Figure 44, Figure 45 and Figure 46.

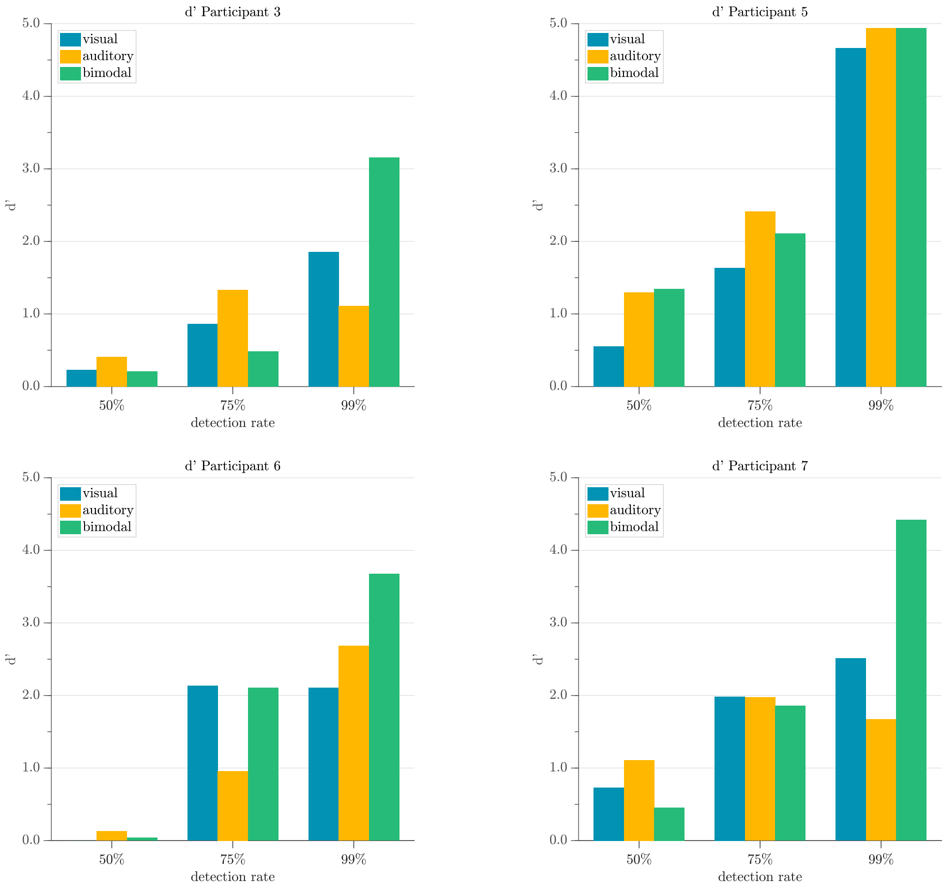

Figure 44: Individual working memory performance of participants 3,5,6 and 7.

