Multisensory Interactions of Visual and Tactile Modalities

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

MohammadAli Nikouei Mahani

aus Kerman/Iran

Tübingen

2020

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	12.03.2021
Stellvertretender Dekan	Prof. Dr. József Fortágh
1. Berichterstatter	Prof. Dr. Rolf Ulrich
2. Berichterstatter	Prof. Dr. Hartmut Leuthold

Abstract

Theories and models of multisensory perception have been drawing attention in the past decade. Understanding the multisensory perception is great of interest for cognitive scientists and is crucial for engineers to develop artificial sensory devices such as wearable and sensory substitution systems. The potential of touch for artificial sensory devices as well as the strong link between touch and vision motivated us to study the interactions of vision and touch with the focus on touch stimulations from a custom designed artificial sensory device. Specifically, we investigated three different type of interactions through three studies: The multisensory perception and calibration for visual-tactile information in a realistic environment (study 1), multimodal task-irrelevant effect (study 2), and finally cross-modal learning and integration of vision and touch under novel tactile experiences (study 3). Study 1 investigated how participants would adjust their perception to the discrepancy among sensory information under varying reliability conditions. Our findings showed participants switched from an integration strategy to a selection or to a calibration strategy according to reliability condition of the sensory information. Our comprehensive simulation and modelling revealed that the optimal causal inference not only depends on the amount of the cue conflict, but also on the relative reliability of stimuli across different modalities. We investigated, through study 2, how task irrelevant information affects the perception of task relevant information in a multimodal setup. Particularly, we extended the well-known Simon experiment to the multimodal edition, in which two source of information, from two modalities, deliver the task irrelevant information. We conducted a visual-tactile and a visual-auditory extension of Simon task. Both experiments showed expected visual congruency effects for task-irrelevant information, such as the slower responses and less accurcy in incongruent than congruent conditions. We additionally observed congruency effects of tactile task-irrelevant information, although the effects were smaller than the visual task-irrelevant information. However, the congruency effects of auditory task-irrelevant information were insufficient and negligible. It revealed that in the co-presence of visual and auditory spatial task-irrelevant information, visual information dominates the auditory congruency effect. A proposed multimodal extension of the Diffusion Model for Conflict Tasks (MDMC) was fitted to the results of both experiments. MDMC provided reasonable fits for the experimental data while keeps the genuine assumptions of DMC. We finally investigated, through study 3, whether and how tactile information from a non invasive sensory feedback device can be learned or even integrated with visual perception. The performance of unimodal tactile perception was improved over seven blocks of training, showed that participants learned the associations between the visual and tactile information. Further analyses showed that participants integrated the unexperienced tactile information with experienced visual information from the very beginning of the training, despite of the fact that the accuracy of tactile modality was initially lower than the visual modality in some conditions. Proposed computational models revealed that participants initially employed a metacognitive oriented decision integration policy. However, they later switched to an optimal Bayesian integration of sensory inputs. We believe our findings connects the topic of multisensory perception to the artificial sensory feedback devices.

Zusammenfassung

Theorien und Modelle der multisensorischen Wahrnehmung haben in den letzten Jahren Aufmerksamkeit erregt. Das Verständnis der multisensorischen Wahrnehmung ist für Kognitionswissenschaftler von großem Interesse und für Ingenieure von entscheidender Bedeutung, um künstliche sensorische Geräte zu entwickeln. Das Berührungspotential für künstliche sensorische Geräte sowie die starke Verbindung zwischen Berührung und Sehen haben uns motiviert, die Wechselwirkungen zwischen Sehen und Berührung zu erforschen, wobei der Schwerpunkt auf Berührungsstimulationen mit einem speziell entwickelten künstlichen sensorischen Gerät liegt. Insbesondere untersuchten wir drei verschiedene Arten von Interaktionen in drei Studien: Die multisensorische Wahrnehmung und Kalibrierung für visuell-taktile Informationen in einer realistischen Umgebung (Studie 1), den multimodalen aufgabenunrelevanten Effekt (Studie 2) und schließlich das modalübergreifende Lernen und Integrieren von Sehen und Berühren unter neuartigen taktilen Erfahrungen (Studie 3). In Studie 1 wurde untersucht, wie sich die Teilnehmer unter verschiedenen Zuverlässigkeitsbedingungen an den Cue-Konflikt zwischen sensorischen Informationen anpassen würden. Unsere Ergebnisse zeigten, dass die Teilnehmer je nach Zuverlässigkeitsbedingung der sensorischen Informationen von einer Integrationsstrategie zu einer Auswahl oder zu einer Kalibrierungsstrategie wechselten. Unsere umfassende Simulation und Modellierung hat gezeigt, dass die optimale kausale Inferenz in einer Umgebung mit variierender Zuverlässigkeit nicht nur vom Ausmaß der multimodalen Diskrepanz abhängt, sondern auch von der relativen Zuverlässigkeit der Stimuli. Analysen der Vertrauensberichte zeigten, dass wir in Studie 2 untersucht haben, wie sich aufgabenunrelevante Informationen auf die Verarbeitung aufgabenrelevanter Informationen in einem multimodalen Setup auswirken. Insbesondere haben wir die bekannte Simon-Aufgabe auf die multimodale Ausgabe erweitert, in der aufgabenunrelevante Informationen aus zwei sensorischen Modalitäten hervorgehen können. Wir haben eine visuell-taktile und eine visuell-auditive Erweiterung der Simon-Aufgabe durchgeführt. Beide Experimente zeigten erwartete visuelle Kongruenzeffekte für aufgabenunrelevante Informationen, wie z. B. langsamere Reaktionen und geringere Genauigkeit bei inkongruenten als bei kongruenten Bedingungen. Wir beobachteten auch Kongruenzeffekte von taktilen aufgabenunrelevanten Informationen, obwohl die Effekte geringer waren als die visuellen aufgabenunrelevanten Informationen. Die Kongruenzeffekte auditorisch aufgabenunrelevanter Informationen waren jedoch unzureichend und vernachlässigbar. Es zeigte sich, dass bei gleichzeitiger Anwesenheit von visuellen und auditorischen räumlich aufgabenunrelevanten Informationen visuelle Informationen den auditorischen Kongruenzeffekt dominieren. Eine vorgeschlagene multimodale Erweiterung des Diffusionsmodells für Konfliktaufgaben (MDMC) wurde an die Ergebnisse beider Experimente angepasst. MDMC lieferte angemessene Anpassungen für die experimentellen Daten, während die echten Annahmen von DMC beibehalten wurden. In Studie 3 haben wir schließlich untersucht, ob und wie künstliche taktile Informationen von einem neuartigen nicht-invasiven Gerät gelernt und in visuelle Informationen integriert werden. Die Leistung der unimodalen taktilen Wahrnehmung wurde über sieben Trainingsblöcke verbessert und zeigte, dass die Teilnehmer die Assoziationen zwischen visuellen und taktilen Informationen lernten. Weitere Analysen zeigten, dass die Teilnehmer die unerfahrenen taktilen Informationen von Beginn des Trainings an mit erfahrenen visuellen Informationen kombinierten, obwohl die Genauigkeit der taktilen Modalität anfangs unter bestimmten Bedingungen geringer war als die visuelle Modalität. Vorgeschlagene Rechenmodelle zeigten, dass die Teilnehmer zunächst eine metakognitiv orientierte Entscheidungsintegrationspolitik anwendeten. Später wechselten sie jedoch zu einer optimalen Bayes'schen Integration sensorischer Eingaben. Wir glauben, dass unsere Ergebnisse das Gebiet der multisensorischen Wahrnehmung mit der Entwicklung künstlicher sensorischer Geräte verbinden.

Contents

Publications

Enclosed Journal Publications

Study 1

Mahani, Mohammad-Ali Nikouei, Saber Sheybani, Karin Maria Bausenhart, Rolf Ulrich, and Majid Nili Ahmadabadi. "Multisensory perception of contradictory information in an environment of varying reliability: Evidence for conscious perception and optimal causal inference." *Scientific reports* 7, no. 1 (2017): 1-15.

Study 2

Mahani, Mohammad-Ali Nikouei, Karin Maria Bausenhart, Majid Nili Ahmadabadi, and Rolf Ulrich. "Multimodal simon effect: A multimodal extension of the diffusion model for conflict tasks." *Frontiers in human neuroscience* 12 (2019): 507.

Study 3

Mahani, Mohammad-Ali Nikouei, Karin Maria Bausenhart, Rolf Ulrich, and Majid Nili Ahmadabadi. "Learning to integrate an artificial sensory device: from decision fusion to sensory integration." *Under review* (2020).

Related Publications

Mohammad-Ali Nikouei Mahani, Saber Sheybani, Karin Bausenhart, Rolf Ulrich, Majid Nili Ahmadabadi, “Visual-Tactile Calibration Conforms Tactile Dominant Adaption”, TEAP 2016.

Mohammadali Nikouei Mahani, Saber Sheybani, Karin Bausenhart, Rolf Ulrich, Majid Nili Ahmadabadi (accepted), "Analysis of ERP components in the multisensory perception task across different reliability conditions", PRNI 2016.

Introduction

Humans evolved to benefit from multiple sensory modalities such as visual, auditory, olfactory, gustatory, and tactile. These sensors allow us to perceive the complex and multimodal environment. In many situations, information of more than one sensory modality contributes to a perceptual decision process. We benefit from multisensory perception in order to improve the reliability and accuracy of perception (Drugowitsch, DeAngelis, Klier, Angelaki, & Pouget, 2014; M.-A. N. Mahani, Sheybani, Bausenhart, Ulrich, & Ahmadabadi, 2017; Rowland, Quessy, Stanford, & Stein, 2007). As a matter of fact, studying the interaction among sensory modalities is crucial to understand and model our perceptual system.

Touch and vision both play crucial role in our perception. Vision is one of the most informative modalities while touch, through the skin, is known as the largest sensory organ(Hertenstein & Weiss, 2011) and acts like the physical interface of our body to the environment.

Many studies reported the strong link/interaction between of visual and tactile modalities. Interactions of vision and touch could be so strong that visual cortical area is active during a normal tactile perception. This phenomenon was first observed in a positron emission tomographic (PET) study (Lacey & Sathian, 2016; Sathian, Zangaladze, Hoffman, Neuroreport, & 1997, n.d.). Another example of visual and tactile interactions is the redundant target effects, in which the perception reaction time to a double visual-tactile target is significantly faster than single/double uni-visual/uni-tactile target (Forster, Cavina-Pratesi, Aglioti, & Berlucchi, 2002). This effect actually shows the mutual influence of simultaneous visual and tactile information.

Furthermore, the rubber hand illusion example (Botvinick, Nature, & 1998, n.d.), and interactions among vision and touch during the development(A Streri, Pownall, & Kingerlee, 1993; Arlette Streri & Gentaz, 2004) bold further the importance of both modalities and their interactions.

In other hands, recent advances in technology highly affected life habits and brought novel sensory experiences to our daily life. Various wearable devices (Dakopoulos & Bourbakis, 2010; Johnson & Higgins, 2006), human in loop assistance systems(*Driver assistance system for driver assistance for consumption controlled driving*, 2011; *Pilot Assistance System*, 2015; Krause, Knott, & Benger, n.d.), and sensory substitution systems (Kristjánsson et al., 2016;

Shull & Damian, 2015) are some examples of these novel sensory experiences. Nevertheless, cognitive aspects of these artificial devices could benefit from more studies.

Vision-touch as novel sensory experience

Touch has a high potential for providing novel sensory experiences through wearable devices. Skin has a large input field that can potentially be utilized by novel sensory feedback devices. It can be used to transfer the processed data to a human-in-loop agent such as a pilot or a driver. Alternatively, it can aid visually/hearing impaired people to perceive the environment through the skin with more details. Entertainment and gaming systems are another potential example in which touch sensory feedback systems can be utilized.

Interactions between the touch-based sensory feedback device and other sensory modalities are crucially important and can affect the perception effectiveness. It specifically becomes important in vision-touch interactions, because vision is the most informative sensory in a wide range of applications, and therefore, the influence of vision on the perception of touch-based sensory feedback device should not be neglected. Moreover, interactions between vision and touch during the development and in adulthood was repeatedly reported in many studies (A Streri et al., 1993; Arlette Streri & Gentaz, 2004). Thus, a high probability of cross-modal visual-tactile interactions, effects, and side-effects is expected to be observed.

Aim of the present study

In the present thesis, we addressed cross-modal interactions between the visual and tactile modalities in adults with the focus on tactile information which delivers by a custom artificial device. Visual and tactile interactions are specifically important since tactile is known as the main source which visual sensory calibrates with and learns from during the development (Gori, Giuliana, Alessandra, Sandini, & Burr, 2010; Gori, Giuliana, Sandini, & Burr, 2012). Furthermore, activation of visual cortical area during touch perception (Sathian et al., n.d.) bolds the importance of visual-tactile interactions. Therefore, cross-modal effect of visual and tactile modalities is expected to be relatively strong (Gentile, Petkova, & Ehrsson, 2011; Keetels & Vroomen, 2008; Ossandón, König, & Heed, 2015; Salzer, Aisenberg, Oron-Gilad, & Henik, 2014).

Touch-vision interactions can be categorized into active or passive interactions, Figure 1. In the present work, active interactions refer to ones that involve some kind of perceptual learning.

Perceptual learning can be an explicit cross-modal learning which is designed and encoded in the experiment instructions. Alternatively, it can be an implicit learning, like calibration, which emerges during the cross-modal interactions. Passive interactions refer to those tasks in which no or very little (non-significant) perceptual learning involve.

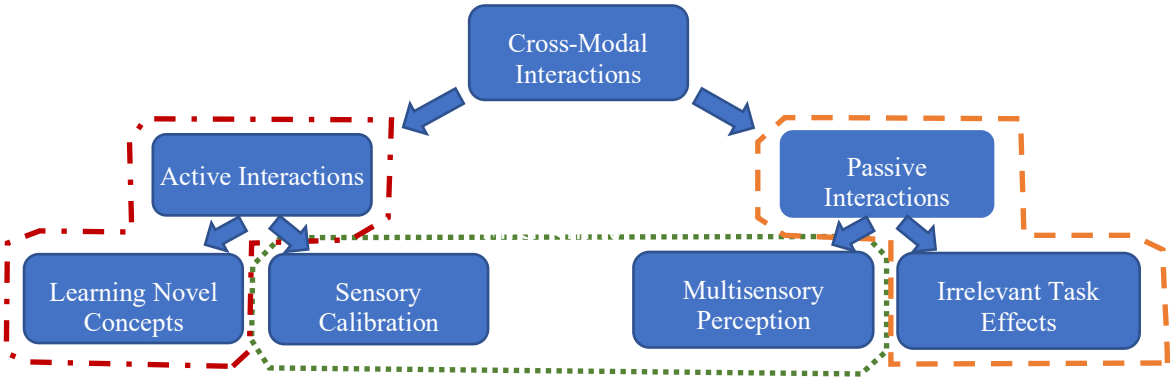

Figure 1. Scope of cross modal interactions in different studies. In the passive interactions, no explicit learning process was involved, while in the active interactions, a learning effect was hypothesized and evaluated in the study.

In the present study, we tried to cover both active and passive interactions by three experiments: multisensory perception (integration and calibration) in an environment of varying reliability, task-irrelevant cross-modal effect of tactile information on visual perception, and finally learnability and integration of novel tactile sensory experiences with existing visual sensory information. For each case a separate experiment was designed, and sufficient experimental data was collected. We statistically analyzed the experimental data and provided various computational models that explain the mechanism underlying the correspondence perception behavior.

In the first study, we investigated the overall visual-tactile perception in a realistic environment of varying reliability information. The goal of this study was to study the multisensory perception in a realistic multimodal environment which covers the passive aspects of interactions across sensory modalities. However, we expect to observe a cross-modal calibration or adaption due to the discrepancy in the information. Therefore, the first study was designed to cover both passive and potential active interactions, Figure 1, green section. In the first part of the experiment, we studied the multisensory perception of visual and auditory information with a discrepancy among visual and auditory information. In the second part of the experiment, we added the tactile information to investigate the multisensory perception in presence of three modalities. The reliability of information varied in the visual and tactile modalities, allows us to study the effect of reliability condition on the perception.

In addition to the experimental results, rich simulations and modeling were provided that shed light on the mechanism and processes undelaying the experimental results. The multisensory perception models are usually proposed for up to two modalities. However, we provided an extension of multisensory models for three (and more) modalities to account for our experimental results.

Covering both active and passive interactions might be closed to the real scenarios, and enables us to better generalize the results to daily perception activities. However, there are also important passive-only or active-only scenarios which we addressed in this thesis in two separate experiments.

In the second study, we investigated the visual-tactile conflict task and proposed a multimodal diffusion model for conflict tasks (DMC). There are different types of conflict tasks such as Stroop task, Eriksen-Flanker task, and the Simon task. The main focus of the second study is a visual-tactile version of Simon task. In a traditional Simon task, the people are instructed to respond to a non-spatial attribute such as color or letter with a spatially defined target (usually left or right hand key press). Although the spatial information is task-irrelevant, responses are faster and more accurate when both the stimulus and the response are on the same spatial side (congruent condition) rather than on different sides (incongruent condition). This study is actually a passive-only experiment in which the processing of simultaneous visual and tactile information is investigated. This study covers two important aspects: 1- whether and how strong is the effect of task-irrelevant tactile information (and also task-irrelevant visual information) on a visual perception task? 2- How a multimodal DMC model can address the cross-modal relations in a multi-modal conflict task?

Unimodal DMC model suggests two parallel processes in a conflict task, an automatic process and a control process. The task-irrelevant information trigs an automatic process, while the control process accumulates task-relevant evidences. The DMC model shows that the decision process can be simply predicted by the superimposition of the automatic and the control process. In the second study, an extension of DMC model for multi-modal stimuli was proposed and a multimodal DMC with the same principles was suggested and fitted to the experimental results.

The main focus in the final study of the thesis is the active interactions across visual and tactile modalities. Specifically, we investigated the explicit cross-modal learning of visual and tactile information. In this study, a custom vibrotactile device was utilized to provide novel tactile

experiences (through novel tactile patterns) to participants. This study would address whether and how an artificial sensory feedback device can be integrated to the sensory and perceptual system of humans. Furthermore, it provides insights on designing and instructions of artificial sensory devices for real world applications.

In the learning phase of the study, participants were exposed to simultaneous visual and tactile information and they were asked to learn the association between novel tactile stimuli and visual stimuli. In the evaluation phase, participants utilized the learned associations to respond to multi-modal or unimodal visual/tactile stimuli with different reliability levels. There were seven blocks of training-evaluation and the learning curve for different reliability conditions were statistically analyzed. We discuss and analyzed the integration-while-learning behavior for different reliability conditions. Furthermore, we provided two computational models that explains the integration of the artificial sensory device. One model is proposed based on the fusion of decisions and the alternative model is aligned with Bayesian integration of sensory information.

Multisensory perception in an environment of varying reliability [Study 1]

The first part of the present thesis addresses the multisensory perception of contradictory information from visual, tactile and auditory modalities in an environment of varying reliability. It is a general consensus when the information of various sensory modalities are congruent and synchronous (Atteveldt, Formisano, …, & 2006, n.d.; Navarra, Vatakis, Zampini, …, & 2005, n.d.) we benefit from a decrease in response time (Drugowitsch et al., 2014; Rowland et al., 2007) an increase in reliability, and an increase in accuracy (Drugowitsch, DeAngelis, Angelaki, & Pouget, 2015; M. O. Ernst & Banks, 2002; Pouget, Beck, Ma, & Latham, 2013). Nevertheless, the incongruent information across different sensory modalities, may lead to a biased percept and more complex scenarios.

It is not surprise if information from an artificial sensory feedback device is not well calibrated with other sensory modalities. It can specially happen at the beginning, when the sensory feedback device delivers new sensory experiences to users. Therefore, information from all sensory modalities, including sensory feedback device, is incongruent and maybe uncalibrated.

Moreover, relative reliability of information across sensory modalities is a crucial factor in multisensory perception as well as multisensory calibration. Different reliability conditions might result in various perceptual behaviors and even some behaviors might emerge only in an environment of varying reliability. In order to understand mentioned cognitive aspects of an artificial sensory device, multisensory perception of incongruent information in an environment of varying reliability was studied.

Multisensory causal inference and multisensory calibration

Multisensory causal inference is an important sub process of perception specially in case of a discrepancy among the modalities. It predicts whether sensory inputs originate from the same cause or different ones (M. Ernst, integration, & 2011, n.d.; Shams & Beierholm, 2010). It has been reported that multimodal causal inference depends on temporal, spatial, contextual features of the stimuli (Woods, Lehet, & Chatterjee, 2012), and also prior knowledge and experiences (Roach, Heron, & McGraw, 2006; Vision & 2007, n.d.). The multisensory causal

inference usually conforms Bayesian rule (Körding et al., 2007), and aimed to minimize the error between true signal and the estimated one by perceptual system.

In case of the inconsistent sensory information, the brain should find out whether the inconsistency is due to a systematic error in perceptual system or just inherited noise from the sensory systems. Some studies showed if the discrepancy is due to a systematic error, multisensory calibration can resolve the persisting discrepancy. In such case, cross-modal calibration increases the accuracy of the multisensory perception. However, conditions in which calibration takes place is not crystal clear. Indeed, multisensory causal inference and multisensory calibration processes are expected to be coupled and should be analyzed together (Körding et al., 2007). Only if signals are received from a same occasion, then calibration of those signals, with respect to each other, is reasonable.

Interactions between multisensory integration process and cross-sensory calibration process is crucial to understand the mechanisms underlying the optimal multisensory perception. This becomes more relevant for real world applications which are often complex, with dynamic flow of information. Dynamic nature of real-world application might reveal behaviors which cannot be captured easily in restricted experimental environment. In order to reduce the gap between the experimental environment and a real environment, we should consider the effect of varying reliability sensory information on perceptual processes such as causal inference and calibration. This is particularly informative to find the required conditions for the calibration process and to study the elasticity of the multisensory calibration. Multisensory causal inference and cross-modal calibration benefit from a history of observations (Körding et al., 2007; Vision & 2007, n.d.). We also plan to investigate the behavior of causal inference in an environment of varying reliability. Specifically, we want to study whether causal inference is independently processed for each reliability condition or whether the causal inference process considers all reliability conditions together.

Varying reliability environment and perceptual strategies

One of the goals we kept in mind during the experiment design was to reduce the gap between the experiment setup and the real-world conditions. Therefore, we designed the experiment in a way to include multiple reliability conditions and various modalities. We divided the experiment into three sections: visual-auditory section (Experiment 1.a), visual-tactile-auditory section (Experiment 1.b), and unimodal section (Experiment 1.c). We introduced various

reliability conditions for multisensory sections 1.b and 1.c in order to make the experiment more realistic. The trials of different reliability conditions were presented in random order. Figure 2 depicts the multisensory sections of the experiment and reliability conditions. Participants received two consecutive stimuli, and the task for participants was to choose whether the second stimulus was on the left or on the right side of the first stimulus.

By embedding a cue conflict between different modalities, we investigated whether and how different modalities interacts with each other to solve the conflict and provide a coherent perception across multiple modalities. Visual-auditory stimuli had a spatial discrepancy with a conflict angle (Δ) of $-4°$ in the experiment 1.a. Tactile vibration was also delivered 4 cm to the left of the audiovisual stimuli in experiment 2.a.

We think that the cue-conflict was not too large, and therefore, participants would consider a common cause for information from different modalities at the beginning and would integrate the information. Over the course of the experiment, if participants find out the systematic repetitive discrepancy, they might alter their perceptual strategy. We expected three different perceptual strategy for the second half of experiments: (1) Participants do not capture the repetitive systematic discrepancy and therefore, continue to integrate the information. (2) Participants capture the systematic discrepancy but still infer the same cause for information from different modalities. They would go for a calibration strategy to resolve or reduce the discrepancy. (3) Participants capture the systematic discrepancy and infer different causes for information. In this case they would select the most reliable source of information.

We have tested these hypotheses for different reliability conditions which we introduced in the experiment. Figure 2 shows the design of artificial sensory feedback device and the experiment procedure. We also modelled the experimental data using multisensory calibration and the multisensory causal inference models and some simulations.