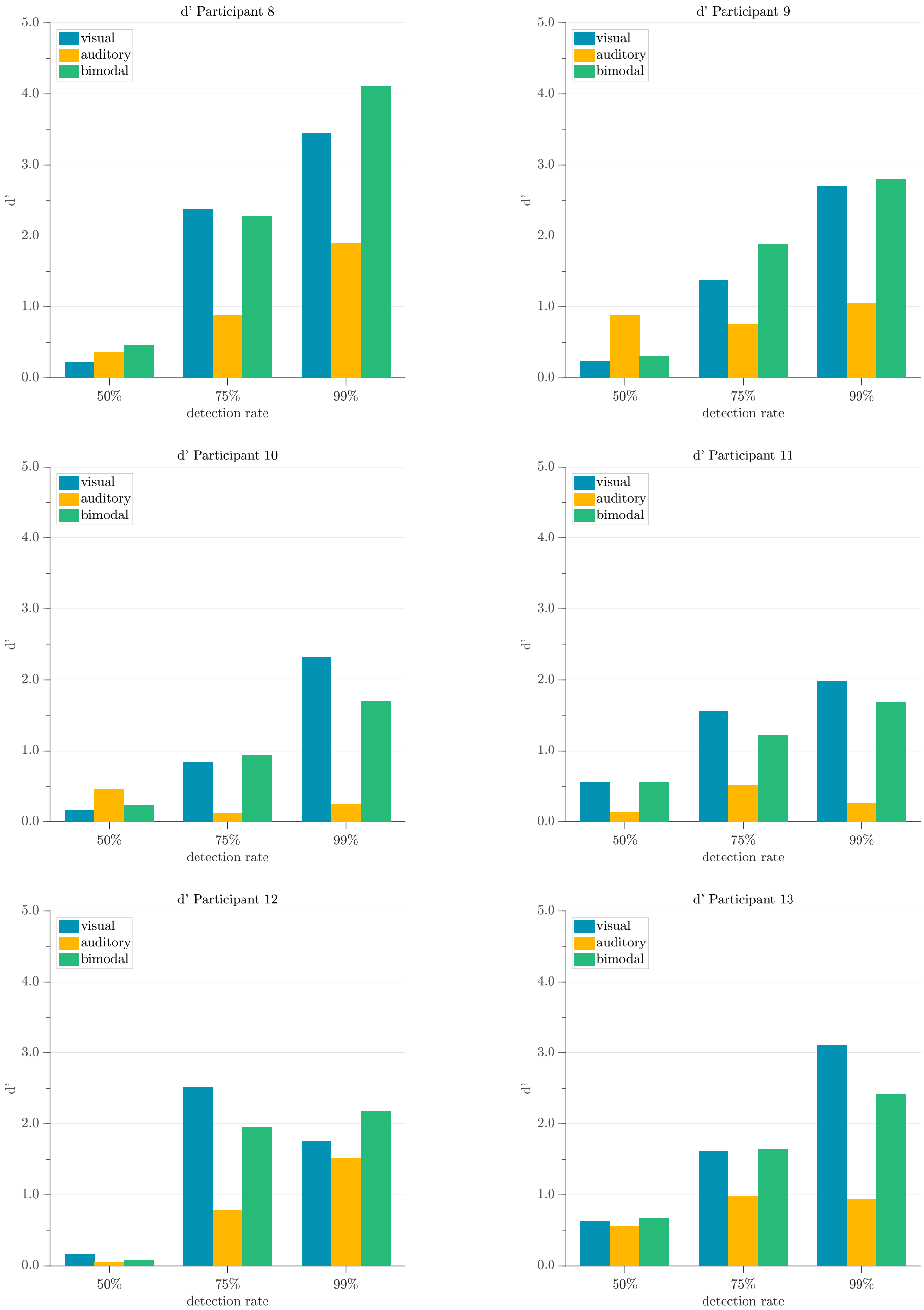

Figure 45: Individual working memory performance of participants 8-13.