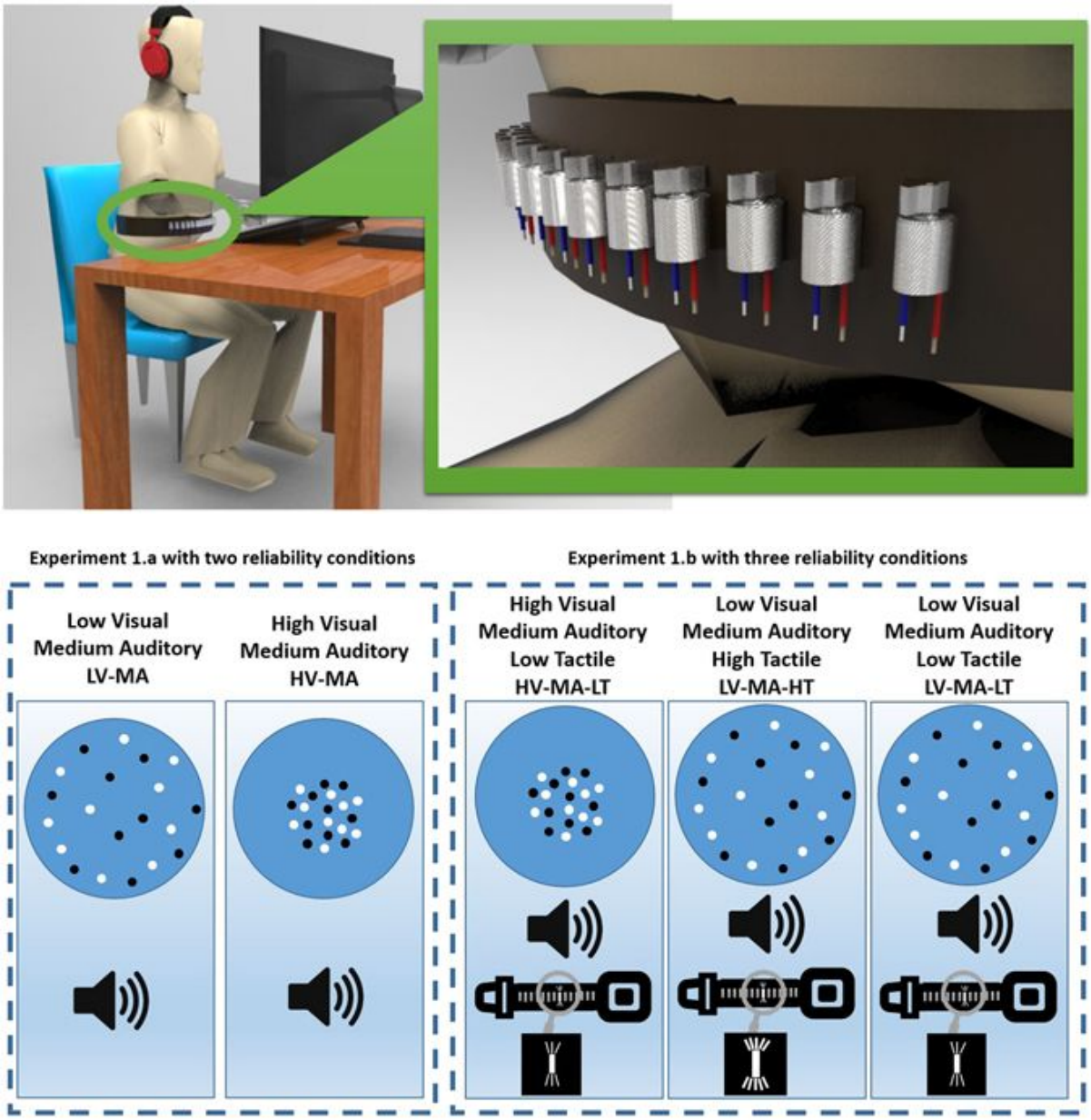

Figure 2. Experiment and stimuli design. Top part of the figure shows the artificial sensory feedback device for the experiment while the bottom part depicts the experiment procedure and reliability conditions. Figure is reprinted from (M.-A. N. Mahani et al., 2017).

Results and discussion

The alteration of perceptual strategy can be assessed through analyses of point of subjective equality (PSE) between the first half and the second half of experiments. Shift of PSE towards visual, auditory, or tactile modalities indicates weighting of each modality in the integration. The data has been analyzed based on individual psychometric functions, which were fitted to for each reliability condition.

In experiment 1.a, subjects received stimuli with two reliability conditions:

- HV-MA: High reliability Visual and Medium reliability Auditory stimuli.
- LV-MA: Low reliability Visual and Medium reliability Auditory.

All trials of two conditions were randomly mixed. We found out that the spatial perception was significantly shifted towards the auditory source in the second half of the LV-MA condition. However, the PSE in the HV-MA condition remained close to the visual stimuli. It suggests that changes in perceptual strategy are specific to different reliability conditions. Moreover, participants reported lower confidence for the LV-MA condition in compare to HV-MA. This can be considered as an evidence for the conscious perception.

We extended the first experiment by adding the tactile stimuli in the second experiment and providing tri-modality information to participants. Obviously, the reliability conditions were also extended to three cases:

HV-MA-LT: High rel. Visual, Medium rel. Auditory, and Low rel. Tactile condition.

LV-MA-HT: Low rel. Visual, Medium rel. Auditory, High rel. Tactile condition.

LV-MA-LT: Low rel. Visual, Medium rel. Auditory, Low rel. Tactile condition.

The PSE was shifted towards tactile in the LV-MA-HT condition and therefore, showed a change in multisensory perception. A similar change was observed in HV-MA-LT condition. No significant change of PSE was observed for the LV-MA-LT condition, probably because the auditory was the most reliable modality in LV-MA-LT reliability condition and was not altered by touch/vision. The slope analyses of the unimodal Experiment (Experiment 1.c) supports this interpretation, see appendix paper for details.

All findings were confirmed statistically, and details are provided in the paper, please see the appendix. Underlying processes of these behavior were further studied by modeling and simulation. We proposed an ideal observer model and performed sophisticated simulations. We tried to provide grounds for various perceptual strategies by jointly studying the causal inference and calibration processes. The model is designed to minimize the overall perception error across all reliability conditions.

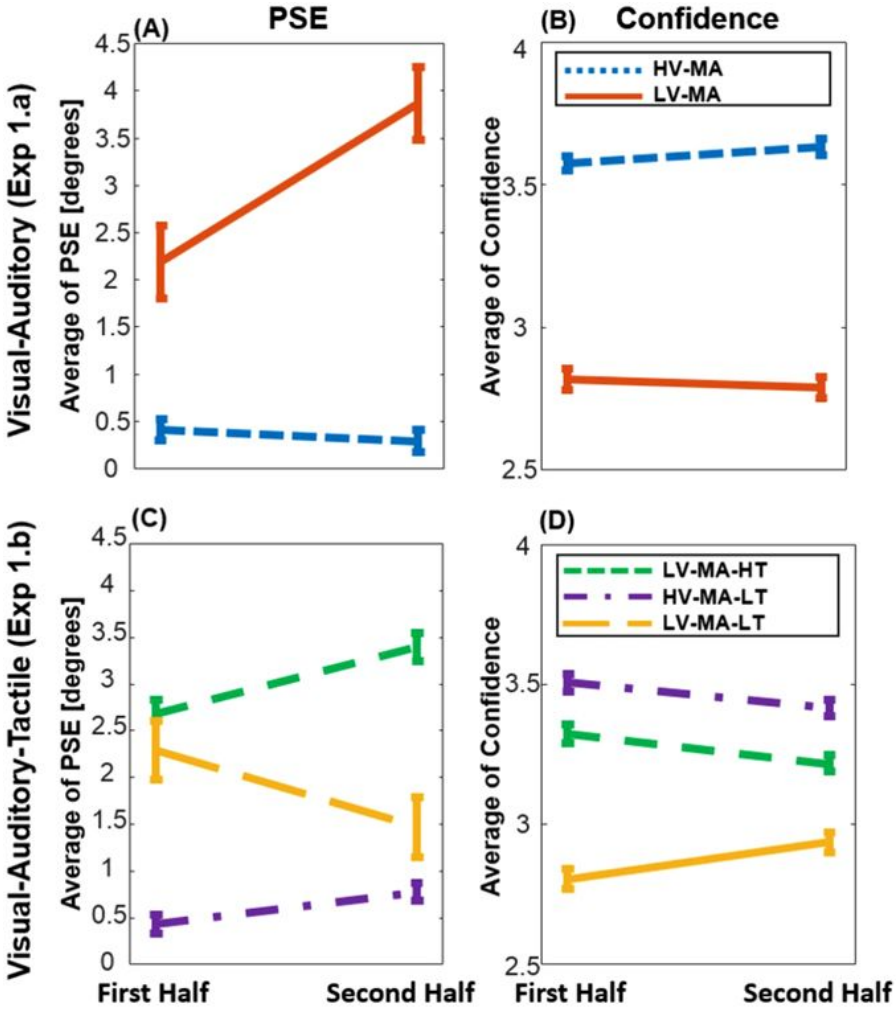

Figure 3. PSE variation and confidence changes between the first and the second half of experiments. (A) and (B) show the PSE alterations while (C) and (D) depict the confidence changes. See paper for more details. Figure is reprinted from (M.-A. N. Mahani et al., 2017)

Perceptual conflicts are the source of some misperceptions, however, they can also initiate perceptual learning processes in the brain. Some studies reported that cross-modal calibration processes initiated conflicts (Van der Burg, Alais, & Cass, 2013; Wozny, Neuroscience, & 2011, n.d.; Zwiers, Opstal, neuroscience, & 2003, n.d.).

We simulated the perception in the presence of conflict stimuli in an environment of varying reliability. Specifically, we simulated three possible multimodal perception strategies and mechanisms which could potentially account for PSE shifts: Collaborative calibration, Modality Dominant (MD) calibration, and selection, please see Figure 4.

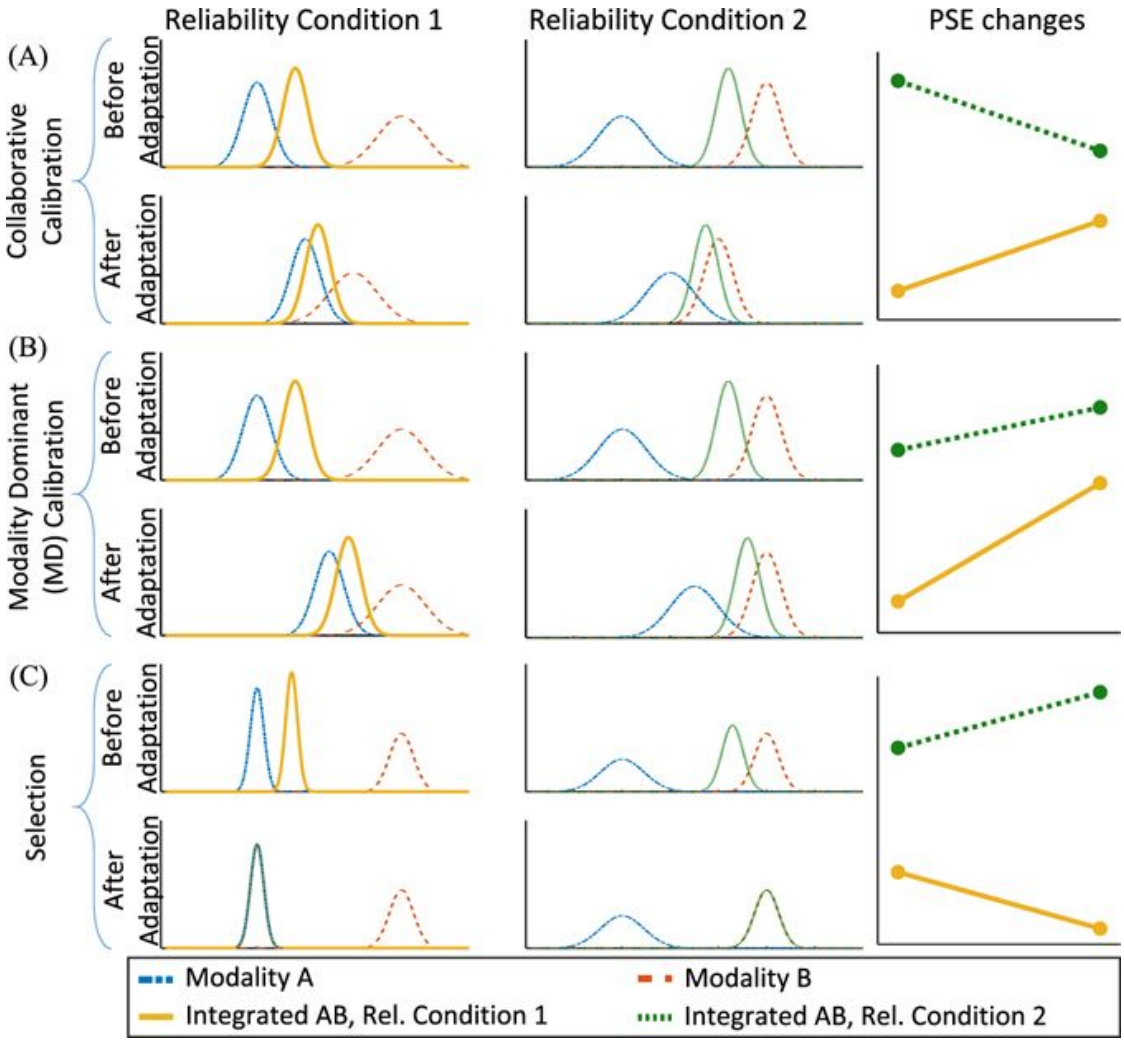

Figure 4. Simulation of different perception strategies under different reliability conditions. Figure is reprinted from (M.-A. N. Mahani et al., 2017)

We can shed light on perceptual strategies, which were taken by participants, by corresponding the PSE shifts between simulations and experimental observations.

By associating the result of Experiment 1.a with the possible simulated strategies, it seems that the perceptual strategy in this section is shifted from integration toward selection strategy; PSEs shifted toward the most reliable source of information. This indicates that, even if participants start with integration behavior, calibration is not the only strategy for resolving/reducing the cue conflict. Actually, under certain reliability conditions, participants might prefer taking the selection strategy instead of performing a cross-modal calibration. We have further investigated the mechanism underlaying this behavior by modeling the causal inference in varying reliability conditions. Please see the paper for more details.

Associating the result of experiment 1.b with the possible simulated strategies showed that participants took MD calibration strategy to resolve/reduce the cue conflict in this section of

the experiment. Specifically, the result revealed that touch calibrated vision, and therefore visual spatial perception was shifted toward the source of tactile information. That is why even the high reliability visual condition (HV-MA-LT) shifted toward the tactile source of information. This type of dominant calibration is also consistent with previous reports (Gori, Sciutti, Burr, & Sandini, 2011).

We provided causal inference modelling and quantitative comparison of all models with experimental results in appendix.

Multimodal Simon effect: A Multimodal Extension of the Diffusion Model for Conflict Tasks [Study 2]

In our daily life, we sometimes suppress irrelevant information and focus only on task-relevant information to improve our efficiency. These situations, where there are task-irrelevant and task-relevant information at the same time, are addressed as conflict tasks. There are several well-known and standard examples for conflict tasks such as the Simon task and the Stroop task (J Richard Simon & Wolf, 1963; Stroop, 1935). In our second study we focused on Simon task and investigated the passive interactions of visual-tactile and visual-auditory information. Having both visual-tactile and visual-auditory experiments would allow us to compare the multisensory interactions for different combination of modalities. We can therefore investigate whether multi-modal interactions are subjected to specific combination of modalities or whether the same interaction behavior can be emerged from other combination of modalities.

In the standard Simon task, participants are asked to respond to a non-spatial attribute of the stimulus (e.g. color) with a spatially coded response key (e.g. left or right). In this experiment, spatial information is considered as task-irrelevant information and the non-spatial attribute would be the task-relevant information. Surprisingly, the findings showed that participants cannot suppress the task irrelevant information. Both the reaction time (RT) and the perceptual error can be influenced by task-irrelevant information. Specifically, when the visual stimulus is presented at the same spatial side as the response key, then the perception is faster and more accurate in comparison with having stimulus and response key on different spatial sides.

This finding is not only limited to visual perception, but also has been observed in other unimodal perceptions like auditory and tactile. Even the task irrelevant information from a modality can influence the perception of another modality. E.g. the auditory information can affect the reaction time of a visual perception task (Donohue, Appelbaum, Park, Roberts, & Woldorff, 2013; J R Simon & Craft, 1970). Similarly, a cross-modal effect of conflictual information has been observed for visual-tactile perception (Kennett, Eimer, Spence, & Driver, 2001; Poole, Couth, Gowen, Warren, & Poliakoff, 2015; Spence, Pavani, & Driver, 2004; Yue, Bischof, Zhou, Spence, & Röder, 2009). Despite of many researches, the findings are still limited to a single source of task irrelevant information; we addressed this limitation in our second part of the thesis.

Diffusion model for conflict tasks (DMC)

Computational models underlaying this behavior suggest that two separate processes acts simultaneously on input stimulus: a Controlled process and an automatic process. The controlled process is responsible for accumulating the task relevant information. Meanwhile, the automatic process works on task irrelevant information. In a recent study, an elaborated diffusion process model for conflict tasks (DMC) (Ulrich, Schroter, Leuthold, & Birngruber, 2015)was suggested. DMC is introduced on top of the standard diffusion models where a decision process accumulates noisy decision relevant information until one of two decision boundaries is hit (R Ratcliff, review, & 2004, n.d.; Roger Ratcliff, 1978; Stone, 1960). DMC extends this model by superimposing a second short-lived process, aka automatic process, for task-irrelevant information. The rest is the same, the superimposed activation accumulated both the controlled and automatic processes until it hits a decision boundary. DMC can reasonably predict the conflict tasks and has been successfully linked to neurophysiological findings (Servant, White, Montagnini, & Burle, 2016).

The present study extends the DMC to support the conflict tasks with two task-irrelevant information sources. Thus, we designed two Simon task experiments, a visual-tactile Simon task and a visual-auditory one. In both experiment both modalities provided task irrelevant information. The relevant task was the visual perception of non-spatial attribute, while simultaneous spatial task irrelevant information was presented from both modalities.

Experiments

As the main focus of the present thesis is visual-tactile interactions, we firstly extended the Simon task to a visual-tactile Simon Task. In this experiment task irrelevant information was delivered by tactile and visual modalities. The main task for participants was to respond to the non-spatial attribute of a visual stimulus with a left/right located key. More specific, participants had to respond to the stimulus letter “H” or “S” with left or right key. At the beginning of the experiment they received an instruction that explained them how left/right key is associated with “H” and “S” letter. The association was fixed for each participant during the experiment, but varied participants. The letter stimulus was accompanied by a vibrotactile stimulus. The location of both stimuli varied across three predefined positions on right, at center, and on left side; Combination of both stimuli and three locations resulted is nine congruency conditions. Please see Figure 5.

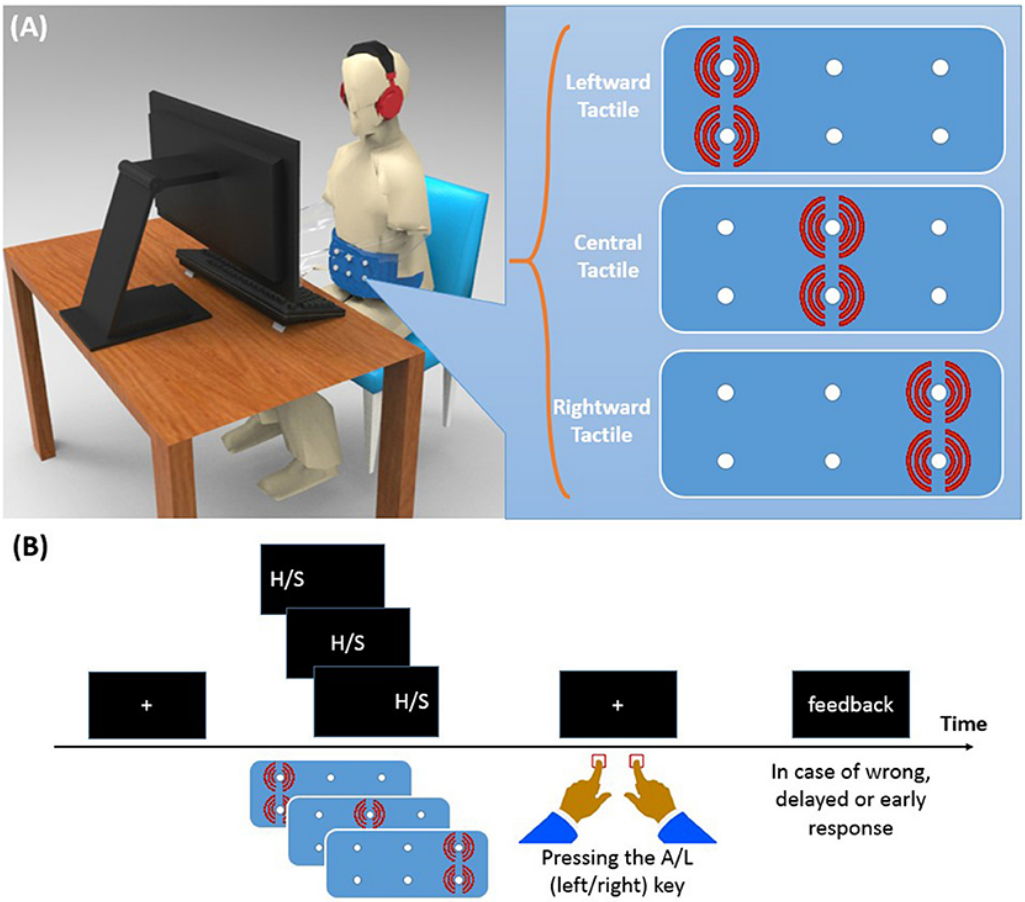

Figure 5. (A) Experimental setup for visual-tactile Simon task. (B) Time course of a trial. The figure is reprinted from (M.-A. N. Mahani, Bausenhart, Ahmadabadi, & Ulrich, 2019)

Figure 6 depicts the result of the both visual-tactile experiment (top row) and visual-auditory one (bottom row). RT was significantly affected by both visual congruency condition and tactile congruency condition. Surprisingly, out further analyses showed that the difference between the effect of visual neutral and the effect of visual congruent is not meaningful. Although tactile irrelevant information influenced the RT, there is clearly no difference between tactile neutral and tactile incongruent conditions.

As it can be seen in Figure 6, the effect of both visual congruency and tactile congruency on mean response error was significant. Similar to RT, the difference between the visual neutral and the visual congruent condition was not meaningful. But, in contrast to RT, the difference between tactile congruent and tactile neutral conditions was not meaningful. Please see the paper in the appendix for statistical details and analyses for these findings as well as for distributional reaction time and conditional accuracy functions. In general, the first experiment revealed that tactile information, even as a second source of irrelevant information, can alter the perception of visual information in terms of reaction time and response errors. However,

effect of tactile information was not exactly the same on RT and response error. We provided more insight in this regard in discussion and modeling sections, please see the appendix.

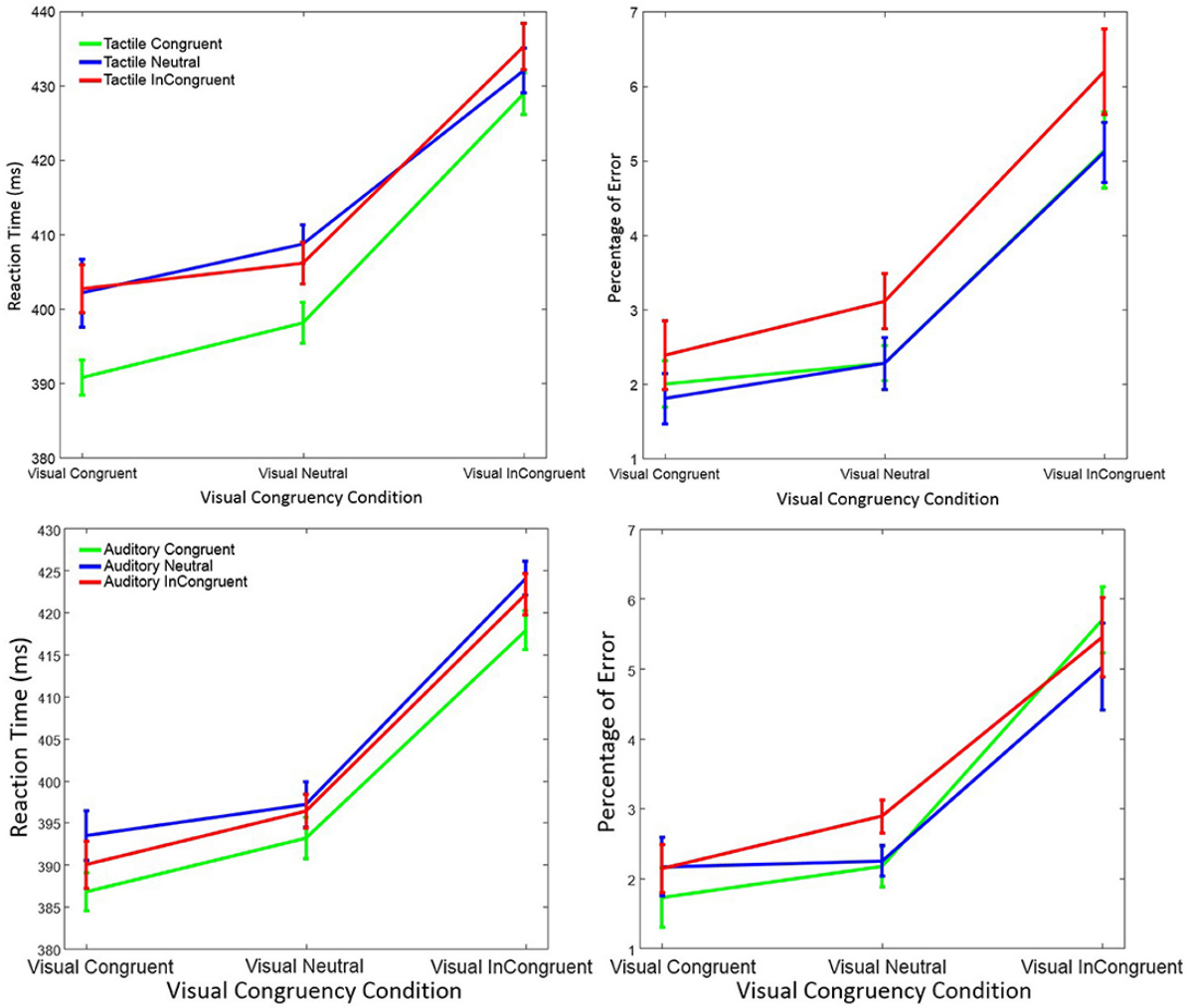

Figure 6. Reaction time (in ms) and mean percentage of response errors in Experiment 1 as a function of visual/tactile congruency. The figure is reprinted from (M.-A. N. Mahani et al., 2019)

We designed and conducted a second experiment by replacing the tactile modality with the auditory one. The second experiment was motivated by two factors: first, we wanted to investigated whether the effect of cross-modal task irrelevant information on visual perception is exclusively valid for tactile modality or it can be observed for other modalities as well. Literature supports the strong effect of touch on vision, but the effect of auditory on vision is not always strong and is rather limited to specific conditions. Secondly, the second experiment would enable us to evaluate our proposed computational model, multimodal DMC, for another combination of modalities rather than visual-tactile. Multimodal DMC is introduced later in this thesis and also in the paper in details. In the second experiment, the effect of task-irrelevant auditory stimulation on performance of visual perception was assessed. Similar to the first

experiment, the results for reaction time, response errors, distribution of the reaction time, and conditional accuracy functions were statistically analyzed and details are available in the paper.

Our analyses showed the effect of visual congruency on RT was again meaningful while auditory congruency had only a mild effect on RT. Similar to the first experiment, the difference between visual neutral and visual congruent conditions was neglectable. Analyses of response errors showed a significant effect of visual congruency on response errors. However, the effect of auditory congruency was not significant. In general, the results indicated a weak effect of task-irrelevant auditory information on visual performance in comparison with the effect of task-irrelevant tactile information. As it can be seen in Figure 6, tactile congruency conditions in visual-tactile experiment is clearly distinguishable. However, auditory congruency conditions have a huge overlap and it is hard to observe a clear effect of auditory congruency condition.

Considering both experiments together, task irrelevant visual information had a meaningful effect on both RT and response error of the visual perception. In both experiments and for both RT and response errors, the difference between visual congruent and visual neutral condition was not meaningful, but visual incongruent condition had a higher RT and a worse response error. This finding was robust and was replicated in both experiments. It has been reported that the visual perception benefits from faster retinal processing at the center of field of view (FOV), in comparison with a stimulus on the left or right side (Osaka, 1976). This might justify why the visual neutral and visual congruent conditions had similar effect. Visual neutral condition can benefit from faster retinal processing and results in similar effect as visual congruent condition. This assumption will be investigated in modelling section.

Nevertheless, tactile and auditory revealed various effects on visual perception. In general, tactile congruency effect on visual RT and response error was meaningful while auditory congruency had a weak effect on RT and no effect on response error. This is consistent with the previous findings which show the dominant of vision over auditory in processing the spatial information. All in all, we can order the effectiveness of various irrelevant information on visual perception as follow: Visual irrelevant information > Tactile irrelevant information >> auditory irrelevant information.

MultiModal DMC model

The multimodal DMC is proposed based on the DMC (Ulrich et al., 2015). In accordance to the DMC, we considered the total reaction time as sum of a decision process (D) and a residual process (R), $RT = D+R$. Residual process is all other processes rather than the decision process.

DMC models the decision process as a standard Wiener diffusion process. At each time stamp t, the decision process, $X(t)$, is defined as superimposition of a controlled process, $X_c(t)$, and an automatic process, $X_a(t)$. DMC assumes that the controlled process and the automatic process are processed independently and in parallel. Similar to the original diffusion model, the decision process accumulates the information until it reaches one of the decision boundaries, $-b$ or b. We formulated the controlled process, $X_c(t)$, at each time stamp t as the following differential equation:

$$X_c(t + \Delta t) = X_c(t) + \mu_c(t) \cdot \Delta t + W_c(t) \cdot \sigma_c \cdot \sqrt{\Delta t} \quad (1)$$

$W_c(t)$ is the standard Wiener diffusion process with mean = 0, and variance = 1.0, and σ_c is the diffusion constant. $\mu_c(t)$ denotes the drift rate of the controlled process, $\mu_c(t) = \mu_c$. We similarly formulated the automatic process as follow:

$$X_a(t + \Delta t) = X_a(t) + \mu_a(t) \cdot \Delta t + W_a(t) \cdot \sigma_a \cdot \sqrt{\Delta t} \quad (2)$$

The time course of an automatic process, $Xa(t)$, is modeled as a pulse-like rescaled Gamma distribution with fixed shape parameter $a = 2$ and the free scale parameter τ. Please refer to the appendix for more details.

The original DMC assumes the decision process as sum of only one automatic process and one controlled process, $X(t) = X_c(t)+X_a(t)$. Multimodal DMC (MDMC) extends the DMC by making the genuine assumption of DMC more general. Specifically, MDMC allows the superimposition of more automatic processes and consider each automatic process as an independent and parallel process to the others, $X(t) = X_c(t)+X_{a1}(t)+X_{a2}(t)$. Of course this assumption was evaluated by the two experiments which were explained in previous section.

Figure 7 exemplifies how MDMC considers the interactions of a control and two automatic processes. In this example the expected values of the decision process $E[X(t)]$ (blue line), the controlled process $E[X_c(t)]$ (red line), and two automatic processes $E[X_{a1}(t)]$, $E[X_{a2}(t)]$ (black and green lines) are plotted. The effect of different combination of congruency conditions on decision process is depicted in four cases. We also assumed that the neutral automatic process does not affect the decision process.

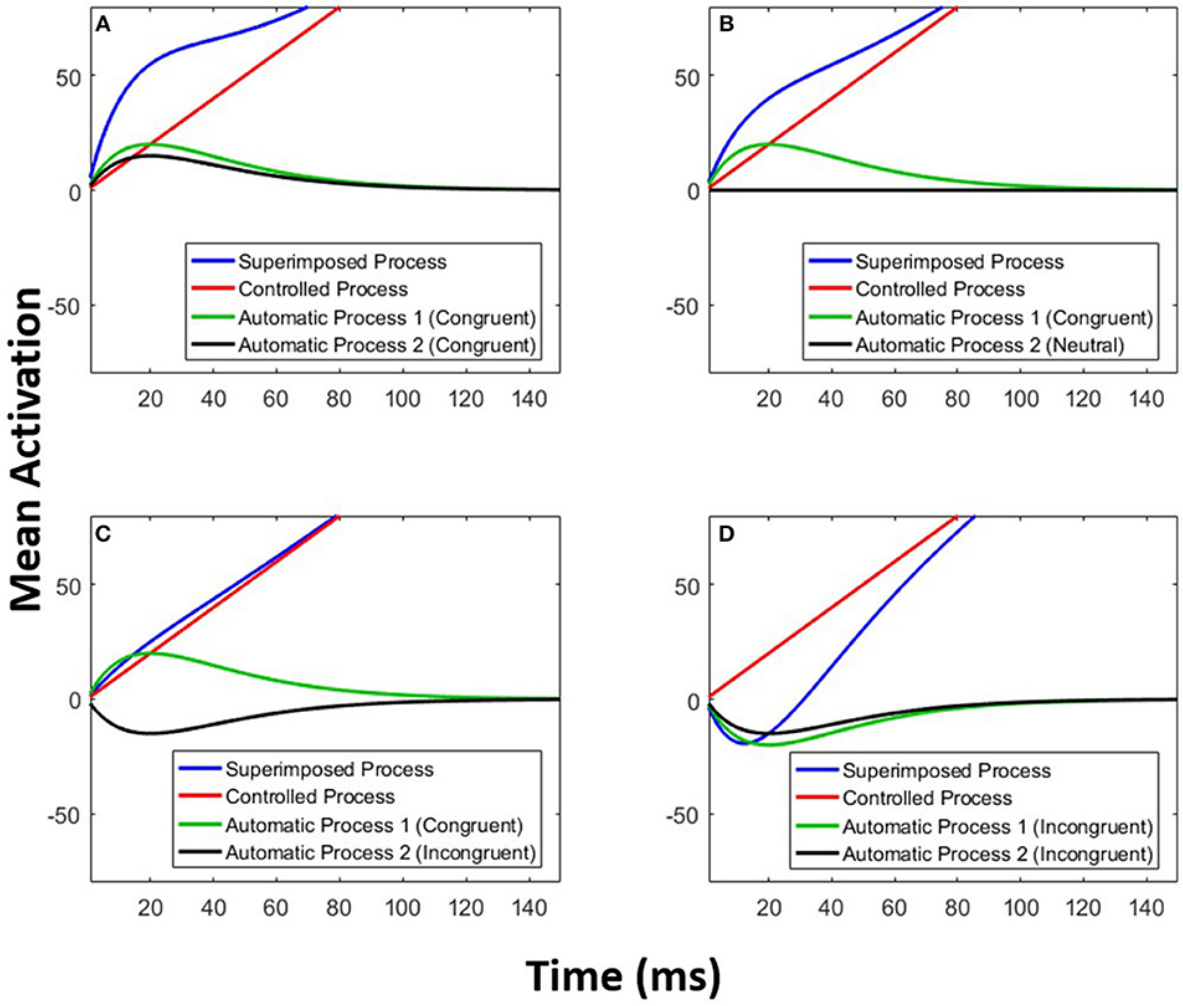

Figure 7. An exemplary illustration of multimodal DMC. The figure is reprinted from .

As it was mentioned in previous section, the visual neutral condition might benefit from a faster retinal processing. Therefore, we considered two variants of MDMC, the first one is the genuine MDMC, as is described so far. The second variant considers a faster process of information for neutral visual condition. We named the second variant as FN-MDMC (Faster Neutral Visual Multimodal DMC).

MDMC was fitted using the similar method described in (Hübner, 2014) and also (Servant et al., 2016). Since the data of individual participants are typically noisy, we fitted the model to the averaged data of all participants and consider this fitting for further analyses. However, the results of individual model fits are available in the supplementary document of the paper.

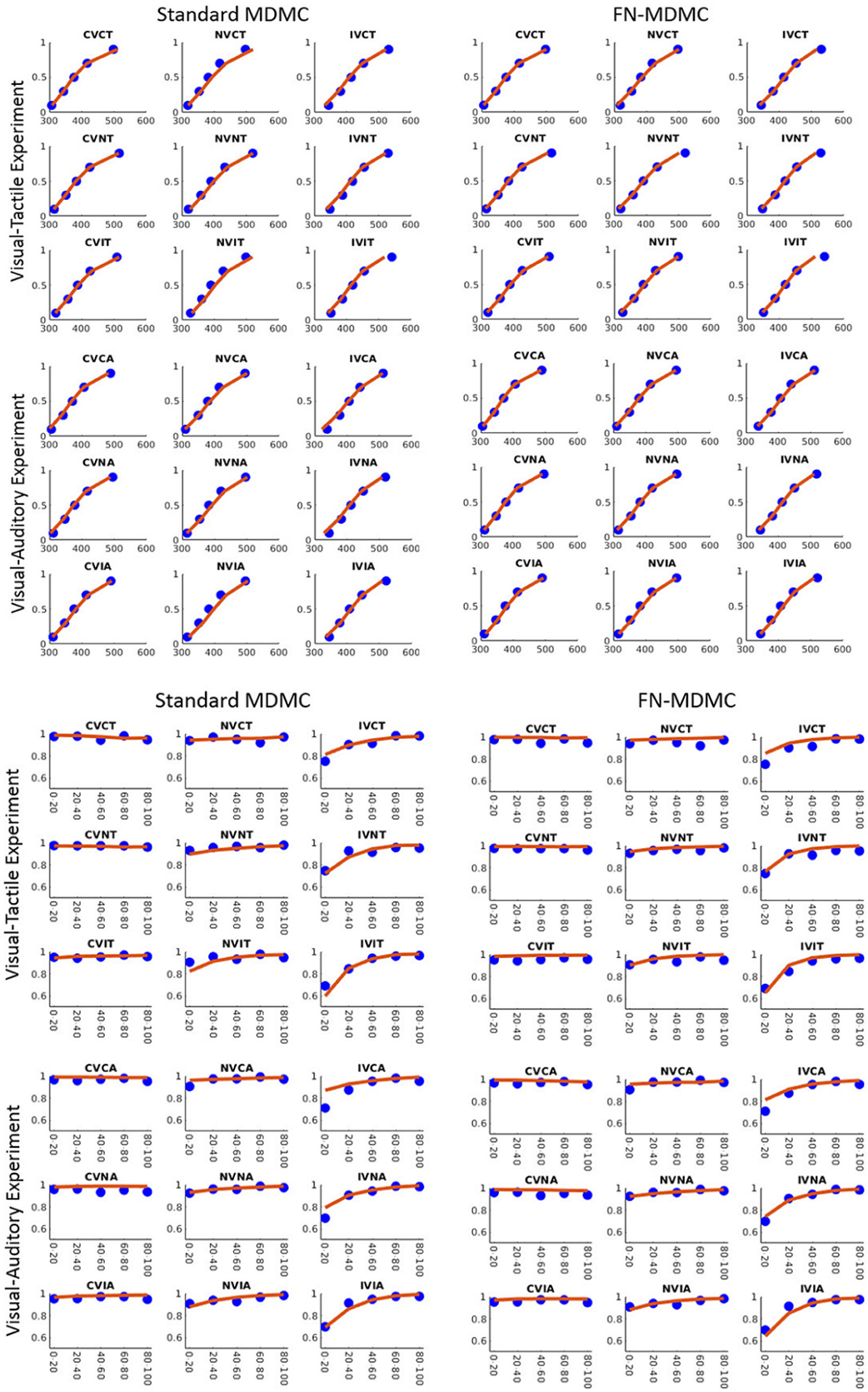

Figure 8. Modelling result of CDFs and CAFs for both experiments. Red lines show the model predictions while the blue dots are the experimental data. The figure is reprinted from (M.-A. N. Mahani et al., 2019)

Figure 8 depicts the results for both cumulative distribution functions (CDFs) and Conditional accuracy functions (CAFs). Our genuine MDMC fits reasonably the experimental data, however, FN-MDMC provides slightly better fits than MDMC. In a Simon task, usually we observe more errors for faster RTs, and it can be also observed in the incongruent visual conditions in our experiments. Modelling results confirm that MDMC could capture this pattern relatively well. The fits of MDMC and FN-MDMC were compared using the paired-sample permutation test across simulated G^2 and BIC values. Please see the paper and appendix for the estimated parameters and model comparison.

Discussion

The present study extends the classical Simon task by considering an additional source of task irrelevant information from a second modality; We studied the effect of additional tactile irrelevant information in Experiment 1 and additional auditory irrelevant information in the Experiment 2. Both experiments were theoretically motivated by MDMC, an extended version of DMC. MDMC keeps the genuine assumption of DMC, which is the independent process of information. i.e. MDMC assumes that the irrelevant information from various modalities are independently processed and therefore, the effect of one modality does not influence the effect of another modality. MDMC could reasonably predict the experimental results in both visual-tactile and visual-auditory experiments.

Similar to the previous studies, the results of both experiments revealed the meaningful effect of task irrelevant visual information on RT and response errors, like as in the classical Simon effect. Beside the effect of task irrelevant visual information, we also studied the effect of second task irrelevant information. Our findings of the first experiment showed that tactile irrelevant information could meaningfully affect the RT and response error of visual perception. However, findings of the second experiment revealed that auditory irrelevant information, in presence of the visual irrelevant information, had a weak effect on RT and no effect on response error of visual perception. Considering all results together, one can conclude that the visual irrelevant information had a strong effect, tactile irrelevant information had a mild effect, and auditory irrelevant information had almost no meaningful effect on processing of task relevant visual information.

Although many studies reported the effects of task-irrelevant information on non-visual decisions, to the best of our knowledge, no one studied the effect of simultaneous task irrelevant

information from both tactile and visual sources on non spatial visual perception. The result of the first experiment was consistent with previous studies and our expectations, since the strong cross-modal interactions of touch on vision was already reported in many studies. However, we were not expected such weak effect of irrelevant auditory information. Indeed, some of the previous studies reported the meaningful effect of irrelevant auditory information on visual perception in the context of Simon task (J R Simon & Craft, 1970). Nevertheless, in that experiment auditory was the only one source of irrelevant information, whereas in our setup, we had two source of irrelevant information, auditory and visual. By looking at the studies in which simultaneous spatial visual and spatial auditory information are processed, we came across similar and consistent findings (Bertelson & Radeau, 1981; Howard & Templeton, 1966; Welch & Warren, 1980). In fact, the effect of auditory irrelevant information on visual perception is expected to be very weak, when it is accompanied by irrelevant visual information.

We proposed and fitted the MDMC, an extension of DMC, to the experimental data. Despite genuine MDMC fitted the observed RT data and response errors reasonably, it was suboptimal with regard to the neutral conditions, probably because the model did not support the contribution of faster foveal processing for neutral conditions. Therefore, the model fit was improved by the FN-MDMC, which considers a potential speedup for central (neutral) visual stimuli. FN-MDMC addresses differences in processing latency/duration for different stimulus locations, but retains our genuine assumption that automatic activation from multiple task irrelevant source of information may act independently and overlap controlled process.

Learning how to integrate an artificial sensory device [Study 3]

Introduction

In everyday life, we perceive our complex environment through our multisensory input. The lack or degradation of inputs can significantly decrease the accuracy of perception. Artificial sensory devices are designed to partially compensate the lack of a sensory modality (Abboud, Hanassy, Levy-Tzedek, Maidenbaum, & Amedi, 2014; Maidenbaum, Abboud, & Amedi, 2014) , or to improve perception by providing complementary and processed information (Shull & Damian, 2015). Therefore, it is important to investigate whether and how an artificially sensory device can be learned and integrate to our multisensory perception system. In the last study of the present thesis, we explicitly imposed a cross-modal visual-tactile learning and investigated how integration behavior might evolve over the course of learning. We designed and utilized a custom artificial sensory device to present vibrotactile stimuli to participants. In general, this experiment covers mainly the active interactions of visual and tactile modalities and completes the sequence of studies we considered for the thesis.

Invasive artificial sensory systems affect directly the neuronal system (Collins et al., 2017) and lead to optimal integration of multisensory information (Dadarlat, O'doherty, & Sabes, 2015). However, non-invasive sensory devices are a better alternative because they are well-developed and more appropriate (and affordable) for realistic applications. Even though, many studies investigated the technological aspects of non-invasive devices such as wearable devices (Iqbal, Aydin, Brunckhorst, Dasgupta, & Ahmed, 2016; Mukhopadhyay, 2015; Son et al., 2014), the cognitive aspects are less studied. Particularly, it is yet open how we utilize input from a non invasive artificial sensory system and if we can integrate it into our multisensory perception.

Many studies have been previously reported that humans optimally integrate multiple sensory inputs, that leads to a significant increase in accuracy and reliability of perception (Butler, Smith, Campos, & Bülthoff, 2010; Drugowitsch, DeAngelis, Angelaki, Elife, et al., 2015; M. Ernst, Nature, & 2002, n.d.; Pouget et al., 2013). The majority of studies have assessed the multisensory integration only for well-experienced sensory information. Among the rare studies which investigated the novel artificial sensory devices, Dadarlat et al. showed that

unfamiliar multichannel intracortical microstimulation (ICMS) signals and proprioceptive input could optimally get integrated in monkeys. However, this issue has not been addressed for non-invasive artificial devices.

We addressed this question by investigating the cross-modal learning and integration of a custom designed wearable device and visual information. Figure 9 depicts the procedure and setup of our experiment. We designed the visual stimulus as a set of motion dots and the tactile stimulus was a unique pattern of vibrations. The experiment had seven blocks of training and evaluation.

Each block of the experiment starts with a training phase. In each training trial, a set of moving dots and a unique vibro-tactile pattern were presented to participants. Tactile patterns were free of any directional movement but were associated to a specific direction of dot motion during the course of experiment. Participants were explicitly asked to learn the associations as best as they could. In each trial of the evaluation phase, either a unimodal visual/tactile stimulus, or a multimodal stimulus was presented. Participants had to perceive and decide on the direction of movement from either the motion dots, or the associated tactile stimulation, or both stimuli in case of the multimodal stimuli. They also reported their confidence in their decisions. The visual stimulus had a constant reliability only in training phase whereas the reliability of tactile stimulation was constant in both training and evaluation phases. In the evaluation phase, we controlled the reliability of visual stimulus in three levels of low, medium and high for both uni-visual and visual-tactile conditions.

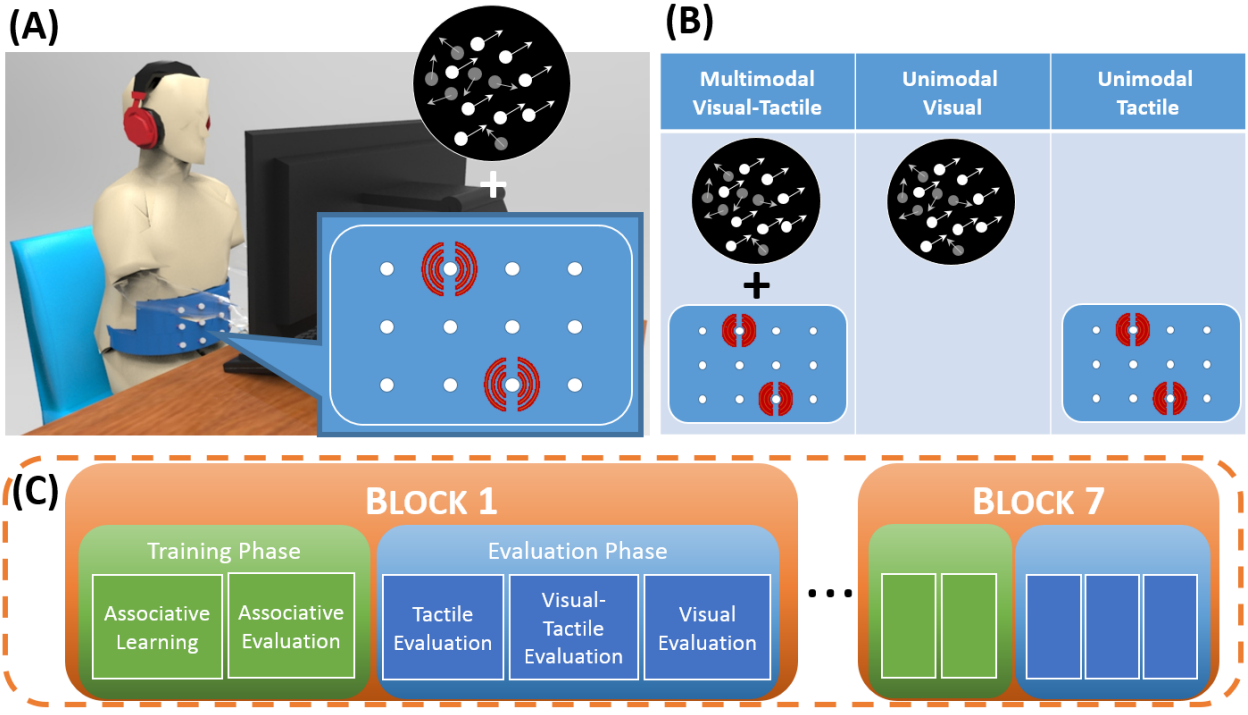

Figure 9. Experimental setup, stimuli design, and learning procedure of the third study. This figure is reprinted from (M. N. Mahani, Bausenhart, & Ulrich, 2020)

Experimental Results

We analyzed the accuracy of perception in all reliability conditions across all learning blocks (Figure 10). The results revealed that the novel vibrotactile patterns were efficiently learned and associated with the moving direction of corresponding visual stimuli.

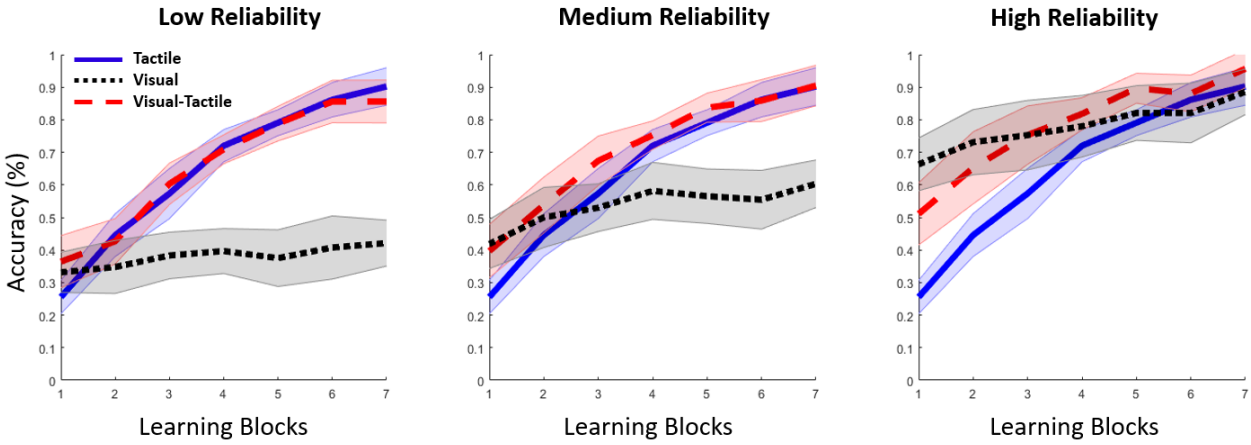

Figure 10. Perception accuracy in all reliability conditions over the learning process. Tactile is replotted for the sake of clarity. The figure is reprinted from (M. N. Mahani et al., 2020)

Our statistical analyses showed the significant effect of learning block and reliability condition on accuracy of perception (see paper for more details). This means first our control of reliability was successful and secondly learning happened over the course of blocks. To understand further the learning of the novel artificial sensor and its possible integration with our visual perception, we statistically analyzed each reliability condition. For the low reliability condition, the effect of both modality and block was significant. The results also showed the dominance of tactile stimulation in visual-tactile perception, which was expected because the reliability of visual information is very low in this condition. For the medium reliability condition, similar to the low reliability condition, the effect of both modality and block was significant. However, our further analyses of medium reliability condition revealed an interesting finding. Perceptual accuracy of visual-tactile exceeded the accuracy of both unimodal stimuli in some of the blocks. This points to a potential interaction of visual and tactile information for boosting the perception, even from the early blocks.