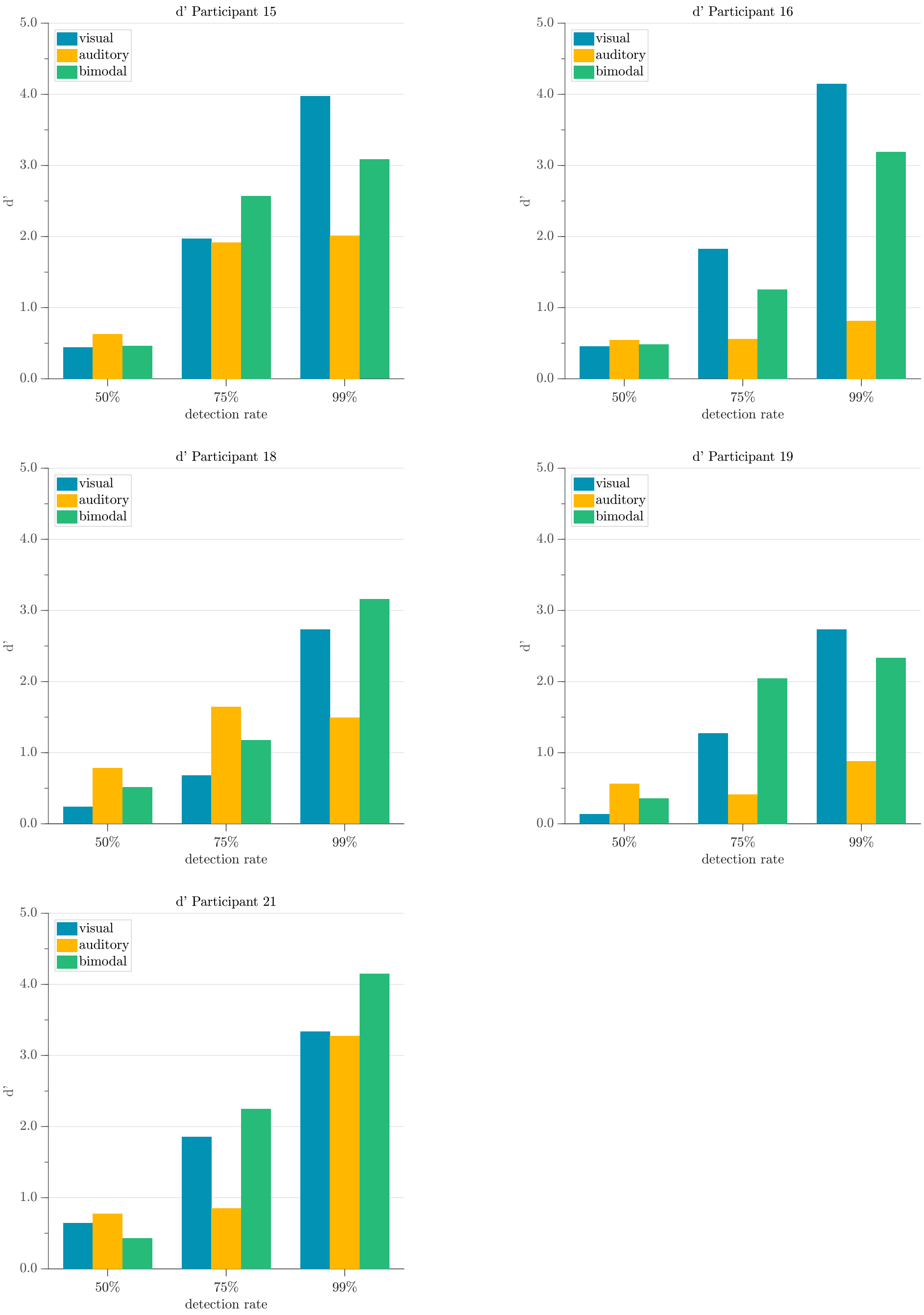

Figure 46: Individual working memory performance of participants 15, 16, 18, 19 and 21.

C Miscellaneous

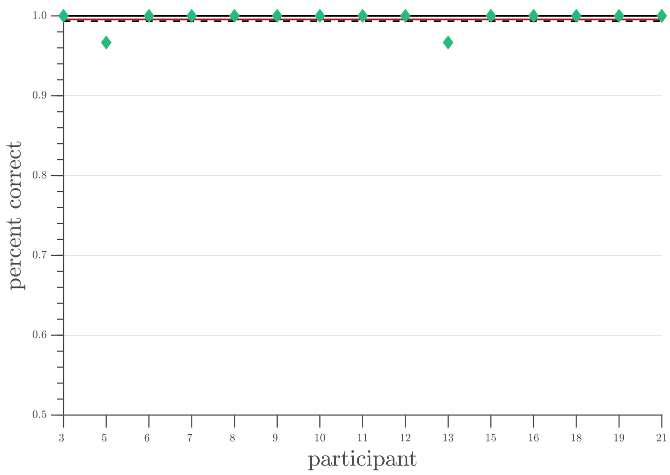

Figure 47: Bimodal performance of each participant for unimodal detection rates at 99%. Each diamond shows the performance of a participant. The dashed black line marks the unimoal pc, the solid black line the bimodal performance computed with the PSM and the red line marks the mean over all participants.

Trial	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Block1	1	1	1	1	3	3	1	1	3	1	1	3	1	1	3	2	3	3
Block2	3	2	1	2	2	2	2	3	3	2	3	1	1	1	1	1	1	2
Block3	2	3	2	2	2	3	1	3	3	3	3	2	2	1	3	1	2	3

Trial	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Block1	3	1	2	1	2	2	1	2	2	1	2	2	1	2	2	3	3	2
Block2	3	1	1	2	2	1	1	2	1	2	3	2	2	1	3	3	3	2
Block3	1	3	1	1	3	2	2	1	3	2	1	1	3	1	2	3	2	1