The accuracy of visual-tactile perception in high reliability condition, shows clearly the mutual contribution of both visual and tactile information. It is obvious by the intermediate accuracy of visual-tactile perception, which lies between the accuracies of visual and tactile unimodal perceptions. This finding, surprisingly, suggests that the visual and tactile information are likely integrated from the beginning of learning, despite of the decreased performance in comparison with the unimodal visual condition. See appendix for post-hoc and statistical tests.

Modeling

In order to understand the mechanisms behind the learning of our artificial sensory device and its integration to our multisensory perception, we proposed and investigated two computational models. The first proposed model assume a decision integration process whereas the second model focused on Bayesian model of sensory integration. A numeric comparison of both models in terms of BIC is also provided.

Decision Integration Model

Before jumping into the proposed model for the fusion of the decisions, we introduce a type of confusion matrix that is calculated by the confidence reports. This confusion matrix, which we called Confidence Confusion Matrix, is required for the decision integration model.

Confidence Confusion Matrix

We extended the normal confusion matrix to second order judgments, which is self-reported confidence in decisions. When a participant reports a confidence for a decision, the confidence reports, which is kind of meta information, can be exploited to compute the meta-accuracy (type II accuracy). A confidence confusion matrix (CCM) is defined similar to confusion matrix as follows:

$$CCM = \begin{bmatrix} C_{1,1} & \cdots & C_{1,N} \\ \vdots & \ddots & \vdots \\ C_{N,1} & \cdots & C_{N,N} \end{bmatrix} \quad (3)$$

$C_{i,j}$ is sum of confidence values for decision $R_{i,j}$, where a ground truth signal i has been precepted as signal j. Meta-accuracy is defined on top of CMM as follows:

$$MetaAccuracy = \frac{\sum_{i=j=1}^{N} C_{i,j}}{\sum_{i=1}^{N}\sum_{j=1}^{N} C_{i,j}} \quad (4)$$

Meta-accuracy reflects the performance of the participants, by weighting the decisions with the confidence reports associated to those decisions. High confidence hits boost the meta-accuracy more than low confidence hits. Similarly, low confidence mistakes are not as bad as high confidence mistakes. CCM can be seen as a general representation where the normal confusion matrix is a sub-form of that with only one level of confidence.

Decision policy

In our decision fusion model, we assumed that the model can only access the first and second order decisional information, which are the decisions and the confidence reports. Considering that perceptual decisions come from only two modalities, we would have two separate CCMs, CCM^A and CCM^B. CCM is actually a table with decision-value pairs. The value of each decision in this table is an internal estimation obtained from confidence reports. An intuitive method for decision making in a multimodal setup, with multiple CCMs, is to choose the highest confidence ratings. Therefore, we defined *Max_CCM*, which selects, for each cell (*i, j*), the max value from the two confidence confusion matrices,

$$Max\_CCM_{i,j} = \max\left(CCM^{A}_{i,j}, CCM^{B}_{i,j}\right) \qquad (5)$$

Assuming *Max_CCM* shows the values of all decisions, we have investigated two decision making policies: linear decision making and parametrized softmax decision making. Please see the paper and appendix for more details and formula of decision-making policies.

If an agent chooses the linear policy, the perceptual accuracy is equivalent to the meta-accuracy obtained from the Max_CCM. The top rows of Figure 11 shows the result of the ideal linear decision maker. The model can successfully explain the accuracy of visual-tactile perception in the first blocks. Nevertheless, it fails to fit the experimental results in the last blocks and systematically underestimates the accuracy of visual-tactile perception. We provided more insight into this systematic underestimation in the paper and appendix.

We proposed a second variant of the model, parametrized softmax model, to address the systematic underestimation issue. This model is inspired from temperature scaling confidence calibration method as well as from softmax models of decision making (Cooper et al., 2014; Daw, O'doherty, Dayan, Seymour, & Dolan, 2006; Reverdy & Leonard, 2016). Please refer to the paper and appendix for more details and formula of this model.

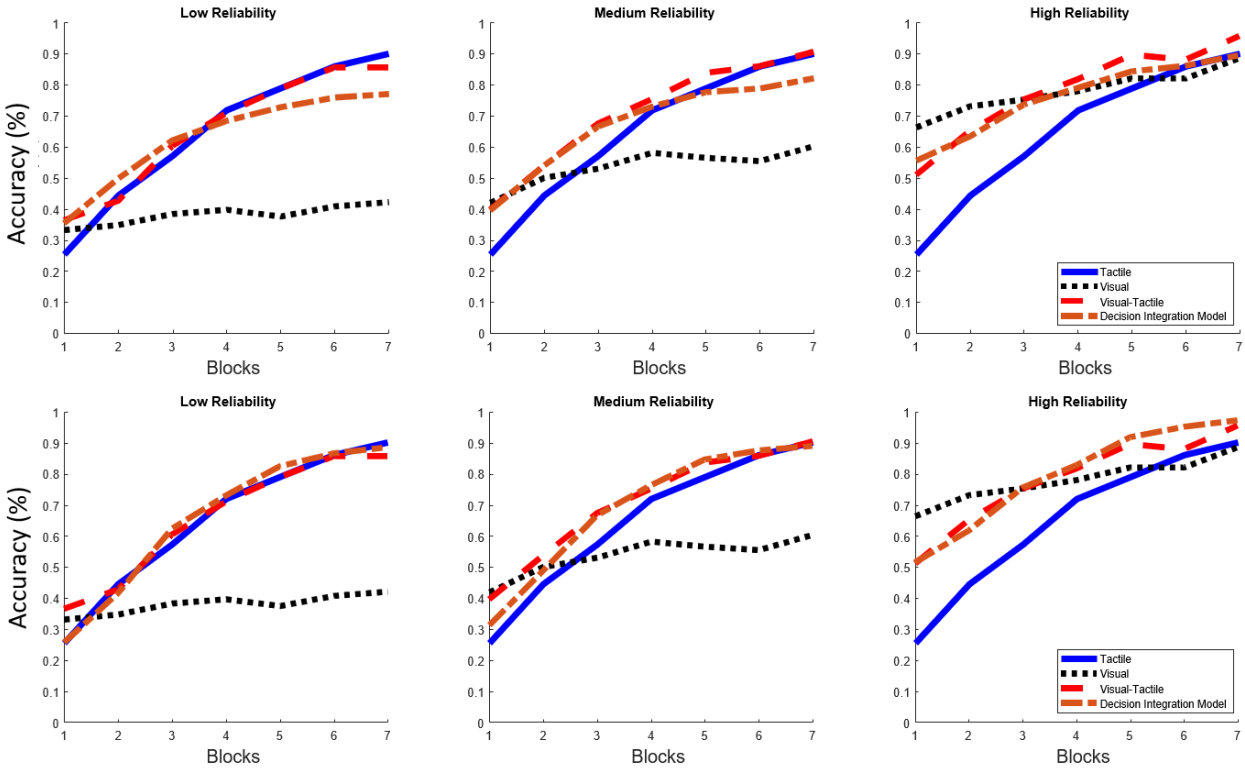

Figure 11. Linear and softmax decision integration models. The top figure depicts the linear model whereas the below one is the softmax model.

Figure 11 depicts and compare both the linear decision-making policy and the softmax decision making policy. The linear ideal observer model predicts reasonable only in the first blocks whereas the softmax model performs better on average and in all blocks. Both models do not consider a selection or a weighted combination of inputs, rather a simple integration of decisional information. Despite of this limitation, both models could reasonably capture the complex pattern of experimental results. Both models can successfully capture some surprising patterns. One of the surprising patterns can be observes in medium reliability condition, second and third blocks, where the accuracy of visual-tactile condition is higher than both unimodal accuracies. As another example, both models are able to capture the accuracy of visual-tactile perception in the first blocks of high reliability condition, where it resides between the accuracies of the unimodal conditions.

Nonetheless, the model fits also revealed the weakness of linear ideal observer; Specifically, model predictions are not adequate in the second half of the experiment, where the systematic underestimation is clearly an issue. This underestimation can happen because of a gradual alteration in the mechanism underlaying the fusion across multiple source of information. By gaining more experiences in novel tactile patterns, the perception strategy can shift from a decision fusion towards a multisensory integration. We propose a Bayesian multisensory

integration model to investigate the possible gradual shift of perception strategy towards the multisensory integration.

Multisensory Integration Model

We assessed whether and how the Bayesian model can explain the behavior of visual-tactile perception. We assumed that the reliability of perception is represented by the angular distance between the ground truth direction and the chosen direction by participants. If the dots move in direction α and while direction β is selected, then we defined the perceptual reliability as follows:

$$d_\theta(\alpha,\beta) = cos^{-1}\big(sin(\alpha)\cdot sin(\beta) + cos\,(\alpha)\cdot cos\,(\beta)\big) \quad (6)$$

$$r_{\alpha\beta} = \pi - d_\theta(\alpha,\beta) \quad (7)$$

$d_\theta(\alpha,\beta)$ is the angular distance between α and β, with an upper limit of π, and $r_{\alpha\beta}$ is defined as the reliability of perception. This definition indicates a less reliability of the perception for a high angular distance.

We utilized the concept of confusion matrix to predict the accuracy of visual-tactile perception, this time in accordance with the by Bayesian principles of sensory integration. We defined an integrated Bayesian confusion matrix as the weighted sum of unimodal confusion matrices. Obviously, the weight of each unimodal confusion matrix was defined as based corresponding relative perceptual reliability.

As it can be observed in Figure 12, the predicted accuracy by the Bayesian model is closed to the observed accuracy of the visual-tactile modality for the medium and high reliability conditions. This is especially prominent in the last 3 to 4 blocks. Nevertheless, the model behaves differently in the low reliability condition and suggests a selection behavior, because the predicted accuracy of the visual-tactile condition follows the observed unimodal tactile accuracy.

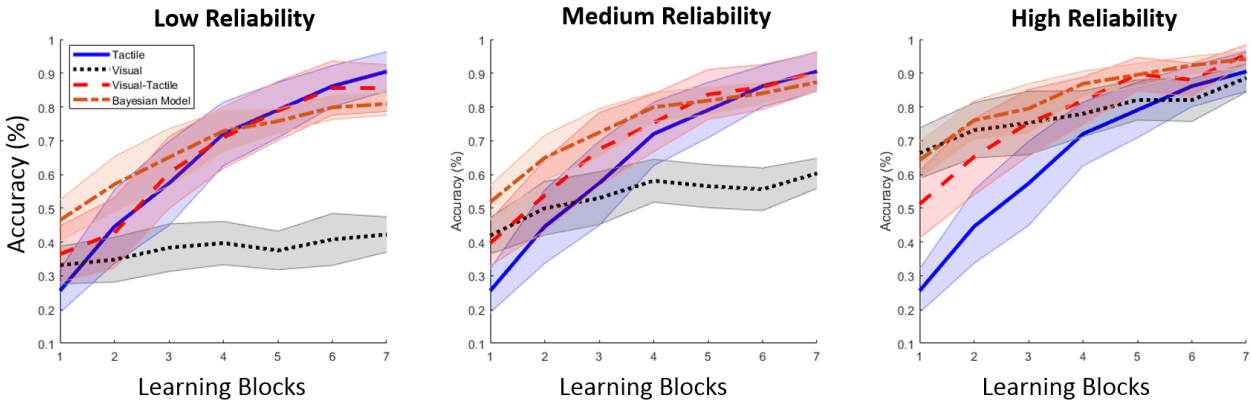

Figure 12. Predicted accuracy of perception by Bayesian sensory integration model. This model was fitted to each individual subject. The figure is reprinted from (M. N. Mahani et al., 2020).

The proposed Bayesian model could explain well the perceptual accuracy of last blocks in medium and high reliability conditions. Nevertheless, it almost failed to predict the perceptual accuracy of visual-tactile in the first blocks, whereas the linear decision fusion model could reasonably predict the accuracy. This shows the complementary predictions of linear decision fusion model (first blocks) and Bayesian sensory integration model (last blocks), and supports the gradual shift of perception strategy from decision fusion towards the multisensory integration. Please see the appendix and paper for statistical comparison of all proposed models.

Discussion

In the third study we investigated how a novel non-invasive artificial sensory system can be learned and possibly integrated within the human multisensory system. Our findings revealed that the perceptual accuracy of unimodal tactile condition raised over the course of the experiment, showing that participants successfully learned tactile patterns. Surprisingly, from the very beginning of the training, participants could integrate the information of novel artificial device into the visual-tactile perception, even though the accuracy of tactile modality was lower than the visual modality. The integration behavior can be clearly observed in the medium and high visual reliability conditions. Nevertheless, tactile was the dominant modality in the low reliability condition, and the perception strategy was selection of tactile rather than integration.

We proposed two computational models to shed light on mechanism underlying the perception and integration of our novel sensory device. The first model was based on the fusion of the

decisions and predicts the accuracy of visual-tactile perception by only accessing to the decisional information of unimodal perceptions, which are first order decisions and confidence reports. The second model was a Bayesian integration model which combines the sensory inputs in accordance with their reliability on a pre-decisional processing level. The second model considers a simple decision process on top of the integrated sensory signal.

The proposed decision fusion model interprets the confidence ratings as the estimation of decision values, and finds a policy that predicts well the observed data. We proposed a linear decision-making policy as well as a parametrized softmax decision-making policy. The linear model reasonably fitted the data only in the first half of the experiment. However, it underestimated accuracy in the second half of the experiment. To address this issue, we propose a new decisions-making policy with more degree of freedom: parametrized softmax policy with the linear equation parameter. The model with parametrized softmax policy could adequately fit the observed data in all training blocks. Both decision fusion models suggest that confidence reports are a decent representation of decision values. The gap between the two policy variants point to a possible gradual alteration of perception mechanism over the course of training. We addressed this possible alteration by exploiting a Bayesian model of sensory integration. The Bayesian model could reasonably fit the observed data in the second half of the experiment, complementary to the linear decision fusion policy. The complementary fits of both models support the gradual alteration of perception mechanism over the course of the learning.

All in all, the third study shows that participants incorporated and integrated symbolic tactile information to improve the accuracy of perception. Participants rely on a linear decision integration process during the initial learning phases, whereas in later phases they shift to Bayesian integration behavior.

General Discussion

We live in a complex and multimodal environment, where our evolved multisensory system allows us to optimally perceive it. With the advance of technology and rapid dynamics of our life habits, artificial sensory systems, such as wearable devices and sensory substitution systems, are becoming a part of our daily life. In the present thesis, we studied the interactions of tactile and visual sensory modalities with a major focus on tactile information generated by an artificial sensory feedback device. We believe that studying how such novel artificial systems are perceived by human is of great interest to scientists and engineers.

The thesis was organized in three studies in order to cover both the active and passive interactions of visual and tactile modalities. We addressed different aspects of visual-tactile interaction, with the minimum overlap across our studies. This diversity would help to acquire a broad vision of cognitive aspects of artificial devices, and their interactions with the visual sensory system.

The first study investigated the visual and tactile interactions in a realistic environment with varying reliability information. This study focused on the integration and calibration of visual and tactile modalities and covers the passive and potential active interactions. Our results show that we use different perception strategies to resolve conflicts in information. We found that participants started the integration strategy, however, they later adapted their perceptual strategy to selection or calibration in order to overcome the conflict. A comprehensive simulation allowed us to map the experimental results to appropriate perceptual strategies. We modeled the perceptual behavior and provided an ideal observer model that sheds light on the rationale for using different strategies. Our modelling revealed that causal inference in an environment of varying reliability depends on the amount of discrepancy as well as the relative reliability of stimuli across sensor modalities. Since the amount of discrepancy was constant across experiments, the reason behind taking different strategies is the difference in relative reliability across experiments. Moreover, we assessed the participants' awareness by analyzing the confidence reports in their judgments. Results show they performed better in high confidence situations than in the low confidence situations, which is an evidence for conscious perception. Interestingly, our findings show that participants did not change their confidence when they shifted from the integration behavior to the selection one. In contrast, they altered their confidence when they took calibration strategy and calibrated their modalities.

The second study of the thesis focused on the interactions of irrelevant cross-modal information. Specifically, we investigated the effect of task-irrelevant tactile/auditory information on processing of visual information. Introducing a novel artificial sensory device to some of our daily tasks, might cause some unwanted effects on other irrelevant tasks. This is thus important to investigate whether and how far could be the effect of such artificial sensory devices on cross-modal irrelevant tasks. On the other hand, the study was theoretically motivated by an extended version of DMC, which we called Multimodal DMC or MDMC. We extended the classical unimodal Simon experiment, which is a well-known task-irrelevant example, to a visual-tactile (experiment 1) and a visual-auditory (Experiment 2) Simon experiments.

The results of both experiments revealed the classical Simon effect, which is the effect of task irrelevant attribute of visual stimuli on response error and reaction time of visual perception. Furthermore, the effect of task irrelevant tactile information on response error and reaction time of visual decisions was significant in the visual-tactile experiment. In the visual-auditry experiment, the task irrelevant auditory stimuli could weakly affect the reaction time of visual perception, but not the response errors. The unreliable effect of auditory information on visual decisions was slightly unexpected. Nonetheless, other studies already found the insufficient effect of task irrelevant auditory stimuli on visual perception where both visual and auditory information are presented together.

Our proposed MDMC model, which is the multimodal extension of DMC model, fitted reasonably the experimental data in both visual-tactile and visual-auditory experiments. We further extend MDMC to the FN-MDMC (Faster Neutral MDMC), which was motivated by considering faster responses in experimental data of visual neutral condition. FN-MDMC was inspired from the studies that showed a visual stimulus presented at center point benefits from faster retinal processing in comparison with stimuli presented in the left or right.

Our modelling revealed that MDMC was not perfect with respect to the neutral conditions, because the MDMC did not consider the contribution of faster visual neutral conditions. The modeling results was thus improved by the FN-MDMC, which considers a faster residual processing for foveal visual stimuli. Our statistical comparison (in terms of BIC) between MDCM and FN-MDMC models also confirms the improvement of FN-MDMC. In general, both models retain our main modeling contribution that independent automatic activations from multiple task-irrelevant information may superimpose the controlled process to shape the

final decision process. MDMC framework helps us understanding the multisensory processing in conflict tasks and can be exploited for further studies in this area.

Last but not least, we addressed cross-modal learning of visual and tactile modalities with a focus on novel sensory experiences from an artificial sensory device. Specifically, we investigated how symbolic information from a non invasive wearable vibrotactile system can be learned and possibly integrated within our visual perception. Shedding light on this question would help us to design artificial sensory devices appropriately, in a way that can optimally integrate to our sensory system. It would also reveal the potentials and limits of such non-invasive sensory systems. We designed and implemented a custom vibrotactile belt which could generate novel and unexperienced tactile patterns. Our experiment had seven blocks of learning and evaluation phases. The main task was to learn the associations between symbolic tactile patterns and the moving direction of visual dot stimuli. In the evaluation phase, we manipulated the reliability of visual stimuli to have three reliability conditions: low, medium, and high. Our findings revealed that participants could learn and utilize novel tactile patterns in unisensory and multisensory perceptions. Surprisingly, they integrated the unexperienced tactile information with experienced visual information from the very beginning of the training. The integration behavior could be observed despite of the fact that the accuracy of tactile modality was initially lower than the visual modality. We even observed the selection of novel tactile information where the reliability of visual information was low. Without knowing the mechanism underlaying this perception behavior, it would be hard to generalize the findings of the study and reuse the obtained knowledge in further practical designs/theoretical studies. We therefore proposed two computational models that explained the observed perception behavior from two different aspects: decision integration, and multisensory integration perspectives. The proposed decision integration model predicts only based on first- and second-order perceptual decisions, whereas the sensory integration model considers a Bayesian integration of multisensory information, in which the sensory inputs integrate on a pre-decisional processing level. We considered a linear policy and a parametrized softmax policy for the decision integration model. The linear decision making policy reasonably predicted the visual-tactile perception only in the first several blocks of the experiment, but it underestimated the accuracy in the last blocks of the experiment. The parametrized softmax policy could fitted reasonably the experimental data in all training blocks because it has enough degree of freedom to consider a potential change in perception strategy during the learning. We hypothesized that the gradual alteration of the perception might represent a transition from a decision integration behavior

towards a sensory integration behavior. We investigated our hypothesis by modelling the experimental data with a Bayesian model of multisensory integration. Complementary to the linear decision fusion model, the Bayesian model reasonably predicted the experimental data only in the last several blocks of the experiment. The complementary predictions of the linear decision making policy and the Bayesian multisensory model supports our proposed hypothesis. Our numeric comparison of the models also confirmed the gradual change of behavior in the perception and integration over the course of training. Our findings also revealed that the confidence reports can be interpreted as a reasonable estimation of the decision values, and the proposed notion of confidence confusion matrix opens a novel horizon for future studies on decision making models with confidence ratings.

Taking everything into account, we designed and conducted three studies to investigate interactions of visual and tactile information. We put a major focus on tactile information delivered by an artificial vibrotactile device, which was designed and implemented for this thesis. We covered both active and passive interactions of visual and tactile modalities and kept studies diverse with almost no overlap. Each of the studies addressed one aspect of visual-tactile interactions and helped to understand the limits and potential of artificial devices. There are of course open points and potential future works which would improve our understanding of visual-tactile interactions as well as the cognitive aspects of artificial sensory devices.

Bibliography

1. Mahani, M.-A. N., Sheybani, S., Bausenhart, K. M., Ulrich, R. & Ahmadabadi, M. N. Multisensory Perception of Contradictory Information in an Environment of Varying Reliability: Evidence for Conscious Perception and Optimal Causal Inference. *Sci. Rep.* **7**, (2017).
2. Rowland, B. A., Quessy, S., Stanford, T. R. & Stein, B. E. Multisensory integration shortens physiological response latencies. (2007).
3. Drugowitsch, J., DeAngelis, G. C., Klier, E. M., Angelaki, D. E. & Pouget, A. Optimal multisensory decision-making in a reaction-time task. *Elife* **3**, (2014).
4. Hertenstein, M. & Weiss, S. *The handbook of touch: Neuroscience, behavioral, and health perspectives*. (2011).
5. Sathian, K., Zangaladze, A., Hoffman, J., Neuroreport, S. G.- & 1997, undefined. Feeling with the mind's eye. *journals.lww.com*.
6. Lacey, S. & Sathian, K. Crossmodal and Multisensory Interactions Between Vision and Touch. in *Scholarpedia of Touch* 301–315 (Atlantis Press, 2016). doi:10.2991/978-94-6239-133-8_25.
7. Forster, B., Cavina-Pratesi, C., Aglioti, S. M. & Berlucchi, G. Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time. *Exp. Brain Res.* **143**, 480–487 (2002).
8. Botvinick, M., Nature, J. C.- & 1998, undefined. Rubber hands 'feel'touch that eyes see. *nature.com*.
9. Streri, A., Pownall, T. & Kingerlee, S. *Seeing, reaching, touching: The relations between vision and touch in infancy.* (1993).
10. Streri, A. & Gentaz, E. Cross-modal recognition of shape from hand to eyes and handedness in human newborns. *Neuropsychologia* **42**, 1365–1369 (2004).
11. Johnson, L. A. & Higgins, C. M. A Navigation Aid for the Blind Using Tactile-Visual Sensory Substitution. in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society* 6289–6292 (IEEE, 2006). doi:10.1109/IEMBS.2006.259473.

12. Dakopoulos, D. & Bourbakis, N. G. Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey. *Appl. Rev.* **40**, 25 (2010).

13. Driver assistance system for driver assistance for consumption controlled driving. (2011).

14. Krause, M., Knott, V. & Benger, K. *Implementing the Tactile Detection Task in a Real Road Experiment to Assess a Traffic Light Assistant*.

15. Pilot Assistance System. (2015).

16. Kristjánsson, ´ Arni *et al.* Designing sensory-substitution devices: Principles, pitfalls and potential 1. *Restor. Neurol. Neurosci.* **34**, 769–787 (2016).

17. Shull, P. B. & Damian, D. D. Haptic wearables as sensory replacement, sensory augmentation and trainer - a review. *J. Neuroeng. Rehabil.* **12**, 59 (2015).

18. Gori, M., Giuliana, L., Sandini, G. & Burr, D. Visual size perception and haptic calibration during development. *Dev. Sci.* no-no (2012) doi:10.1111/j.1467-7687.2012.01183.x.

19. Gori, M., Giuliana, L., Alessandra, S., Sandini, G. & Burr, D. Calibration of the visual by the haptic system during development. *J. Vis.* **10**, 470 (2010).

20. Salzer, Y., Aisenberg, D., Oron-Gilad, T. & Henik, A. In touch with the Simon effect. *Exp Psychol* **61**, 165–179 (2014).

21. Gentile, G., Petkova, V. I. & Ehrsson, H. H. Integration of visual and tactile signals from the hand in the human brain: an FMRI study. *J Neurophysiol* **105**, 910–922 (2011).

22. Ossandón, J. P., König, P. & Heed, T. Irrelevant tactile stimulation biases visual exploration in external coordinates. *Sci. Rep.* **5**, (2015).

23. Keetels, M. & Vroomen, J. Temporal recalibration to tactile-visual asynchronous stimuli. *Neurosci Lett* **430**, 130–134 (2008).

24. Atteveldt, N. Van, Formisano, E., … L. B.-C. & 2006, undefined. The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *academic.oup.com*.

25. Navarra, J., Vatakis, A., Zampini, M., … S. S.-F.-C. B. & 2005, undefined. Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. *Elsevier*.

26. Pouget, A., Beck, J. M., Ma, W. J. & Latham, P. E. Probabilistic brains: knowns and unknowns. *Nat. Neurosci.* **16**, 1170 (2013).

27. Drugowitsch, J., DeAngelis, G. C., Angelaki, D. E. & Pouget, A. Tuning the speed-accuracy trade-off to maximize reward rate in multisensory decision-making. *Elife* **4**, (2015).

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

29. Shams, L. & Beierholm, U. R. Causal inference in perception. *Trends Cogn. Sci.* **14**, 425–432 (2010).

30. Ernst, M., integration, M. D. L.-S. cue & 2011, undefined. Multisensory perception: from integration to remapping. *books.google.com*.

31. Woods, A. J., Lehet, M. & Chatterjee, A. Context Modulates the Contribution of Time and Space in Causal Inference. *Front. Psychol.* **3**, 371 (2012).

32. Roach, N. W., Heron, J. & McGraw, P. V. Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. *Proc. R. Soc. B Biol. Sci.* **273**, 2159–2168 (2006).

33. Vision, M. E.-J. of & 2007, undefined. Learning to integrate arbitrary signals from vision and touch. *tvst.arvojournals.org*.

34. Körding, K. P. *et al.* Causal Inference in Multisensory Perception. *PLoS One* **2**, e943 (2007).

35. Zwiers, M., Opstal, A. Van, neuroscience, G. P.-N. & 2003, undefined. Plasticity in human sound localization induced by compressed spatial vision. *nature.com*.

36. Wozny, D., Neuroscience, L. S.-J. of & 2011, undefined. Recalibration of auditory space following milliseconds of cross-modal discrepancy. *Soc Neurosci.*

37. Van der Burg, E., Alais, D. & Cass, J. Rapid recalibration to audiovisual asynchrony. *J. Neurosci.* **33**, 14633–7 (2013).

38. Gori, M., Sciutti, A., Burr, D. & Sandini, G. Direct and Indirect Haptic Calibration of Visual Size Judgments. *PLoS One* **6**, e25599 (2011).

39. Stroop, J. R. Studies of interference in serial verbal reactions. *J. Exp. Psychol.* **18**, 643 (1935).

40. Simon, J. R. & Wolf, J. D. Choice reaction time as a function of angular stimulus-response correspondence and age. *Ergonomics* **6**, 99–105 (1963).

41. Eriksen, B. A. & Eriksen, C. W. Effects of noise letters upon the identification of a target letter in a nonsearch task. *Percept. Psychophys.* **16**, 143–149 (1974).

42. Ulrich, R., Schroter, H., Leuthold, H. & Birngruber, T. Automatic and controlled stimulus processing in conflict tasks: Superimposed diffusion processes and delta functions. *Cogn Psychol* **78**, 148–174 (2015).

43. Stone, M. Models for choice-reaction time. *Psychometrika* **25**, 251–260 (1960).

44. Ratcliff, R. A theory of memory retrieval. *Psychol. Rev.* **85**, 59 (1978).

45. Ratcliff, R., review, P. S.-P. & 2004, undefined. A comparison of sequential sampling models for two-choice reaction time. *psycnet.apa.org*.

46. Servant, M., White, C., Montagnini, A. & Burle, B. Linking Theoretical Decision-making Mechanisms in the Simon Task with Electrophysiological Data: A Model-based Neuroscience Study in Humans. *J Cogn Neurosci* **28**, 1501–1521 (2016).

47. Simon, J. R. & Craft, J. L. Effects of an irrelevant auditory stimulus on visual choice reaction time. *J Exp Psychol* **86**, 272–274 (1970).

48. Donohue, S. E., Appelbaum, L. G., Park, C. J., Roberts, K. C. & Woldorff, M. G. Cross-modal stimulus conflict: the behavioral effects of stimulus input timing in a visual-auditory Stroop task. *PLoS One* **8**, e62802 (2013).

49. Kennett, S., Eimer, M., Spence, C. & Driver, J. Tactile-visual links in exogenous spatial attention under different postures: Convergent evidence from psychophysics and ERPs. *J. Cogn. Neurosci.* **13**, 462–478 (2001).

50. Spence, C., Pavani, F. & Driver, J. Spatial constraints on visual-tactile cross-modal distractor congruency effects. *Cogn Affect Behav Neurosci* **4**, 148–169 (2004).

51. Yue, Z., Bischof, G. N., Zhou, X., Spence, C. & Röder, B. Spatial attention affects the processing of tactile and visual stimuli presented at the tip of a tool: An event-related potential study. *Exp. Brain Res.* **193**, 119–128 (2009).

52. Poole, D., Couth, S., Gowen, E., Warren, P. A. & Poliakoff, E. Adapting the crossmodal congruency task for measuring the limits of visual–tactile interactions within and between groups. *Multisens. Res.* **28**, 227–244 (2015).

53. Osaka, N. Reaction time as a function of peripheral retinal locus around fovea: effect of stimulus size. *Percept. Mot. Skills* **42**, 603–606 (1976).

54. Hübner, R. Does attentional selectivity in global/local processing improve discretely or gradually? *Front. Psychol.* **5**, 61 (2014).

55. Howard, I. P. & Templeton, W. B. Human spatial orientation. (1966).

56. Welch, R. B. & Warren, D. H. Immediate perceptual response to intersensory discrepancy. *Psychol Bull* **88**, 638–667 (1980).

57. Bertelson, P. & Radeau, M. Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Percept Psychophys* **29**, 578–584 (1981).

58. Slater-Hammel, A. T. Reaction time to light stimuli in the peripheral visual field. *Res. Quarterly. Am. Assoc. Heal. Phys. Educ. Recreat.* **26**, 82–87 (1955).

59. Abboud, S., Hanassy, S., Levy-Tzedek, S., Maidenbaum, S. & Amedi, A. EyeMusic: Introducing a "visual" colorful experience for the blind using auditory sensory substitution. *Restor. Neurol. Neurosci.* **32**, 247–257 (2014).

60. Maidenbaum, S., Abboud, S. & Amedi, A. Sensory substitution: closing the gap between basic research and widespread practical visual rehabilitation. *Neurosci. Biobehav. Rev.* **41**, 3–15 (2014).

61. Collins, K. L. *et al.* Ownership of an artificial limb induced by electrical brain stimulation. *Proc. Natl. Acad. Sci.* **114**, 166–171 (2017).

62. Dadarlat, M. C., O'doherty, J. E. & Sabes, P. N. A learning-based approach to artificial sensory feedback leads to optimal integration. *Nat. Neurosci.* **18**, 138 (2015).

63. Iqbal, M. H., Aydin, A., Brunckhorst, O., Dasgupta, P. & Ahmed, K. A review of wearable technology in medicine. *J. R. Soc. Med.* **109**, 372–380 (2016).

64. Mukhopadhyay, S. C. Wearable sensors for human activity monitoring: A review. *IEEE Sens. J.* **15**, 1321–1330 (2015).

65. Son, D. *et al.* Multifunctional wearable devices for diagnosis and therapy of movement disorders. *Nat. Nanotechnol.* **9**, 397 (2014).

66. Drugowitsch, J. *et al.* Tuning the speed-accuracy trade-off to maximize reward rate in multisensory decision-making. *Elife* **4**, (2015).

67. Ernst, M., Nature, M. B.- & 2002, undefined. Humans integrate visual and haptic information in a statistically optimal fashion. *nature.com*.

68. Butler, J. S., Smith, S. T., Campos, J. L. & Bülthoff, H. H. Bayesian integration of visual and vestibular signals for heading. *J. Vis.* **10**, 23 (2010).

69. Reverdy, P. & Leonard, N. E. Parameter estimation in softmax decision-making models with linear objective functions. *IEEE Trans. Autom. Sci. Eng.* **13**, 54–67 (2016).

70. Daw, N. D., O'doherty, J. P., Dayan, P., Seymour, B. & Dolan, R. J. Cortical substrates for exploratory decisions in humans. *Nature* **441**, 876 (2006).

71. Cooper, J. A. *et al.* Training attention improves decision making in individuals with elevated self-reported depressive symptoms. *Cogn. Affect. Behav. Neurosci.* **14**, 729–741 (2014).

Appendix

Study 1

The official citation that should be used in referencing this material is:

Mahani, Mohammad-Ali Nikouei, Saber Sheybani, Karin Maria Bausenhart, Rolf Ulrich, and Majid Nili Ahmadabadi. "Multisensory perception of contradictory information in an environment of varying reliability: Evidence for conscious perception and optimal causal inference." Scientific reports 7, no. 1 (2017): 1-15.

https://doi.org/10.1038/s41598-017-03521-2

The final publication is available here:

https://www.nature.com/articles/s41598-017-03521-2

Author contributions

All authors contributed in the main manuscript text. M.A.N.M. prepared figures. M.A.N.M. and S.S. collected the data. All authors contributed in the data analysis. M.A.N.M., M.N.A., and S.S. designed the study experiment and all authors reviewed the manuscript.

SCIENTIFIC REPORTS

OPEN

Multisensory Perception of Contradictory Information in an Environment of Varying Reliability: Evidence for Conscious Perception and Optimal Causal Inference

Received: 25 August 2016
Accepted: 1 May 2017
Published online: 09 June 2017

Mohammad-Ali Nikouei Mahani[1,2,3], Saber Sheybani[1], Karin Maria Bausenhart[2], Rolf Ulrich[2] & Majid Nili Ahmadabadi[1,3]

Two psychophysical experiments examined multisensory integration of visual-auditory (Experiment 1) and visual-tactile-auditory (Experiment 2) signals. Participants judged the location of these multimodal signals relative to a standard presented at the median plane of the body. A cue conflict was induced by presenting the visual signals with a constant spatial discrepancy to the other modalities. Extending previous studies, the reliability of certain modalities (visual in Experiment 1, visual and tactile in Experiment 2) was varied from trial to trial by presenting signals with either strong or weak location information (e.g., a relatively dense or dispersed dot cloud as visual stimulus). We investigated how participants would adapt to the cue conflict from the contradictory information under these varying reliability conditions and whether participants had insight to their performance. During the course of both experiments, participants switched from an integration strategy to a selection strategy in Experiment 1 and to a calibration strategy in Experiment 2. Simulations of various multisensory perception strategies proposed that optimal causal inference in a varying reliability environment not only depends on the amount of multimodal discrepancy, but also on the relative reliability of stimuli across the reliability conditions.

Optimal integration of multisensory information received from different sensory organs is crucial for a coherent perception of the complex environment. Many studies have reported benefits of multisensory compared to unisensory perception in psychophysical tasks. Integration of different modalities when they are congruent and synchronous[1–3] leads to a significant decrease in response time[4–6], an increase in reliability, and an increase in accuracy[7–14]. However, when information is incongruent across different sensory modalities, integration may lead to a biased percept. Various models of information processing, including causal inference and calibration, have been suggested to describe perception when different modalities receive inconsistent information[15,16].

Multisensory causal inference is the process of deciding which sensory inputs originate from the same cause[17,18]. It plays a crucial role in multisensory perception, especially when there is a discrepancy among the modalities. Previous studies have reported that the probability of assuming a common cause decreases with the increase of this discrepancy[19]. It has also been shown that causal inference across different modalities not only depends on temporal, spatial, and contextual features of the stimuli[20] but also on prior knowledge and experiences[21,22]. According to ideal observer models[19], causal inference in multisensory perception is often performed in line with Bayes' rule[16] and in such a way to minimize the estimation error, defined as the mean squared error between the true value and the estimated value of the perceptual system.

In order to cope with inconsistent sensory information, the brain has to discover whether the discrepancy is due to a random noise in the sensory information, or to a systematic error, potentially in one of the sensory

[1]Cognitive Systems Lab, School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran. [2]Cognition and Perception, Department of Psychology, University of Tübingen, Tübingen, Germany. [3]School of Cognitive Science, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran. Correspondence and requests for materials should be addressed to M.-A.N.M. (email: nikouei@ut.ac.ir)

systems[18]. Recent studies have suggested that if the source of the discrepancy is a systematic error, calibration is beneficial to resolve the persisting discrepancy. Meanwhile, the accuracy of the sensory systems is maintained throughout calibration[18,23]. Calibration models describe how unimodal sensory percepts could be changed so that the inconsistency between unimodal percepts decreases. Classical calibration models indicated that all the sensory modalities are calibrated corresponding to vision, known as visual dominance[24,25]. Recent studies have proposed general calibration models which involve other modalities as well. In such models, all of the modalities are calibrated in accordance with each other based on the relative reliability[15,26] of their corresponding cues, assuring minimum-variance sensory estimates over time[15,27]. However, Zaidel *et al.*[28] reported independence of calibration from cue reliabilities in the context of a visual-vestibular setup. They proposed a fixed-ratio calibration model in which each modality contributes to calibration with a certain weight that remains constant irrespective of variations of cue reliability.

Although calibration adjusts perception in the case of systematic errors, it is still not clear under which conditions calibration takes place. It seems that multisensory causal inference and multisensory calibration are tied together, that is, calibration of different modalities according to each other would be reasonable only when signals originate from the same external event. Even though the interaction of causal inference and calibration is still unclear, the minimal error in the integration of multimodal contradictory information can be achieved if causal inference and Bayesian integration processes are considered jointly[19].

Consequently, several processes underlying multisensory perception of incongruent information have been suggested. In order to better understand the mechanisms of perception on one hand and calibration in a multimodal environment on the other hand, it is important to study interactions across these processes. However, some of these interactions might emerge especially under more realistic conditions rather than in the restricted environment that is usually studied in multisensory perception experiments. In natural environments, the reliability of the information received from each organ is often not stationary but can change within a short period of time. For example, when one drives in foggy (or rainy) weather, the reliability of the visual information can change within a short period when the density of the fog varies. Thus, the reliability of the information is not stationary in natural scenarios. Therefore, when studying the interaction of different perceptual processes such as causal inference and calibration, one should consider the role of varying reliabilities. This might be especially informative, for example, in order to identify the conditions which are required for the calibration process, or to investigate the flexibility of the processes underlying multisensory calibration. The multisensory processes, such as causal inference and calibration, require experiences that are acquired through previous observations[22,29]. In an environment of varying reliability, the question is raised whether the causal inference problem is solved based on the aggregated information across all reliability conditions, or whether it is solved for each reliability condition separately.

To the best of our knowledge, multisensory calibration and multisensory causal inference of contradictory information have never been studied under conditions where the reliability of the sensory modalities varies randomly from trial to trial within a single session. Therefore, the present study investigates multisensory perception of contradictory spatial information with a concurrent variation of reliabilities. Specifically, we designed an experiment comprised of three sections administered in the following order: a visual-auditory section, a visual-tactile-auditory section, and a unimodal section. In every trial of each section, participants received two consecutive section-specific stimuli presented from varying spatial locations. They were asked to determine whether the second stimulus was perceived spatially on the left or on the right side of the first stimulus (see figures in methods and apparatus section). In the first section, the reliability of the visual stimuli was high in half of the trials, while the reliability of the auditory stimuli was constant in all trials (see Fig. 1). Visual-auditory stimuli had a spatial discrepancy; the auditory stimulus was always presented on the left side of the visual stimulus with a conflict angle (Δ) of $-4°$. In the visual-tactile-auditory section, tactile stimuli (delivered to the abdomen) accompanied the visual-auditory stimuli. The reliability of the various modalities changed according to three different reliability conditions (see Fig. 1). In this second section, the tactile stimulus was presented 4 cm to the left of the spatially congruent audiovisual stimuli, where the central position of the tactile stimulus was the vertical body axis. The term cue-conflict refers to the spatial conflict between different individual modalities in our study. In this section, cue-conflict means that tactile stimulus was shifted two motors (4 cm) to the left. The third and final section was a unimodal section in which the reliability of each modality was evaluated separately for each participant.

Since the cue-conflict was not large, it was expected that the participants would initially assume a common cause, and hence integrate the different modalities at the beginning of a multimodal section. However, during the course of the experiment, one could expect that participants would realize that there is a systematic conflict rather than unsystematic noise. Accordingly, during each section, the participants' integration strategy might change. We therefore investigated how their spatial perception varied between the first half and the second half of each section. We hypothesized that three different perceptual strategies could emerge in this setup: (1) Assuming the same cause for the modalities and continuing to integrate the different modalities without any calibration or any change in perception. (2) Assuming the same cause for the modalities, but calibrating the individual modalities in order to resolve the cue-conflict. (3) Inferring that there are several causes for the modalities and thus selecting the most reliable one in each trial.

The first strategy implies that there is no significant change in the perception, specifically between the first and the second half of each section in the experiment. However, the second and the third possibilities represent new strategies of perception. In this study, for the sake of simplicity, we use the term "adaptation" to indicate such possible perceptional changes between the first and the second half of a section, independent of the specific perception strategy. Crucially, we were interested in determining whether these adaptation strategies would depend on the varying reliability of the signals.

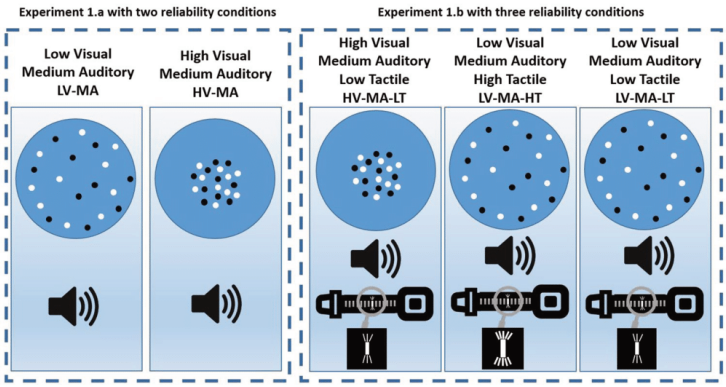

Figure 1. Reliability conditions in Experiment 1.a and Experiment 1.b. Experiment 1.a was a visual-auditory experiment with two reliability conditions (the two left-most figures). The spread of the visual dots was wider in the LV-MA condition than in the HV-MA condition. Experiment 1.b had three reliability conditions (the three right-most figures). In the HV-MA-LT condition, the visual input had the highest reliability, while in the LV-MA-HT condition the tactile stimulus was the most reliable stimulus. In the LV-MA-LT condition, the reliability of both the visual and tactile stimuli was low. The reliability of the auditory stimulus was the same in all the three conditions of Experiment 1.b.

As the environment becomes more dynamic, the uncertainty in decisions should increase. Even though it is reported that people are usually aware of their own uncertainty during their perception[30], there is no consensus on this issue in the literature. Faivre *et al.*[31] showed that people integrate multimodal information unconsciously when the stimuli were subliminal and after undergoing a conscious learning procedure. Nevertheless, they did not ask their participants to report their confidence about their decisions. Such confidence ratings assess a participant's awareness of his/her perceptual performance. Specifically, participants demonstrate a high level of monitoring accuracy when their confidence and perceptual performance are positively correlated. Confidence ratings have also been used to assess whether a brain process is conscious or unconscious[32]. For example, if the variation in perceptual performance is not accompanied by consistent changes in confidence, then the perceptual processes and the resulting decisions are likely to be unconscious[33]. In our study, we also assessed the participants' confidence ratings in order to examine whether the participants were aware of the perceptual difficulty associated with the varying reliabilities of the stimuli.

Results

Psychometric function. As outlined above, comparing the point of subjective equality (PSE) for each reliability condition between the first half and the second half of the experiment assesses the strategy adopted by the participant during the experiment. This comparison addresses PSE shifts towards either the visual, auditory, or tactile cues and thus reveals the relative weighting of each modality in the integration process for each reliability condition. The observed psychometric functions measured the proportion of rightward choices against the stimulus angle. Therefore, PSE $= -4$ means a shift of four degrees to the right and PSE $= +4$ a shift of four degrees to the left. We used generalized linear model (GLM) regression in order to fit all psychometric functions. These are maximum likelihood models with the binomial distribution and the probit as the link function (the inverse of the transformation is the link function). The PSE was calculated as the threshold value of the psychometric function at 0.5. Figure 2 illustrates the fitted psychometric function (averaged over participants) for each reliability condition; the solid blue lines show the performance before adaptation and dashed red lines show the performance after adaptation. The mean squared error of these fits is 0.0058 ± 0.0049.

Comparison of the PSEs before and after adaptation (see the magnified plots) illustrates that the PSE changed in some reliability conditions. In a further analysis, psychometric functions were fitted to the data of each participant in each reliability condition. All further analyses and modeling were based on these individual psychometric functions. The average mean squared error of these individual psychometric functions was 0.018 ± 0.016 which indicates a reasonable fit. The average correlation between individual psychometric functions and data points was 0.902 ± 0.085 (average p-values. 014 ± 0.020).

Visual-Auditory (Experiment 1.a). In this part of the experiment, the strategy adopted by participants to cope with the visual-auditory perception in an environment of varying reliability was studied. Participants were exposed to two reliability conditions in this section: the high reliability visual and medium reliability auditory stimuli (HV-MA), and the low reliability visual and medium reliability auditory (LV-MA). All trials of the two conditions were randomly mixed. Figure 1 illustrates all reliability conditions in Experiment 1.a and Experiment 1.b.

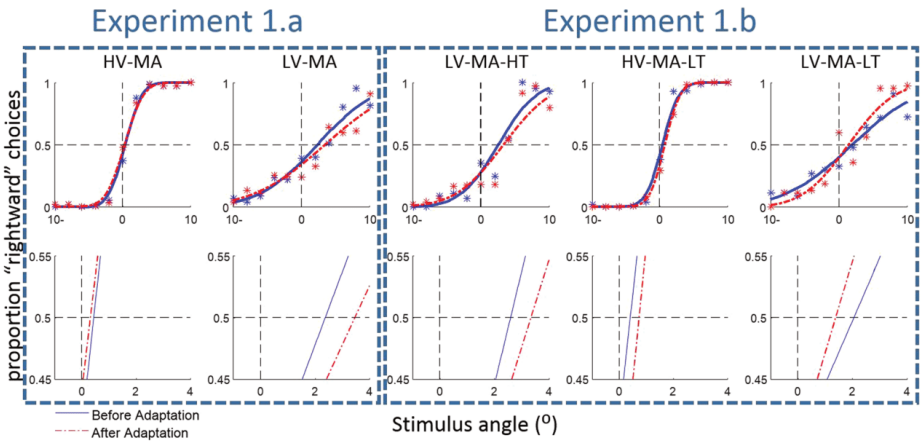

Figure 2. Fitted psychometric functions to all trials of all participants. The first row shows the psychometric functions before and after adaptation in all reliability conditions. The second row shows the magnified plots of the first row around the midpoint.

In order to examine the potential effect of *reliability condition* (LV-MA vs. HV-MA) and of *adaptation* (first half vs. second half) on perception, the PSEs were analyzed using 2 × 2 within-subjects ANOVA (Bayesian analyses are available in the Supplementary material). This analysis showed a significant interaction effect, $F(1,18) = 5.52$, $p = 0.030$, and a significant effect of reliability condition, $F(1,18) = 72.52$, $p < 0.001$; while factor adaptation was not significant, $F(1,18) = 3.43$, $p = 0.080$ (see Fig. 3A). Since the interaction effect was significant, the PSE changes were assessed for each reliability condition. According to a two-sided paired-sample t-test, the effect of adaptation on the PSE was not significant in the HV-MA condition, $t(18) = 0.53$, $p = 0.599$, but significant in the LV-MA condition, $t(18) = 2.17$, $p = 0.043$. Therefore, in the second half of the LV-MA condition, participants' judgments about the multisensory location were significantly biased towards the auditory location. However, the PSE in the HV-MA condition did not change and remained close to the location of the visual stimuli. This pattern of results indicates that perceptual changes are specific to different reliability conditions. We will investigate the source of this change later by modeling these effects.

The slope of psychometric functions was analyzed with the same ANOVA design as the PSE. The results showed that the slope in the HV-MA condition was significantly greater than in the LV-MA condition, $F(1,18) = 23.59$, $p < 0.001$. Neither the factor adaptation, $F(1,18) = 0.17$, $p = 0.682$, nor the interaction of adaptation × reliability condition yielded a significant effect, $F(1,18) = 0.12$, $p = 0.733$. The average slope of the psychometric function was 2.09 ± 2.25 and 1.85 ± 2.14 in the first half and second half of the HV-MA condition, respectively. In the LV-MA condition, the average slope was 0.171 ± 0.08 and 0.148 ± 0.09 in the first half and second half. The difference between the slopes in the HV-MA and LV-MA conditions suggests that our manipulation of the reliability condition was successful.