Trial	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
Block1	2	3	2	3	3	3	2	2	2	1	2	2	3	3	3	3	1	1
Block2	3	3	1	2	2	3	1	2	1	1	2	3	1	3	3	3	3	3
Block3	2	1	2	3	1	1	3	1	3	1	2	1	3	2	2	2	1	3

Table 2: Example for three blocks with different stimulus assignment. Go Trials are marked with a grey background

P1	A	B	I	C	H	D	G	E	F
P2	B	C	A	D	I	E	H	F	G
P3	C	D	B	E	A	F	I	G	H
P4	D	E	C	F	B	G	A	H	I
P5	E	F	D	G	C	H	B	I	A
P6	F	G	E	H	D	I	C	A	B
P7	G	H	F	I	E	A	D	B	C
P8	H	I	G	A	F	B	E	C	D
P9	I	A	H	B	G	C	F	D	E
P10	F	E	G	D	H	C	I	B	A
P11	G	F	H	E	I	D	A	C	B
P12	H	G	I	F	A	E	B	D	C
P13	I	H	A	G	B	F	C	E	D
P14	A	I	B	H	C	G	D	F	E
P15	B	A	C	I	D	H	E	G	F
P16	C	B	D	A	E	I	F	H	G
P17	D	C	E	B	F	A	G	I	H
P18	E	D	F	C	G	B	H	A	I

Table 3: Williams Design for nine conditions leading to 18 different task orders.

3AFC Task

In 3 verschiedenen Experimenten werden Sie mit verschieden Stimuli, Bildern, Tönen oder Bilder/Töne gleichzeitig, konfrontiert:

- Das Bild kann nach links gedreht, gerade/nicht gedreht oder nach rechts gedreht sein.

- Der Ton, ein Dreiklang auf einer Gitarre gespielt, kann von links, vorne oder rechts ertönen.
- Wenn Töne und Bilder zusammen präsentiert werden, dann immer wie folgt:
    - Ein Ton von links mit dem nach links gedrehten Bild.
    - Ein Ton von vorne mit dem geraden Bild.
    - Ein Ton von rechts mit dem nach rechts gedrehten Bild.

Alle drei Experimente sind wie folgt aufgebaut:

- In jedem Experiment ist die Reihenfolge der Bilder und/oder Töne zufällig.
- Jedoch werden alle (links, gerade/vorn, rechts) gleich oft gezeigt.
- Ein Stimulus wird immer 1 Sekunden lang präsentiert.
- **Ihre Aufgabe ist es**, bei jedem Stimulus zu entscheiden welches Bild und/oder welcher Ton gezeigt wurde.

… bitte wenden!

- Um zu antworten drücken Sie bitte die linke, mittlere oder rechte Maustaste wie folgt:
    - Ein Ton von links und/oder eins nach links gedrehten Bild.
      ➔ **Linke Maustaste**
    - Ein Ton von vorne und/oder ein gerades Bild.
      ➔ **Mittlere Maustaste**
    - Ein Ton von rechts und/oder eins nach rechts gedrehten Bild.
      ➔ **Rechte Maustaste**

- Die Bilder und Töne sind in jedem Durchgang mehr oder weniger stark mit Rauschen verdeckt. In manchen Durchgängen ist das Rauschen so stark, dass man den Stimulus kaum erkennen kann. In diesen Durchgängen müssen Sie trotzdem versuchen richtig zu antworten. Wenn Sie den Stimulus gar nicht mehr wahrnehmen können müssen Sie raten.
- Es muss immer geantwortet, also eine Maustaste gedrückt werden. Der Nächste Durchgang startet nur, wenn Sie eine Maustaste gedrückt haben.
- Nach jedem Block oder wenn ein neues Experiment gestartet wird können Sie bei Bedarf eine Pause einlegen.

Bei auftretenden Fragen oder Unklarheiten wenden Sie sich bitte an den Versuchsleiter.

Bitte führen Sie die Aufgabe so genau wie möglich aus…

Viel Spaß! ☺

Figure 48: Instruction of the threshold estimation tasks.

<u>2 - back Task</u>

In 9 verschiedenen Experimenten werden Sie mit je einer Sequenz bestehend aus der Abfolge von Stimuli konfrontiert. Dabei handelt es sich um die gleichen wie in den Experimenten zuvor:

- Das Bild, das nach links gedreht, gerade/nicht gedreht oder nach rechts gedreht ist.
- Der Ton, der von links, vorne oder rechts ertönt.
- Die Bilder und Töne wie eben gepaart.
- Die Töne und Bilder sind schwach, mäßig oder stark verrauscht. Dies Ändert sich in jedem Experiment.