An analogous ANOVA was performed on confidence ratings. This analysis revealed no significant interaction between reliability condition and adaptation, $F(1,18) = 1.73$, $p = 0.205$. There was also no significant effect of adaptation, $F(1,18) = 0.14$, $p = 0.712$. Nevertheless, the effect of the reliability condition on confidence was highly significant, $F(1,18) = 112.28$, $p < 0.001$, see Fig. 3B. These results showed that participants reported significantly lower confidence when the reliability of stimuli was also low (the LV-MA condition) which supports the view that they were aware of their performance. However, confidence ratings did not significantly differ between the first and second half of the experiment, even when they changed their judgments in LV-MA conditions.

Visual-Tactile-Auditory (Experiment 1.b). The second section of the experiment had the same structure as the first section, except that the tactile modality was added and three varying reliability conditions were implemented: a high reliability visual (HV-MA-LT) condition, a high reliability tactile (LV-MA-HT) condition, and a medium reliability auditory (LV-MA-LT) condition (see Fig. 1 and also methods section for more details). Although many studies have mentioned interactions between touch and the other modalities in the context of adaptation[15,35,36], those which have studied the spatial adaptation are rare[37]. Similar to the first section, the difference in the PSEs between the first half and the second half of the experiment is crucial to study the adaptation strategy among the modalities.

The PSE results of this section are shown in Fig. 3C. The 3 × 2 within-subjects ANOVA on PSEs demonstrated that the interaction effect of factor adaptation and reliability (HV-MA-LT, LV-MA-HT, and LV-MA-LT) was significant, $F(2,36) = 4.53$, $p = 0.018$. It also indicated that the effect of adaptation on PSE was not significant, $F(1,18) = 0.082$, $p = 0.777$, while the effect of the reliability condition on PSE was significant, $F(2,36) = 41.64$,

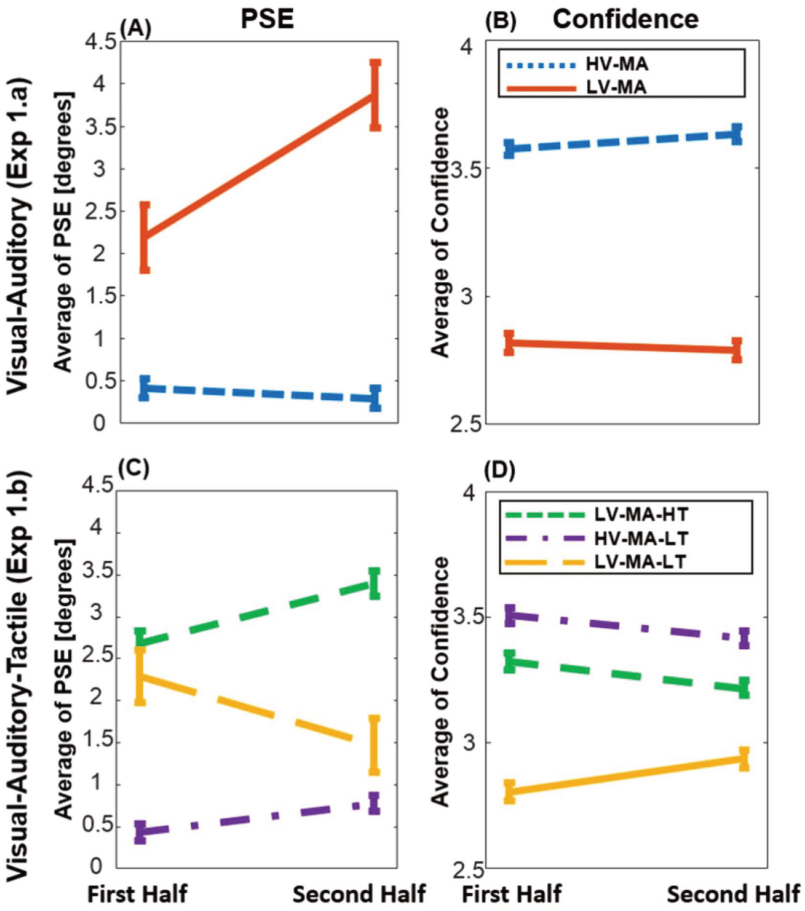

Figure 3. PSE variations as well as confidence changes between the first and the second half of each section across different reliability conditions. The error bars were computed by the method described in ref. 34 and show ± 1SE. (**A**) PSE variation in the visual-auditory section: The PSE was shifted toward the auditory location in the LV-MA condition, while it remained close to the visual location in the HV-MA condition. (**B**) Confidence alteration in the visual-auditory section: there was no significant change in the confidence, however the confidence in the HV-MA condition is significantly higher than in the LV-MA. (**C**) PSE variation in the visual-tactile-auditory section: the PSEs in the HV-MA-LT and the LV-MA-HT conditions moved toward the tactile location significantly. Nevertheless, it was not significant in the LV-MA-LT condition (yellow line). (**D**) Confidence alteration in the visual-tactile-auditory section: All confidence ratings changed significantly, a significant increase in the LV-MA-LT condition, and significant decrease in the HV-MA-LT and the LV-MA-HT conditions.

$p < 0.001$. The significant interaction effect indicates that adaptation varied across reliability conditions. Therefore, we conducted a separate 2×2 within-subjects ANOVA without the LV-MA-LT condition, since the LV-MA-LT condition had a different adaptation trend. This second ANOVA revealed that the interaction between HV-MA-LT and LV-MA-HT was not significant, whereas the effect of reliability condition on PSE, $F(1,18) = 138.93$, $p < 0.001$, and also the effect of adaptation on PSE, $F(1,18) = 7.23$, $p = 0.015$, were significant. The paired-sample t-test also showed that the effect of adaptation on the PSE was not significant in the LV-MA-LT condition, $t(18) = 1.30$, $p = 0.209$.

The PSE in the LV-MA-HT condition was shifted towards the tactile location, suggesting an alteration of multisensory perception. The same alteration was seen in the condition in which the high reliability stimuli was visual (HV-MA-LT), even though confidence was higher in the LV-MA-HT condition. However, for the LV-MA-LT

condition, there was no indication of such variation in the PSE since audition was not affected by touch/vision and the auditory stimuli were the most reliable stimuli in the LV-MA-LT reliability condition (see the slope analyses in the "unimodal study Experiment 1.c section").

An analogous ANOVA on the slope of the psychometric function revealed that the effect of the reliability condition was significant; $F(2,36) = 12.08$, $p < 0.001$. However, neither factor adaptation, $F(1,18) = 0.83$, $p = 0.374$, nor the interaction of adaptation x reliability condition, $F(2,36) = 2.39$, $p = 0.106$, was significant. Like in Experiment 1.a, the manipulation of the reliability conditions was successful, since the effect of the reliability condition was significant. The slope of psychometric functions in the first half and second half were, respectively, as follows: 0.30 ± 0.15, 0.49 ± 1.08 in the LV-MA-HT condition, 2.08 ± 2.6, 1.14 ± 1.49 in the HV-MA-LT condition, and 0.14 ± 0.10, 0.21 ± 0.09 in the LV-MA-LT condition.

The same ANOVA design was used to study the confidence ratings. When all three modalities are considered together, the 3×2 repeated two-way ANOVA on confidence rating showed an significant interaction between reliability and adaptation, $F(2,36) = 11.47$, $p < 0.001$, while factor adaptation was not significant, $F(1,18) = 0.43$, $p = 0.519$, and the effect of reliability condition was significant, $F(2,36) = 39.05$, $p < 0.001$, see Fig. 3D. However, an ANOVA without the LV-MA-LT condition showed that there was no interaction between reliability condition and adaptation, $F(1,18) = 0.10$, $p = 0.750$. Factor adaptation, $F(1,18) = 6.90$, $p = 0.017$, and reliability condition, $F(1,18) = 7.27$, $p = 0.015$, produced significant effects on confidence. Paired-sample t-tests showed that the effect of adaptation on the confidence was also significant in the LV-MA-LT condition, $t(18) = 2.49$, $p = 0.023$, but in the opposite direction than in the HV-MA-LT and LV-MA-HT conditions. Figure 3D illustrates the changes of the confidence ratings in each reliability condition.

As in experiment 1.a, participants reported significantly different confidence in various reliability conditions suggesting that they were aware of their performance (additional analyses of confidence ratings are provided in the discussion section). In contrast to experiment 1.a, they changed their confidence between the first and the second half of the experiment. Thus, in this section the bias in participants' perception changed together with their confidence.

Unimodal study (Experiment 1.c). In order to model multisensory perception, we need to estimate the unimodal perceptual reliabilities of the participants. Therefore, the last section was a unimodal experiment, designed to estimate the reliability of the stimuli in each of the modalities for each of the participants separately. One-sample t-tests were conducted for each stimulus type in order to examine whether there were any persisting biases in location perception for unimodal stimuli. The average of PSEs, as well as the p-values of the t-test, in each of the reliability conditions were as follows: -0.22 ± 0.96, $t(18) = 0.99$, $p = 0.335$, for high reliability visual, 0.62 ± 3.9, $t(18) = 0.69$, $p = 0.498$, for low reliability visual, -0.90 ± 1.8, $t(17) = 2.08$, $p = 0.052$, for high reliability tactile, 3.2 ± 4.6, $t(18) = 3.01$, $p = 0.007$, for low reliability tactile, and -0.04 ± 1.1, $t(18) = 0.15$, $p = 0.880$, for medium reliability auditory conditions. One PSE value was eliminated from the analysis in the high reliability tactile condition since it was outside of the average $\pm 3\sigma$ range. Thus, the t-test for this condition was conducted with eighteen data instead of nineteen.

The PSE for low reliability tactile was significantly biased. Specifically, participants displayed a partial bias to the left for the low reliability tactile condition. Since the unimodal experiment was performed after experiment 1.b, this may reflect an acquired shift in location perception corresponding to the previously experienced conflict angle. Presumably, the shift was significant in the low reliability condition because the certainty was lower when the reliability of the stimuli was also lower. Thus, participants probably used their prior information about the conflict angle more when the uncertainty of the stimuli was higher.

A one-way within-subjects ANOVA revealed that there is a significant difference across slopes of the unimodal perceptions, $F(4,72) = 4.42$, $p = 0.003$. The average of the slopes was 0.57 ± 0.25 in the high reliability visual condition, 0.075 ± 0.026 in the low reliability visual condition, 0.58 ± 1.16 in the high reliability tactile condition, 0.12 ± 0.13 in the low reliability tactile condition, and 0.39 ± 0.18 in the medium reliability auditory condition.

The average of the reported confidence was 3.43 ± 0.39 in the high reliability visual condition, 2.72 ± 0.54 in the low reliability visual condition, 3.39 ± 0.37 in the high reliability tactile condition, 2.32 ± 0.75 in the low reliability tactile condition, and 3.02 ± 0.48 in the medium reliability auditory condition. A one-way ANOVA showed that there is a significant difference in confidence across unimodal conditions, $F(4,72) = 37.74$, $p < 0.001$. Post-hoc Tukey test showed that all unimodal confidence ratings were significantly different from each other except two of them: the difference between the high reliability visual and high reliability tactile conditions ($p = 0.988$), as well as the difference between the low reliability tactile and low reliability visual ($p = 0.110$) situations.

In order to account for the results of the experiment, we developed a mathematical model and conducted simulations that explain different strategies by considering the causal inference and calibration processes simultaneously. This model is inspired by the Bayesian framework of multisensory perception and minimizes the overall perception error across all reliability conditions.

Discussion

Adaptation strategies in an environment of varying reliability. Our brain faces many perceptual conflicts every day, especially in multimodal perception. Although conflicts seem to be the source of some misperceptions, they often trigger perceptual learning processes. Several studies suggested that cross-modal calibration processes are triggered by cross-modal cue-conflicts[38–40]. The present experiments examined how perception is modified in the presence of cross-modal conflict stimuli in an environment of varying reliability. To enable a better understanding of the results, we simulated different possible multimodal perception strategies in order to model the mechanism that may cause the PSE shifts in an environment of varying reliability (see Fig. 4).

Consider a multimodal stimulus S being perceived through two different modalities A (blue line), and B (red line) and let $\hat{S}_A = f_A(S)$ and $\hat{S}_B = f_B(S)$ denote their internal estimations, respectively. Furthermore, several studies[8]

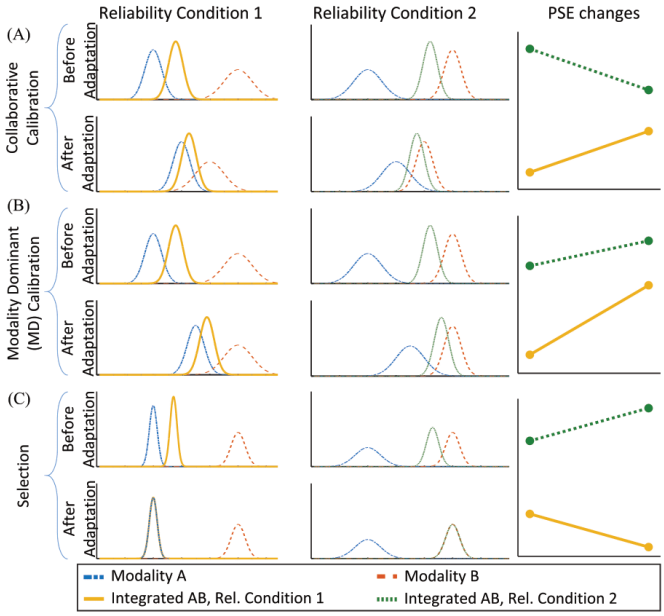

Figure 4. Simulation of different perception strategies in an environment of varying reliability. (**A**) stimulus S is perceived through two sensory modalities A (blue) and B (red) under two different reliability conditions. In the reliability condition 1, the reliability of modality A is higher than the reliability of modality (**B**) while it is vice versa in reliability condition 2. We simulated the variation of integrated PSEs in different adaptation strategies: (**A**) Collaborative calibration, (**B**) Modality Dominant (MD) calibration, and (**C**) Selection.

showed that the brain integrates unimodal estimations, $\hat{S}_{AB} = w_A \cdot \hat{S}_A + w_B \cdot \hat{S}_B$, in an optimal Bayesian fashion, where the summation of all individual weights (in this case the summation of w_A and w_B) equals 1. Facing an environment of varying reliability with two reliability conditions, two multisensory stimuli (yellow and green lines) are perceived with the same unimodal but different integrated estimations:

$$\begin{cases} \textit{reliability condition } 1\text{: } \hat{S}_{1AB} = w_{1A} \cdot \hat{S}_A + w_{1B} \cdot \hat{S}_B \\ \textit{reliability condition } 2\text{: } \hat{S}_{2AB} = w_{2A} \cdot \hat{S}_A + w_{2B} \cdot \hat{S}_B \end{cases} \quad (1)$$

where $\hat{S}_{iAB}$ denotes the integrated estimation of stimuli A and B in the i^{th} reliability condition. The difference between the integrated estimations, $\hat{S}_{iAB}$ and $\hat{S}_{2AB}$, in the two reliability conditions is due to the various reliabilities of unimodal stimuli. However, the unimodal estimations, $\hat{S}_A$ and $\hat{S}_B$, do not change across these reliability conditions. The weights are calculated as follow:

$$w_{iA} = \frac{\frac{1}{\sigma_{Ai}^2}}{\frac{1}{\sigma_{Ai}^2} + \frac{1}{\sigma_{Bi}^2}}, \; w_{iB} = \frac{\frac{1}{\sigma_{Bi}^2}}{\frac{1}{\sigma_{Ai}^2} + \frac{1}{\sigma_{Bi}^2}} \quad (2)$$

where σ_{Ai}^2 and σ_{Bi}^2 are the variance of the modality A and modality B in the i^{th} reliability condition.

Comparing the PSE variations in the simulation study and the observed PSE shifts informs us about the strategies taken by participants in the present psychophysical study. Assume that modality A receives the more reliable stimulus in reliability condition 1 and modality B receives the more reliable stimulus in reliability condition 2, similar to our experimental design. Our simulation shows how the integrated PSEs change dependent upon different adaptation strategies. As discussed before, we assume that three different strategies might occur: Calibration, selection, and no change in the perception. Here, we simulated the selection behavior as well as two modes of calibration. We did not simulate the case when there is no change in the behavior since the PSEs simply would not change. The simulated strategies are as follow: (A) Collaborative calibration: assumes that both of the conflicting modalities shift toward each other in a calibration process. (B) Modality dominant (MD) calibration: one modality is calibrated according to the other modality, so one modality remains steady, called the supervisor

modality, and the other modality adapts to the supervisor modality. (C) Selection: the perceptual system selects the most reliable modality, instead of integrating both inputs.

The discrepancy between the modalities is denoted by $\Delta = \hat{S}_A - \hat{S}_B$ before the adaptation. In a calibration process the unimodal estimators, $f_A(S)$ and $f_B(S)$, are updated in order to decrease the conflict Δ. The collaborative calibration, as well as the MD calibration, were defined as follow:

$$Collaborative\begin{cases} f_A^{New} = f_A^{Old} - \Delta C_A \\ f_B^{New} = f_B^{Old} + \Delta C_B \end{cases}, \quad MD\begin{cases} f_A^{New} = f_A^{Old} - \Delta C_A \\ f_B^{New} = f_B^{Old} \end{cases} \tag{3}$$

where f_j^{Old} is the old estimator, f_j^{New} is the updated estimator of the j^{th} modality, and C_j is the calibration coefficient of the j^{th} modality. In our simulations, C_A and C_B are equal and positive values in the collaborative calibration while C_B is zero in the MD calibration. After the adaptation, the multisensory stimuli were calculated according to Equation (3) in the collaborative calibration and MD calibration models.

By comparing the simulation and the experimental result (see Figs 3 and 4), it becomes clear that calibration is not the sole strategy the brain takes to overcome a conflict, at least in a varying reliability multisensory environment. Experiment 1.a suggests that when the participants received conflicting visual-auditory stimuli in an environment of varying reliability, they started to select between the visual and the auditory stimulus (Fig. 3A), exhibiting the selection strategy (Fig. 4C). In this situation, the PSEs corresponding to the multimodal stimuli moved toward the location of the unimodal stimuli during the experiment, each one toward the most reliable unimodal stimulus. The integrated PSE of the LV-MA condition (red line) moved toward the auditory source. However, the integrated PSE of the HV-MA condition (blue line) reflected the visual source. This pattern is consistent with the view that participants selected the auditory stimulus in the low reliability visual (LV-MA), and the visual stimulus in the high reliability visual (HV-MA) condition. Undoubtedly, this trend in the changes of the PSEs is inconsistent with each of the two calibration processes and is best described by a selection strategy.

The results of Experiment 1.b correspond to the Modality Dominant calibration strategy (Fig. 4B), suggesting that vision was calibrated by touch which is also consistent with previous findings[35,41]. According to the results (Fig. 3C), the PSEs in both the HV-MA-LT (purple line) condition and the LV-MA-HT (green line) condition moved toward the tactile source, meaning that touch affected both reliability conditions, probably by calibrating vision.

Furthermore, the three models were quantitatively compared to experimental results. To this end, we minimized the sum of the absolute error between empirical and predicted PSEs using the genetic algorithm[42]. For the data of Experiment 1.a, this sum was 0.01 for the selection model, 0.12 for the MD calibration model, and 1.55 for the collaborative calibration model. Therefore, the results of the Experiment 1.a are best described by the selection model. Likewise, for Experiment 1.b this sum was 0.34 for the selection model, 0.18 for the MD calibration model, and 1.06 for the collaborative calibration model. Accordingly, MD calibration was the dominant strategy in Experiment 1.b.

Merging the integration model and the selection model. Our experimental results together with the simulation results provide evidence that participants rely on both selection and calibration strategies when facing conflict stimuli in an environment of varying reliability. We have fitted a perceptual model to experimental data that enables us to assess the probability of different perception strategies in experiment 1.a and experiment 1.b.

According to previously suggested models[19], causal inference in multisensory perception is performed in line with the model averaging approach[16,17]. In the model averaging approach, the optimal estimate is a weighted average of two estimates: one derived under the assumption that two stimuli originate from different sources, the other derived under the assumption that two stimuli originate from the same source.

We extended the model averaging approach in order to compare the probability of the common cause between experiment 1.a and experiment 1.b. For the sake of simplicity, the prior probability of all unimodal stimuli is assumed to be uniform in our study. According to the model averaging approach, the perception in an environment of varying reliability with two reliability conditions would be as follows:

$$\begin{cases} reliability\ condition\ 1\text{: } \hat{S}_1 = P_C \cdot \hat{S}_{1AB} + (1 - P_C) \cdot \hat{S}_A \\ reliability\ condition\ 2\text{: } \hat{S}_2 = P_C \cdot \hat{S}_{2AB} + (1 - P_C) \cdot \hat{S}_B \end{cases} \tag{4}$$

where P_C denotes the probability of the common cause, $\hat{S}_i$ is the averaged estimation in the i^{th} reliability condition, and $\hat{S}_{1AB}$ and $\hat{S}_{2AB}$ are Bayesian integrated stimuli which are calculated according to Equation (1). $\hat{S}_A/\hat{S}_B$ is assumed to be the most reliable stimulus in reliability condition 1/2. The ideal observer integrates $\hat{S}_A$ and $\hat{S}_B$ with the probability P_C, and selects the most reliable stimulus between $\hat{S}_A$ and $\hat{S}_B$ with the probability $(1-P_C)$.

We have fitted the model of averaging to the average performance of participants. In the experiment 1.a, we assumed P_{C1}/P_{C2} (probability of common cause in the first/second half of the experiment), S_{V1}/S_{V2} (location of visual stimulus in the first/second half of the experiment), and S_{A1}/S_{A2} (location of auditory stimulus in the first/second half of the experiment) are free parameters. In the experiment 1.b, S_{T1}/S_{T2} (location of tactile stimulus in the first/second half of the experiment) were also added to the list of free parameters. The slope of the unimodal perceptions (which were measured in experiment 1.c) was used as the reliability of the individual modalities in the averaging model. We have minimized the sum of the square of the difference between the estimated and observed PSEs using the genetic algorithm (GA)[42]. The values of P_{C1} and P_{C2} were limited between 0 and 1 in the minimization procedure. We imposed the experimental perceptual reliabilities on the model and let the model fit the other parameters to the experimental data. Therefore, comparison between P_{C1} and P_{C2} shows how the model

	P_{C1}	P_{C2}	S_{V1}	S_{V2}	S_{A1}	S_{A2}	S_{T1}	S_{T2}	Fitting error
Experiment 1.a	0.56	0.18	−0.18	0.00	2.41	3.97	NA	NA	<0.0000
Experiment 1.b	0.49	0.47	−0.20	0.44	2.39	1.23	2.85	3.99	<0.0000

Table 1. Fitted parameters in experiments 1.a and 1.b.

predicts the change of strategy. Furthermore, comparison of location of the individual modalities between the first and the second half would reveal the effect of calibration on the perception.

Table 1 illustrates the fitted parameters in experiment 1.a and 1.b where the fitting error were minimum. The model of the experiment 1.a shows a decrease in the P_C. It also illustrates that the perception of the auditory stimulus was shifted to the auditory location while the perception of the visual stimulus was almost unchanged. The fitted parameters in experiment 1.a are in favor of changing the strategy from integration to selection, specifically since P_C decreased from first half to second half. In contrast, the P_C was almost constant in experiment 1.b while location of the individual modalities varied between the first half and second half. This pattern is more consistent with the calibration strategy rather than selection strategy. The model proposed that the visual and tactile modalities were shifted toward the tactile location while the auditory modality was shifted toward the location of the visual and auditory stimuli. The possible account for auditory shift is that the biased perception of the auditory stimulus in experiment 1.a affected the perception of the auditory in experiment 1.b. Thus, the auditory stimulus was shifted back to the unbiased location in experiment 1.b.

The parameters of the fitted models confirm the findings which were provided by simulation to some extent. However, the modeling reveals that the probability of the integration was about fifty percent at the beginning of experiments 1.a and 1.b. Thus, participants did not integrate the modalities perfectly. This imperfect integration may explain why the reliability of the multisensory perceptions were not always greater than the reliability of the unimodal perceptions as is expected in the Bayesian integration model. The fitted model predicted that the location of the auditory stimulus was shifted in experiment 1.a. Therefore, participants adjusted their auditory perception while they altered their strategy. It seems that the perception of the auditory stimulus was affected by the discrepancy at the beginning of the experiment when the probability of a common cause of the auditory stimulus and visual stimulus was higher. However, the participants seem to have adjusted their perception of the auditory stimulus later when they decided not to integrate the information.

Even though the results show that participants used the selection strategy in experiment 1.a and the calibration strategy in experiment 1.b, other perceptual processes seemed to be involved in both experiments. Therefore, further investigation is required to rule out the effect of other processes and to confirm our findings.

Model averaging explains the optimality in conflict solving strategy. We have investigated the multisensory perception of contradictory information in an environment of varying reliability and showed that participants use different strategies when facing different situations. However, the rationale for using different strategies in experiments 1.a and 1.b was not yet clear. Therefore, we investigated why it might be optimal to use different strategies in these two experiments. To this end, an ideal observer was modeled that minimized the perceptual estimation error in an environment of varying reliability based on the model averaging approach (equation 4). Since all reliability conditions were mixed and presented randomly, the ideal observer should infer the causality in such a way to minimize the overall error across all reliability conditions. Therefore, the overall estimation error, $Error_E$, for the two reliability conditions is defined as follows:

$$Error_E = [P_C(S - \hat{S}_{1AB}) + (1 - P_C)(S - \hat{S}_A)]^2 + [P_C(S - \hat{S}_{2AB}) + (1 - P_C)(S - \hat{S}_B)]^2 \quad (5)$$

Assuming that P_C is not changing in each trial, an optimal estimation of the common cause probability, P_C, is the one which minimizes the $Error_E$ (see the supplementary material for the proof and derivation):

$$P_C = \frac{\hat{S}_A + \hat{S}_B - 2S}{\Delta(1 - w_{1A} - w_{2A})} \quad (6)$$

According to Equation (6), increasing the discrepancy, Δ, leads to decrease in P_C; that is, the larger the discrepancy, the more probable that selection strategy is chosen by an ideal observer, which is consistent with previous studies investigating multisensory perception[19]. Furthermore, minimizing the error function in an environment of varying reliability reveals a term we call discrepancy weight,

$$w_\Delta = (1 - w_{1A} - w_{2A}), \quad (7)$$

This discrepancy weight shows the gain effect of the discrepancy (weight of the effect of the discrepancy) on P_C. Since $w_{iA} = 1w_{iB}$, the discrepancy weight, w_Δ, depends on the relative reliability of both stimuli in both reliability conditions. Consequently, the ideal observer chooses between the selection or integration strategies not only according to the amount of the discrepancy Δ, but also based on the relative reliability of the stimuli in all reliability conditions, represented by discrepancy weight w_Δ.

Similar to Equations (5) and (6), the optimal P_C and the discrepancy weight for three modalities and three reliability conditions, as in experiment 1.b, was derived as follows (see the supplementary material for error function, derivation and the proof).

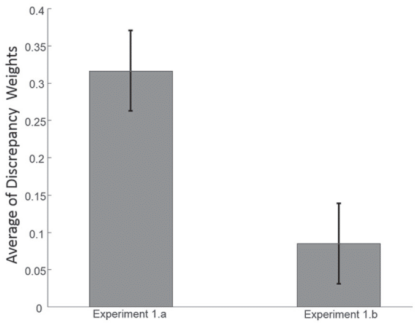

Figure 5. Discrepancy Weights. The discrepancy weights depend on the relative reliability of the cues across the different reliability conditions. Discrepancy weights are inversely correlated to the probability of the common cause P_C. Thus, the lower discrepancy weight in the experiment 1.b indicates a higher probability of an assumed common cause, which is consistent with the observed behavior in that experiment. In contrast, the higher discrepancy weight in the experiment 1.a points to a lower common cause probability that explains the selection strategy in experiment 1.a.

$$P_C = \frac{\hat{S}_A + \hat{S}_B + \hat{S}_C - 3S}{\Delta(2 - w_{1A} - w_{2A} - w_{3A} - w_{1B} - w_{2B} - w_{3B})} \tag{8}$$

$$w_\Delta = (2 - w_{1A} - w_{2A} - w_{3A} - w_{1B} - w_{2B} - w_{3B}) \tag{9}$$

Equation (9) indicates the gain effect of the discrepancy on P_C for three modalities across three reliability conditions.

Discrepancy weights explain the dissimilarity in perception strategies. Since the amount of the discrepancy (Δ) was the same in experiment 1.a and experiment 1.b, we hypothesize that the discrepancy weight is the source of taking different multimodal perception strategies. To investigate this, we compared the w_Δ in experiment 1.a and in experiment 1.b using Equation (7) and Equation (9) respectively. The values of $\hat{\boldsymbol{\sigma}}$ in Equation (2) are obtained by fitting psychometric functions to the results of the unimodal experiment (experiment 1.c). Figure 5 illustrates the average of the discrepancy weights over all participants in experiments 1.a and 1.b. The paired-sample t-tests showed a significant difference, $t(17) = 2.14$, $p = 0.047$, suggesting that the dissimilarity in taking different strategies is related to discrepancy weight (excluding one outlier whose weights outside of the range average $\pm 3\sigma$). The lower discrepancy weights in experiment 1.b points to the higher probability of the common cause P_C in that experiment. Therefore, the participants preferred to integrate the information in the visual-tactile-auditory experiment and tended to solve the conflict by calibration. In contrast, the discrepancy weight is higher in the visual-auditory experiment (experiment 1.a), and therefore the probability of the common cause is lower than in the visual-tactile-auditory experiment. As a result, the participants preferred to use the selection strategy to decrease the overall error in the visual-auditory experiment.

Evidence for conscious adaptation in multisensory environment. Confidence reporting is a metacognitive process which represents the cognition about one's own judgments[32]. In multisensory perception, it illustrates how much an individual is aware of his/her perceptual performance when coming across a multisensory stimulus. The confidence ratings together with slopes of the psychometric functions provide a measure of the participants' awareness of their performance[43–45]. In order to investigate the degree of correspondence between the self-reported confidence and the perceptual performance, we partitioned the confidence ratings into high confidence and low confidence trials. The high confidence trials include all trials in which the reported confidence level was equal to or greater than 3 and the low confidence trials consist of the trials with confidence level equal to or less than 2. For multisensory sessions (experiments 1.a and 1.b), a 2×5 (high/low confidence level $\times$ five reliability conditions) within-subjects ANOVA showed that there is a main effect of confidence level on the slope of the fitted psychometric functions, $F(1,18) = 7.89$, $p = 0.012$, as well as a significant effect of reliability condition, $F(4,72) = 13.63$, $p < 0.001$. However, no significant interaction effect was observed, $F(4,72) = 1.93$, $p = 0.115$. The same result was obtained for the unimodal experiment (experiment 1.c). The 2×5 within-subjects ANOVA again showed a main effect of confidence level on the slope, $F(1,18) = 12.35$, $p = 0.002$, as well as a significant effect of reliability condition on the slope, $F(4,72) = 8.45$, $p < 0.001$, but no significant interaction effect, $F(4,72) = 1.38$, $p = 0.249$. The psychometric functions for each experiment were plotted separately for high and low confidence trials (see Fig. 6). If participants were aware of their discrimination performance, better performance (a steeper slope) should be observed in high confidence trials than in low confidence trials. This expectation was clearly

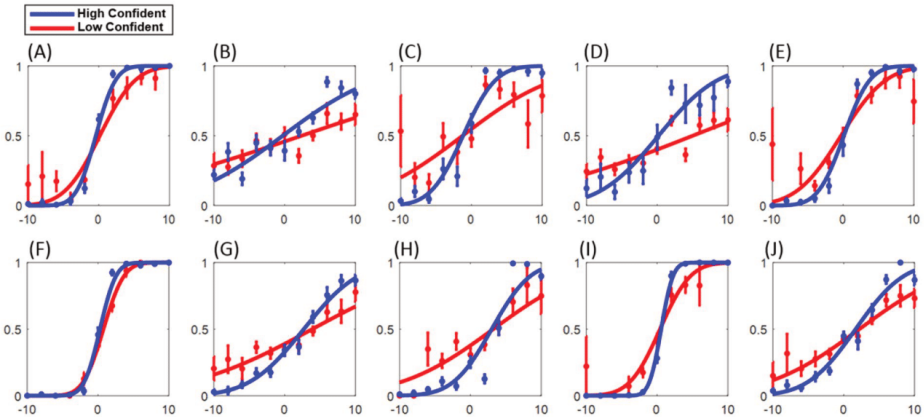

Figure 6. The Performance of overall subjects in high confidence and low confidence groups: (**A**)- Unimodal, high reliability visual. (**B**)- Unimodal, low reliability visual. (**C**)- Unimodal, high reliability tactile. (**D**)- Unimodal, low reliability tactile. (**E**)- Unimodal, medium reliability auditory. (**F**)- Multimodal high reliability visual-medium reliability auditory. (**G**)- Multimodal, low reliability visual-medium reliability auditory. (**H**)- Multimodal, high reliability tactile-low reliability visual-medium reliability auditory. (**I**)- Multimodal, low reliability tactile-high reliability visual-medium reliability auditory. (**J**)- Multimodal, low reliability tactile-low reliability visual-medium reliability auditory. The error bars show the standard deviation for running the bootstrap with 1000 iterations.

confirmed by the above analysis. The corresponding variation between the perceptual performance and the confidence reporting is seen for both unimodal (Fig. 6A–E) and multimodal judgments (Fig. 6F–J).

Furthermore, the variation of the confidence and the PSE in experiment 1.b (Fig. 3C,D) favor conscious calibration; that is, the changes of PSEs regarding the calibration in experiment 1.b were accompanied by a corresponding variation in confidence, thus providing evidence for conscious calibration. However, there was no significant variation in the confidence of experiment 1.a, in which participants used the selection strategy.

Conclusion. The brain employs multiple perceptual processes such as integration of multimodal stimuli, causal inference, and calibration in order to achieve a coherent perception of the complex and dynamic environment. Even though many studies have investigated and modeled these perceptual processes in multisensory environments, multisensory perception has never been studied when the reliability of the sensory modalities varies within a single session. We investigated the multisensory perception of contradictory information in such an environment of varying reliability. Particpiants were tested in a visual-auditory experiment, where they were asked to discriminate the direction of conflicting visual-auditory stimuli with two different reliability conditions. They also performed the same task in a visual-tactile-auditory experiment with three different reliability conditions.

Our results demonstrate that participants initially started to integrate the information, but later they changed their perception strategy in order to overcome the conflict. They chose the selection strategy in the visual-auditory experiment and the calibration strategy in the visual-tactile-auditory experiment. To understand the rationale for using different strategies, we modeled the perceptual mechanisms in an environment of varying reliability and also provided an ideal observer model. Our model suggests that causal inference in an environment of varying reliability depends not only on the amount of discrepancy but also on the reliability of stimuli across all reliability conditions. Thus, the rationale for using different strategies is probably mediated by the difference in weights of stimuli across reliability conditions. We also investigated the participants' awareness during the experiment by analyzing their confidence in their judgments. Their performance was better in high confidence trials than in the low confidence trials, indicating that participants were aware of their performance levels during the experiment. Moreover, participants did not change their confidence in Experiment 1.a, when they switched from integration behavior to the selection behavior. However, they changed their confidence in Experiment 1.b in which they calibrated their modalities.

In conclusion, the present study demonstrates that humans employ various strategies in a multimodal environment of varying reliability to cope with inconsistent information. Modeling of the results obtained suggests that it is optimal to utilize different strategies based on the amount of inconsistent sensory information and relative cue reliability. The results also indicate that humans engage consciously in these various perceptual strategies in such an environment.

Figure 7. Experimental setup. Participants received tactile stimuli through a vibrotactile belt around their abdomen. The belt consists of 14 minatory vibration motors, which are fixed horizontally on the belt. Running a vibration motor in a specific location on the belt causes a tactile stimulus to a subject. They simultaneously received auditory stimuli through a headphone and visual stimuli through a monitor.

Methods and Apparatus

Participants. 20 healthy male subjects (21.61 ± 0.63 years old) participated in the experiment. The data of one participant was excluded from data analysis because his PSE was outside average $\pm 3\sigma$. None of the participants reported having any history of neurological disorders. They all reported normal or corrected to normal vision, normal hearing, normal tactile sense, and no neurological problems.

All procedures and experimental protocols are approved by the ethical committee board of the University of Tehran, Tehran, Iran. All methods were carried out in accordance with the approved guidelines. A written informed consent was also obtained from all participants prior to data collection and they were compensated 10000 Rials per hour (approx. 3 USD) for their participation.

Stimuli and Materials. Participants were seated in an armchair in a sound-attenuated room in front of a 19″ LCD screen with 60 Hz refresh rate, on which the visual stimuli were presented. The distance between participants' eyes and the screen was 50 cm, corresponding to approximately 40° visual angles.

Auditory stimuli were delivered to the participant through headphones (Sennheiser HD419). Tactile stimuli were delivered by a custom designed belt, which was tightened over the participant's abdomen under the bottom of the sternum. The belt, which is called vibrotactile belt, was designed in the Cognitive Robotics Lab at University of Tehran, see Fig. 7.

Visual stimuli. The visual stimulus was a random dot cloud made up of ten white and ten black dots against a grey background (#7d7d7d). The diameter of each dot was approximately 0.4 degrees. The average position of all dots typifies the position of the visual stimulus and the spread of the dots corresponds to the reliability of the visual stimulus. The more wide-spread the dots, the lower the reliability of the stimulus. A standard deviation of 0.1 of the total screen size (4 degrees of the visual angle) was used to produce the high reliability visual stimuli while a standard deviation of 0.4 of the screen size (16 degrees of the visual angle) was used for the low reliability stimuli.

Auditory stimuli. The auditory stimulus was a white noise sound played through stereo headphones. The Head Related Transform Function (HRTF) simulated the position of the origin of the sound. The auditory stimulus was generated by SLAB which is an open source real-time virtual acoustic environment rendering system developed by the Spatial Auditory Displays Lab at NASA Ames Research Center. The intensity of the auditory stimulus was approximately 70 db (SPL). The reliability of the auditory stimuli did not change in the experiment.

Tactile stimuli. The vibrotactile belt consists of 14 vibration motors which are fixed horizontally on a cotton canvas tape. Running a vibration motor in a specific location on the belt causes a tactile stimulus to the subject. The reliability of the tactile stimuli can be adjusted by a separated PWM (Pulse Width Modulation) level for each vibration motor. The vibration motors are controlled by a microcontroller with a 32 MHz clock. According to a pilot test, the values 100%, and 60% were used as the PWM levels to produce high reliable and low reliable vibrations.

Experimental Procedure. The experiment included two multisensory and one unimodal spatial perception tasks which consisted of delivering visual, auditory and tactile stimuli to the participants. Each of these sections consisted of trials which only differed in the reliability of the information provided to the stimulated modalities. For example, in one trial, a high reliability visual stimulus may have accompanied a medium reliability auditory stimulus, while in the next trial a low reliability visual stimulus may have accompanied a medium reliability auditory stimulus. This setup, therefore, provided an opportunity to investigate the mechanisms underlying multisensory integration under more realistic conditions than previous setups since it provides an environment of varying reliability.

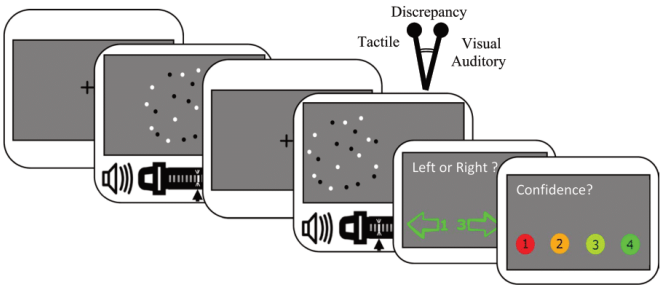

Figure 8. Time course of an exemplary trial: Subjects were presented a fixation screen for 800 ms, and then the first multimodal stimulus for 100 ms. Afterwards, another fixation screen was presented for 600 ms, and the second multimodal stimulus with a cue-conflict (spatial discrepancy of −4° visual angle) was presented for 100 ms. Subjects were asked to report whether the second stimulus was to the left or right of the first one. They also reported their confidence on a scale of 1 (low confidence) to 4 (high confidence).

In multimodal trials, the stimuli from the different modalities were presented simultaneously. In each trial, first a fixation cross was presented for 800 ms. Afterward, two consecutive stimuli were presented, each for 100 ms, separated by an interval of 600 ms, in which again a fixation cross was presented (Fig. 8). After the presentation of the stimuli, the subject was asked if the second stimulus (i.e., the comparison stimulus) was perceived on the left or the right of the first stimulus (i.e., the standard stimulus). The standard stimulus was always presented at the central position, and the location of the comparison stimulus varied between 11 spatial points evenly distributed around fixation along the horizontal axis. In multisensory trials, the judgment should be based on a combined impression from all stimulated senses. After judging the relative position, the participant was asked to report their confidence about the judgment, on a scale of 1 (low confidence) to 4 (high confidence), by pressing the corresponding number on the numeric keypad. The inter-trial interval was 800 ms for all trials.

Experiment 1.a. Experiment 1a was a visual-auditory spatial perception session. In each trial of this section, the first visual stimulus was always presented at the fixation cross. The second visual stimulus could be presented at 11 different stimulus locations distributed along the horizontal axis on the screen (the distance between fixation and the most left/rightwards presented stimulus was 10° visual angle), with a higher density at the middle of the range, where spatial discrimination should be harder. Each visual stimulus was accompanied by an auditory stimulus. There was no cue conflict between vision and audition for the first stimulus pair (standard stimulus). For the second stimulus pair (comparison stimulus), the auditory stimulus was always presented at a cue-conflict of −4 ° angle from the visual stimulus. There were two sets of trials with two different reliability conditions: In the high-reliability set, visual stimuli were presented at a high reliability and auditory stimuli at a medium reliability (**HV-MA**), and in the low-reliability set, visual stimuli were presented with low reliability and auditory with medium reliability (**LV-MA**). Inside each set, there were 115 trials, covering a symmetric range of 11 spatial points for spatial locations. These points were adjusted to be denser in the middle of the range, where differentiating between the first and the second stimulus in a pair becomes more challenging (15 trials for 5 midpoints). In order to avoid any bias resulting from the order of the trials, both trial sets were randomly mixed.

Experiment 1.b. The second section was a visual-tactile-auditory spatial perception session, which was completed right after the first section with no break. As in the first section, it consisted of several sets of trials, with varying reliability of the visual and/or tactile stimulation. Auditory stimuli were always presented with medium reliability. For the standard stimulus, the visual, tactile, and auditory stimuli were presented concurrently at the central position. For the comparison stimulus, auditory stimuli were presented at the same spatial position as the visual stimuli, while tactile stimuli were presented at a conflict angle of −4° degrees visual angle from the audiovisual ones. Specifically, three different stimulus sets were presented: **LV-MA-HT:** visual (low reliability), auditory (medium reliability), tactile (high reliability). **HV-MA-LT:** visual (high reliability), auditory (medium reliability), and tactile (low reliability). **LV-MA-LT:** visual (low reliability), auditory (medium reliability), tactile (low reliability). Similar to the first section, there were 115 trials in each trial set, covering a symmetric range of 11 spatial locations. All three trial sets were randomly mixed. After the second section, the subject had a break for about 5 minutes.

Experiment 1.c. The third section consisted of 575 unimodal trials, comprising five sets of 115 trials each. The stimuli specification in each set, as well as the order of the presentation, was as follows: **1st set:** auditory (medium reliability), **2nd set:** tactile (high reliability), **3rd set:** tactile (low reliability), **4th set:** visual (high reliability), **5th set:** visual (low reliability). The 2nd and the 3rd set were presented together and in random order, just as the 4th and the 5th set.

Practice trials. Before the experiment, all participants performed a number of practice trials with instructions provided by an experimenter. The practice section consisted of 50 trials, but it was terminated as soon as the participant felt familiarized with the task. The practice section included only visual-auditory trials.

References

1. van Atteveldt, N. M., Formisano, E., Blomert, L. & Goebel, R. The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cereb. Cortex* **17**, 962–74 (2007).
2. Navarra, J. *et al*. Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. *Brain Res. Cogn. Brain Res.* **25**, 499–507 (2005).
3. Senkowski, D., Talsma, D., Grigutsch, M., Herrmann, C. S. & Woldorff, M. G. Good times for multisensory integration: Effects of the precision of temporal synchrony as revealed by gamma-band oscillations. *Neuropsychologia* **45**, 561–71 (2007).
4. Drugowitsch, J., DeAngelis, G. C., Klier, E. M., Angelaki, D. E. & Pouget, A. Optimal multisensory decision-making in a reaction-time task. *Elife* **3** (2014).
5. Rowland, B. A., Quessy, S., Stanford, T. R. & Stein, B. E. Multisensory integration shortens physiological response latencies. *J. Neurosci.* **27**, 5879–84 (2007).
6. Diederich, A. & Colonius, H. Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Percept. Psychophys.* **66**, 1388–404 (2004).
7. Drugowitsch, J., DeAngelis, G. C., Angelaki, D. E. & Pouget, A. Tuning the speed-accuracy trade-off to maximize reward rate in multisensory decision-making. *Elife* **4**, e06678 (2015).
8. Ernst, M. O. & Banks, M. S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429–33 (2002).
9. Knill, D. C. & Saunders, J. A. Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Res.* **43**, 2539–2558 (2003).
10. Butler, J. S., Smith, S. T., Campos, J. L. & Bülthoff, H. H. Bayesian integration of visual and vestibular signals for heading. *J. Vis.* **10**, 23 (2010).
11. Yuille, A. & Bülthoff, H. H. Bayesian decision theory and psychophysics in *Perception as Bayesian inference* 123–162 (Cambridge University Press, 1996).
12. Pouget, A., Beck, J. M., Ma, W. J. & Latham, P. E. Probabilistic brains: knowns and unknowns. *Nat. Neurosci.* **16**, 1170–8 (2013).
13. Ter Horst, A. C., Koppen, M., Selen, L. P. J. & Medendorp, W. P. Reliability-Based Weighting of Visual and Vestibular Cues in Displacement Estimation. *PLoS One* **10**, e0145015 (2015).
14. Fetsch, C. R., Turner, A. H., DeAngelis, G. C. & Angelaki, D. E. Dynamic reweighting of visual and vestibular cues during self-motion perception. *J. Neurosci.* **29**, 15601–12 (2009).
15. Burge, J., Girshick, A. R. & Banks, M. S. Visual-haptic adaptation is determined by relative reliability. *J. Neurosci.* **30**, 7714–21 (2010).
16. Kayser, C. & Shams, L. Multisensory causal inference in the brain. *PLoS Biol.* **13**, e1002075 (2015).
17. Shams, L. & Beierholm, U. R. Causal inference in perception. *Trends Cogn. Sci.* **14**, 425–32 (2010).
18. Ernst, M. O. & Di Luca, M. Multisensory Perception: From Integration to Remapping in Sensory Cue Integration 224–250 (2011).
19. Körding, K. P. *et al*. Causal inference in multisensory perception. *PLoS One* **2**, e943 (2007).
20. Woods, A. J., Lehet, M. & Chatterjee, A. Context modulates the contribution of time and space in causal inference. *Front. Psychol.* **3**, 371 (2012).
21. Roach, N. W., Heron, J. & McGraw, P. V. Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. *Proc. Biol. Sci.* **273**, 2159–68 (2006).
22. Ernst, M. O. Learning to integrate arbitrary signals from vision and touch. *J. Vis.* **7**, 7.1–14 (2007).
23. Gori, M., Sandini, G., Martinoli, C. & Burr, D. Poor haptic orientation discrimination in nonsighted children may reflect disruption of cross-sensory calibration. *Curr. Biol.* **20**, 223–5 (2010).
24. Rock, I. & Victor, J. Vision and Touch: An Experimentally Created Conflict between the Two Senses. *Science (80-.)* **143**, 594–596 (1964).
25. Bertelson, P. & Radeau, M. Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Percept. Psychophys.* **29**, 578–584 (1981).
26. Burge, J., Girshick, A. R. & Banks, M. S. Visuo-haptic adaptation: the role of relative reliability. *J. Vis.* **7**, 67–67 (2010).
27. Morgan, M. L., Deangelis, G. C. & Angelaki, D. E. Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron* **59**, 662–73 (2008).
28. Zaidel, A., Turner, A. H. & Angelaki, D. E. Multisensory calibration is independent of cue reliability. *J. Neurosci.* **31**, 13949–62 (2011).
29. Bruns, P. & Röder, B. Sensory recalibration integrates information from the immediate and the cumulative past. *Sci. Rep* **5**, 12739 (2015).
30. de Gardelle, V. & Mamassian, P. Does confidence use a common currency across two visual tasks? *Psychol. Sci.* **25**, 1286–8 (2014).
31. Faivre, N., Mudrik, L., Schwartz, N. & Koch, C. Multisensory integration in complete unawareness: evidence from audiovisual congruency priming. *Psychol. Sci.* **25**, 2006–16 (2014).
32. Proof, O. U. *Behavioral methods in consciousness research*. 101–102 (OUP Oxford, 2015).
33. Vlassova, A., Donkin, C. & Pearson, J. Unconscious information changes decision accuracy but not confidence. *Proc. Natl. Acad. Sci. USA* **111**, 16214–8 (2014).
34. Morey, R. D. Confidence intervals from normalized data: A correction to Cousineau. *Reason* **4**, 61–64 (2005).
35. Gori, M., Sciutti, A., Burr, D. & Sandini, G. Direct and indirect haptic calibration of visual size judgments. *PLoS One* **6**, 1–5 (2011).
36. Gori, M., Giuliana, L., Sandini, G. & Burr, D. Visual size perception and haptic calibration during development. *Dev. Sci* **15**, 854–862 (2012).
37. Bruns, P., Spence, C. & Röder, B. Tactile recalibration of auditory spatial representations. *Exp. brain Res.* **209**, 333–44 (2011).
38. Zwiers, M. P., Van Opstal, A. J. & Paige, G. D. Plasticity in human sound localization induced by compressed spatial vision. *Nat. Neurosci.* **6**, 175–81 (2003).
39. Van der Burg, E., Alais, D. & Cass, J. Rapid recalibration to audiovisual asynchrony. *J. Neurosci.* **33**, 14633–7 (2013).
40. Wozny, D. R. & Shams, L. Recalibration of auditory space following milliseconds of cross-modal discrepancy. *J. Neurosci.* **31**, 4607–12 (2011).
41. Berkeley, G. *An essay towards a new theory of vision* (1709).
42. Houck, C. R. & Kay, M. G. A Genetic Algorithm for Function Optimization: A Matlab Implementation. *Ncsuie Tr* **95**, 1–14 (2008).
43. de Gardelle, V. & Mamassian, P. Does confidence use a common currency across two visual tasks? *Psychol. Sci.* **25**, 1286–8 (2014).
44. de Gardelle, V. & Mamassian, P. Weighting mean and variability during confidence judgments. *PLoS One* **10**, e0120870 (2015).
45. de Gardelle, V., Le Corre, F. & Mamassian, P. Confidence as a Common Currency between Vision and Audition. *PLoS One* **11**, e0147901 (2016).