Die Stimuli sind zufällig über die gesamte Sequenz verteilt.

- Jedes Experiment besteht aus einem Block mit je 54 Stimuli.
- Ein Stimulus wird immer 1 Sekunden lang präsentiert, danach folgt eine Pause von 1 Sekunde in der Sie einen grauen Bildschirm sehen.
- **Ihre Aufgabe ist es**, bei jedem Stimulus zu entscheiden ob der gerade präsentierte Stimulus identisch ist mit dem, welcher **<u>zwei</u>** Schritte zuvor gezeigt wurde. Dies ist natürlich erst ab dem 3. Stimulus möglich.
- Ist der Ton, das Bild oder Bild/Ton **<u>gleich</u>** mit dem zwei Stellen zuvor, muss die **<u>linke Maustaste</u>** gedrückt werden.
- Ist der Ton, das Bild oder Bild/Ton **<u>nicht gleich</u>** mit dem zwei Stellen zuvor, muss die **<u>rechte Maustaste</u>** gedrückt werden.
- Es muss also immer eine Maustaste gedrückt werden. Dies kann während der Stimulus-Präsentation geschehen, oder auch in der Pause danach.
- Bsp.-Sequenz: 1 - 3 - 1 - 2 - … Bei der Präsentation des 3. Stimulus (1) muss die linke Maustaste gedrückt werden, da zwei Stellen davor auch eine 1 gezeigt wurde. Beim 4. Stimulus (2) dagegen die rechte Maustaste, da zwei Stellen vorher eine 3 präsentiert wurde.

… bitte wenden!

- Wenn eine Bedingung beendet ist, startet automatisch die nächste Bedingung. Jedoch können Sie nach Bedarf eine Pause einlegen, bevor Sie mit der Durchführung beginnen.
- Das gesamte Experiment dauert ca. 25 Minuten.

Bei auftretenden Fragen oder Unklarheiten wenden Sie sich bitte an den Versuchsleiter.

Bitte führen Sie die Aufgabe so genau wie möglich aus…

Viel Spaß! ☺

Figure 49: Instructions of the working memory tasks.

Fragebogen

ID:

Zum Abschluss noch ein kurzer Fragebogen mit 4 Fragen. Bitte beantworten Sie die Fragen der Reihe nach. Kehren Sie nicht zu einer vorherigen Frage zurück um diese zu ändern.

1. **Wie schwer fanden Sie das Arbeitsgedächtnis Experiment am heutigen Tag?**

einfach — schwer

2. **Haben Sie eine besondere Strategie verfolgt? Falls ja bitte erläutern Sie kurz Ihre Strategie.**

◯ Nein

◯ Ja: ______________________________

____

3. **Fiel Ihnen einer oder zwei der drei Aufgabentypen einfacher als die anderen bzw. der andere? Mehrfachnennungen möglich.**

◯ Alle gleich schwer

◯ Visuell (nur Bilder) einfacher

◯ Auditiv (nur Töne) einfacher

◯ Bimodal (Bilder + Töne) einfacher

4. **Worauf haben Sie sich bei den bimodalen (Bilder + Töne) Aufgaben konzentriert.**

nur Töne — eher Töne — gleich stark — eher Bilder — nur Bilder

Questionnaire

ID:

Finally a questionnaire with four short questions. Please answer the questions in order. Do not go back to an earlier question to change it!

1. **How would you rate the difficulty of the working memory task today?**

easy — difficult

2. **Did you have a strategy? If so, please describe it in a few words.**

◯ No

◯ Yes______________________________

____

3. **Did you think that one (or two) of the below mentioned types were easier than the others? You can pick more than one answer.**

◯ All the same difficulty

◯ Visual (only pictures) easier

◯ Auditory (only tones) easier

◯ Bimodal (pictures + tones) easier

4. **In the bimodal task (pictures + tones), did you concentrate on pictures or tones?**

only tones — rather tones — both equal — rather pictures — only pictures

Figure 50: Questionnaire that each participant had to fill out at the end of the experiment (German and English version). All participants filled out the German version.