Acknowledgements
We are grateful for the constructive comments of three anonymous reviewers. We also thank Donna Bryce for her help.

Author Contributions
All authors contributed in the main manuscript text. M.A.N.M. prepared figures. M.A.N.M. and S.S. collected the data. All authors contributed in the data analysis. M.A.N.M., M.N.A., and S.S. designed the study experiment and all authors reviewed the manuscript.

Additional Information
Supplementary information accompanies this paper at doi:10.1038/s41598-017-03521-2

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Study 2

The official citation that should be used in referencing this material is:

Mahani, Mohammad-Ali Nikouei, Karin Maria Bausenhart, Majid Nili Ahmadabadi, and Rolf Ulrich. "Multimodal simon effect: A multimodal extension of the diffusion model for conflict tasks." Frontiers in human neuroscience 12 (2019): 507.

doi: 10.3389/fnhum.2018.00507

The final publication is available here:

https://www.frontiersin.org/articles/10.3389/fnhum.2018.00507/full

Author contributions

All authors contributed in the main manuscript text. M-AM and KMB collected the data. M-AM and RU contributed to the data analysis. RU has designed the study experiment and all authors reviewed the manuscript.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 09 January 2019
doi: 10.3389/fnhum.2018.00507

Multimodal Simon Effect: A Multimodal Extension of the Diffusion Model for Conflict Tasks

Mohammad-Ali Nikouei Mahani[1,2]*, *Karin Maria Bausenhart*[1], *Majid Nili Ahmadabadi*[2] and *Rolf Ulrich*[1]

[1] Cognition and Perception, Department of Psychology, University of Tübingen, Tübingen, Germany, [2] Cognitive Systems Lab, School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran

In conflict tasks, like the Simon task, it is usually investigated how task-irrelevant information affects the processing of task-relevant information. In the present experiments, we extended the Simon task to a multimodal setup, in which task-irrelevant information emerged from two sensory modalities. Specifically, in Experiment 1, participants responded to the identity of letters presented at a left, right, or central position with a left- or right-hand response. Additional tactile stimulation occurred on a left, right, or central position on the horizontal body plane. Response congruency of the visual and tactile stimulation was orthogonally varied. In Experiment 2, the tactile stimulation was replaced by auditory stimulation. In both experiments, the visual task-irrelevant information produced congruency effects such that responses were slower and less accurate in incongruent than incongruent conditions. Furthermore, in Experiment 1, such congruency effects, albeit smaller, were also observed for the tactile task-irrelevant stimulation. In Experiment 2, the auditory task-irrelevant stimulation produced the smallest effects. Specifically, the longest reaction times emerged in the neutral condition, while incongruent and congruent conditions differed only numerically. This suggests that in the co-presence of multiple task-irrelevant information sources, location processing is more strongly determined by visual and tactile spatial information than by auditory spatial information. An extended version of the Diffusion Model for Conflict Tasks (DMC) was fitted to the results of both experiments. This Multimodal Diffusion Model for Conflict Tasks (MDMC), and a model variant involving faster processing in the neutral visual condition (FN-MDMC), provided reasonable fits for the observed data. These model fits support the notion that multimodal task-irrelevant information superimposes across sensory modalities and automatically affects the controlled processing of task-relevant information.

Keywords: conflict processing, simon task, multimodal congruency effect, diffusion model for conflict tasks (DMC), reaction time, multisensory processing

OPEN ACCESS

Edited by:
Juliana Yordanova, Institute of Neurobiology (BAS), Bulgaria

Reviewed by:
Ling Wang, South China Normal University, China
Roberta Sellaro, Leiden University, Netherlands

***Correspondence:**
Mohammad-Ali Nikouei Mahani
nikouei@ut.ac.ir

Received: *23 March 2018*
Accepted: *05 December 2018*
Published: *09 January 2019*

Citation:
Mahani M-AN, Bausenhart KM, Ahmadabadi MN and Ulrich R (2019) Multimodal Simon Effect: A Multimodal Extension of the Diffusion Model for Conflict Tasks. Front. Hum. Neurosci. 12:507. doi: 10.3389/fnhum.2018.00507

INTRODUCTION

People sometimes need to suppress task-irrelevant information to minimize interference with processing of task-relevant information. Standard examples for the empirical investigation of such situations are conflict tasks, such as the Stroop task, the Eriksen-Flanker task, and the Simon task (Stroop, 1935; Simon and Wolf, 1963; Eriksen and Eriksen, 1974). For instance, in the Simon task

participants are instructed to respond to a non-spatial stimulus attribute (e.g., color, form, letter, or pitch) with a spatially defined response (e.g., a key press of the left or the right hand). Although the location of the stimulus presentation is task-irrelevant, it influences task performance. Specifically, responses are faster and more accurate when both the stimulus and the response are on the same spatial side (congruent condition) rather than on different sides (incongruent condition) (Simon and Wolf, 1963). Such congruency effects have been reported not only for the visual but also for other modalities (Simon and Rudell, 1967; Simon et al., 1970; Cohen and Martin, 1975; McClain, 1983; Jerger et al., 1988; Medina et al., 2014; Salzer et al., 2014). Thus, the effects of task-irrelevant information are not limited to a single modality but rather reveal a general phenomenon that presumably emerges from an amodal processing mechanism.

It has been suggested that this mechanism involves two separate processes acting simultaneously on stimulus input. More specifically, it is assumed that task-relevant information is processed by a controlled process, whereas task-irrelevant information is mediated by an automatic process. Moreover, these two processes are often assumed to operate in parallel and in independent pathways (Logan, 1980; Coles et al., 1985; Cohen et al., 1990; Hommel, 1993; Ridderinkhof, 2002). A recent quantitative account of this processing architecture is provided by an elaborated diffusion process model, called the Diffusion Model for Conflict Tasks (Ulrich et al., 2015). DMC is based on standard diffusion models according to which a decision-making process accumulates noisy task-relevant information until one of two decision boundaries is hit and the corresponding response is selected (Stone, 1960; Ratcliff, 1978; Ratcliff and Smith, 2004). This accumulation process is typically modeled as a standard Wiener diffusion process (Ratcliff, 1978). DMC extends this framework by adding a second process to incorporate processing of task-irrelevant information, which leads to a short-lived activation. This automatically triggered activation superimposes the activation from the controlled process, which operates on the task-relevant information. The superimposed activation from both processes triggers either the correct or the incorrect response. DMC can predict various phenomena associated with common conflict tasks, including reaction time (RT) patterns, the shape of conditional accuracy functions, and the shape of delta functions (Ulrich et al., 2015; Ellinghaus et al., 2018; White et al., 2018). Furthermore, DMC has been successfully linked to neurophysiological findings (Servant et al., 2016).

DMC's core assumption of independently operating controlled and automatic processes receives particular support from studies demonstrating that task-relevant and task-irrelevant information needs not to stem from the same stimulus source. Specifically, task-irrelevant information may impede performance even if it stems from a different modality than the task-relevant information. For example, task-irrelevant congruent auditory information decreases RT to visual stimuli compared to task-irrelevant incongruent auditory information (Simon and Craft, 1970; Donohue et al., 2013; Schupak et al., 2015). Similar cross-modal conflict effects have been reported for visual-tactile conflict tasks (Kennett et al., 2001; Spence et al., 2004; Yue et al., 2009; Wesslein et al., 2014; Poole et al., 2015). These results strengthen the notion that automatic and controlled processes act independently. Nevertheless, these results are limited to single sources of task-irrelevant information. Hence the question is, whether DMC accounts for the processing of multiple sources of task-irrelevant information as well.

The goal of the present study was thus to examine whether DMC can be extended to conflict tasks with two task-irrelevant information sources. Within DMC, it seems most reasonable that these conflicting sources are processed by two independent automatic processes, with each process generating separate activation and superimposing the controlled process. By orthogonally manipulating the compatibility of two independent task-irrelevant sources, it is possible to put this assumed superimposition within DMC to a comprehensive test.

To this end, we conducted two Simon task experiments with two rather than one task-irrelevant information sources (i.e., task-irrelevant location information was provided by two modalities). The first (second) experiment was a typical visual Simon task with additional task-irrelevant tactile (auditory) information. In both experiments, the spatial congruency of these task-irrelevant information sources (i.e., visual and tactile/auditory stimulus location) and the response side varied orthogonally. Specifically, stimulus location in both modalities could be congruent, neutral, or incongruent with the side of the correct response. DMC's architecture was extended to address the contribution of these two independent sources of automatic activation and the extended DMC was fitted to the experimental results of the two experiments. These experiments emulated the Simon task of the original DMC study (Ulrich et al., 2015) in order to facilitate the comparison of results between the present experiments and this former study.

EXPERIMENT 1

This experiment examines whether task-irrelevant tactile stimulation influences speeded decisions in a visual Simon task. The spatial position of the tactile stimulation was congruent, incongruent, or neutral with the required response to a letter appearing to the left, to the right, or at the central position of the fixation point.

Method

Participants

Thirty participants (26 women and 4 men) volunteered in this experiment (23.5 ± 3.5 years of age). They all reported normal or corrected-to-normal vision, normal tactile sense, and no neurological problems. All procedures and experimental protocols are approved by the ethical committee board of the University of Tehran and all methods were carried out in accordance with the approved guidelines. A written informed consent was also obtained from all participants prior to data collection and they either received 8 € per hour or course credit for their participation.

Stimuli and Apparatus

Participants were seated in a sound-attenuated room in front of a 19" CRT screen on which the visual stimuli were presented

(see **Figure 1A**). The distance between participants' eyes and the monitor screen was about 50 cm. Like in Ulrich et al. (2015), we employed letters as imperative stimuli. Specifically, these were the letter "H" and "S" (Font: Arial; letter size: 48 pt, ∼1.5° visual angle) which were presented in white color against a black background. One of these two letters appeared either to the left or to the right (8° visual angle) from of the center of the screen, or at the center. The tactile stimulus came from a custom canvas belt consisting of six vibration motors. Two motors were placed on the left, two at the center, and two on the right of the belt. Running vibration motors on the left/center/right side of the belt causes a tactile stimulus to the left/center/right side along the participant's horizontal body plane. Since participants varied in waist size, we designed two belts: a medium size belt and a large size belt. In the medium (large) size belt, the horizontal distance of motors was 16.5 (19.0) cm and the vertical distance was 1.3 (1.5) cm. The vibration motors were controlled by a XMEGA microcontroller with a 32 MHz clock, the same device as used in a previous study (Mahani et al., 2017). Stimulus durations of both visual and tactile stimuli were 200 ms. Oscilloscopic measurements were conducted to ensure simultaneity of the visual stimulation and the maximum vibration amplitude. Both the experiment and the microcontroller program were written in C++ language. Left and right responses were recorded with the "A" and "L" keys, respectively.

Procedure

Each trial of Experiment 1 started with the presentation of a fixation cross for 500 ms at the center of the screen (see **Figure 1B**). Then, the visual stimulus and the tactile stimulus were presented simultaneously for 200 ms. In both modalities, and independent of each other, stimuli were presented either on the left, on the right, or at the central position. For half of the participants, the stimulus "H" was associated with the left response key and "S" with the right response key. A reverse setting was used for the remaining participants. Participants were asked to ignore the location of the stimulation and to respond to the letter identity quickly within 1,500 ms, but also to avoid errors as much as possible. They received visual feedback for 1,000 ms when their RT was longer than 1,500 ms, or shorter than 150 ms, or if their response was wrong. The inter-trial delay was 1,000 ms.

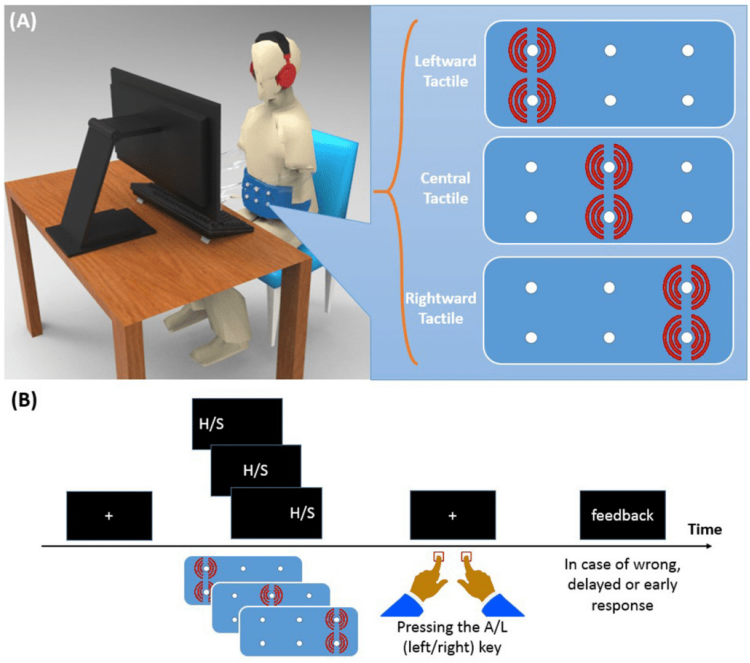

FIGURE 1 | (A) Experimental setup and vibrotactile belt. Running two vibration motors on the left/center/right side of the belt causes a tactile stimulus to the left/center/right side of the participant's waist. **(B)** Time course of a trial. A left/center/right visual stimulus was presented along with a left/center/right tactile stimulus. Participants were asked to identify the visual stimulus (H or S) with a left/right key press and to ignore the location of the visual stimulus.

Design

The combination of three visual positions (left, center, and right), three tactile positions (left, center, and right), and two letters ("H" and "S") resulted in 18 trial types. Each trial type was repeated five times per block, and trials were presented in random order. Overall, participants completed six blocks. Note that the side of visual and tactile stimulation could be either congruent (same side), incongruent (opposite side), or neutral (central position) to the side of the correct response. Thus, from the orthogonal combination of the two within-subject factors Visual Congruency and Tactile Congruency, nine different conditions emerged: (1) Congruent visual and congruent tactile stimulation CVCT, (2) congruent visual and neutral tactile CVNT, (3) congruent visual and incongruent tactile CVIT, (4) neutral visual and congruent tactile NVCT, (5) neutral visual and neutral tactile NVNT, (6) neutral visual and incongruent tactile NVIT, (7) incongruent visual and congruent tactile IVCT, (8) incongruent visual and neutral tactile IVNT, and (9) incongruent visual and incongruent tactile IVIT. Each participant received each of the nine congruency conditions 60 times.

Results

Trials with RTs $>$ 1,200 ms or $<$ 150 ms were discarded (0.68%) from data analysis. However, statistical results were virtually identical when we kept those trials. Separate 3 x 3 within-subject ANOVAs with factors *visual congruency* (congruent, neutral, and incongruent) and *tactile congruency* (congruent, neutral, and incongruent) were performed on RT and on response errors. **Figure 2** depicts mean RT and the percentage of response errors as a function of visual and tactile congruency.

Reaction Time

RT was significantly affected by visual congruency, $F_{(2,\ 58)} = 62.74$, $p < 0.001$, $\eta_p^2 = 0.68$, and tactile congruency, $F_{(2,\ 58)} = 13.94$, $p < 0.001$, $\eta_p^2 = 0.33$ (η_p^2 indicates the partial eta-squared). However, the interaction of the two factors was not significant, $F_{(4,58)} = 0.95$, $p = 0.43$, $\eta_p^2 = 0.03$. *Post-hoc* Tukey tests showed that RTs were longer in the visual incongruent than in the visual neutral ($p < 0.001$), and in the visual congruent ($p < 0.001$) conditions. No significant difference was observed between the visual neutral and visual congruent condition ($p = 0.20$). *Post-hoc* Tukey tests for the tactile congruency conditions showed that RTs were shorter in the tactile congruent than in the tactile neutral ($p < 0.001$), and in the tactile incongruent ($p < 0.001$) conditions. The difference between the tactile incongruent and tactile neutral condition was not significant ($p = 0.98$).

Response Error

There were also significant main effects of visual congruency $F_{(2,58)} = 29.02$, $p < 0.001$, $\eta_p^2 = 0.50$, and tactile congruency, $F_{(2,58)} = 7.93$, $p = 0.001$, $\eta_p^2 = 0.22$, on mean response error. The interaction of visual and tactile congruency was not significant, $F_{(4,58)} = 0.39$, $p = 0.81$, $\eta_p^2 = 0.01$. *Post-hoc* Tukey tests on visual congruency showed that the percentage of error was higher in the visual incongruent than in the visual neutral ($p < 0.001$), and in the visual congruent ($p < 0.001$) conditions. However, there was no difference between the visual neutral and the visual congruent condition ($p = 0.36$). The same Tukey test on tactile congruency revealed similar results; the percentage of errors was higher in the tactile incongruent than in the tactile neutral ($p = 0.016$), and in the tactile congruent ($p = 0.002$) conditions. The difference between the tactile congruent and tactile neutral conditions was not significant ($p = 0.90$).

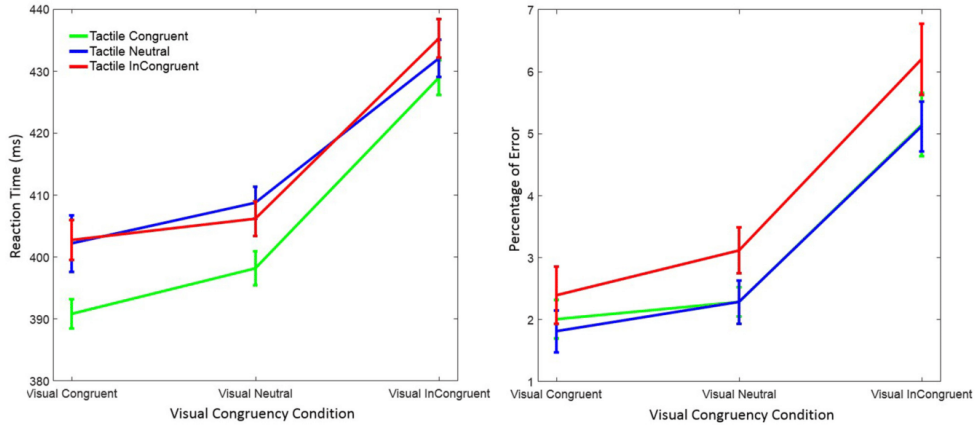

FIGURE 2 | Mean reaction time (left figure) and mean percentage of response errors (right figure) in Experiment 1 as a function of visual and tactile congruency. Error bars were computed according to Morey's method (Morey, 2008).

Distributional Analysis of Reaction Time

RT percentiles (10, 30, 50, 70, 90%) for each congruency condition and for each participant were estimated. Percentiles were analyzed by a three-way ANOVA with factors percentile, visual congruency, and tactile congruency. As one expects, the main effect of percentile was significant, $F_{(4,116)} = 228.74$, $p < 0.001$, $\eta_p^2 = 0.89$. There was also a significant main effect of tactile congruency, $F_{(2,58)} = 12.82$ $p < 0.001$, $\eta_p^2 = 0.31$, and visual congruency, $F_{(2,58)} = 64.91$, $p < 0.001$, $\eta_p^2 = 0.69$. The three-way visual congruency x tactile congruency x percentile interaction was significant, $F_{(16,\ 464)} = 1.76$, $p = 0.034$, $\eta_p^2 = 0.06$. **Figure 3** illustrates the cumulative distribution functions (CDFs) for all congruency conditions, as well as delta functions for the visual and tactile modalities. The CDF of each of the visual congruency conditions was averaged over all tactile congruency conditions (e.g., the visual congruent CDF is the average of the CVCT, CVNT, and CVIT conditions). The same approach was used to calculate CDFs for the tactile congruency conditions. Delta (Δ) functions show the percentile difference between the congruent and the incongruent condition for each modality.

Analysis of Conditional Accuracy Functions

Conditional accuracy functions (CAFs) depict response accuracy given response speed (**Figure 8**). As in previous investigations on conflict tasks, we have analyzed CAFs for each of the congruency conditions. All RTs of a given congruency condition were sorted from fastest to slowest. Thereafter, the RT distribution was split into five equal bins (0–20, 20–40, 40–60, 60–80, 80–100%) and the percentage of correct responses was calculated for each bin.

A three-way ANOVA with factors bin, visual congruency, and tactile congruency was used to analyze the CAFs. This analysis revealed a main effect of visual congruency, $F_{(2,\ 58)} = 12.30$, $p < 0.001$, $\eta_p^2 = 0.30$, and a main effect of tactile congruency, $F_{(2,\ 58)} = 4.00$, $p = 0.024$, $\eta_p^2 = 0.12$. However, the effect of bin on CAFs was not significant, $F_{(4,\ 116)} = 1.88$, $p = 0.12$, $\eta_p^2 = 0.06$. The three-way bin x visual congruency x tactile congruency

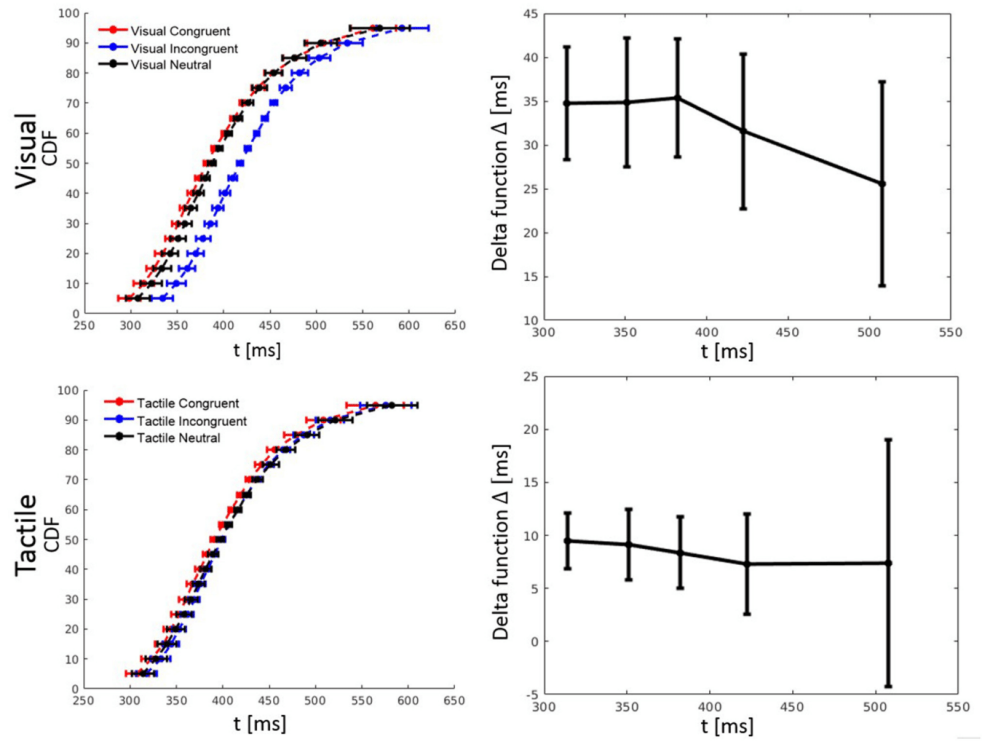

FIGURE 3 | Cumulative distribution functions (CDFs) and delta (Δ) functions for percentiles (5, 10, 15, . . ., 95%) in Experiment 1. Error bars show 95% confidence intervals and are calculated according to Morey (2008). Each of the visual (tactile) CDFs was calculated as the average over all tactile (visual) congruency conditions. For example, the visual congruent CDF is the average of CVCT, CVNT, and CVIT conditions. Delta functions show the difference between the congruent and incongruent CDFs as a function of response time.

interaction, $F_{(16, 464)} = 0.75$, $p = 0.73$, $\eta_p^2 = 0.03$, and all of the two-way interactions were not significant (all Fs < 0.85 and all ps > 0.56).

Discussion

We extended the Simon task to study the effect of simultaneous task-irrelevant tactile and task-irrelevant visual information on speeded visual decisions. The results show that both visual and tactile congruency significantly affected the task-relevant processing of letter identity. In general, visual and tactile incongruent stimulus locations produced longer RTs and more response errors than visual and tactile congruent stimulus locations, reflecting the typically expected pattern of results in the Simon task. In addition, *post-hoc* analyses showed that the effects regarding the neutral condition were not the same for the visual and tactile modalities. There was no significant difference between the visual congruent and the visual neutral condition in terms of both RT and response errors. Thus, only the visual incongruent information significantly increased RT and response errors. In contrast, there was no meaningful difference between the tactile incongruent and tactile neutral information in terms of RT while they had significantly different effects on the response errors. That is, for the tactile modality, the neutral condition was the same as the incongruent condition in terms of RT while it was the same as the congruent condition in terms of response errors.

EXPERIMENT 2

This experiment assesses the effect of task-irrelevant auditory stimulation instead of task-irrelevant tactile stimulation on letter processing performance in the visual Simon task. Task-irrelevant tones were presented to the left or to the right ear, or to both ears simultaneously. Otherwise the experimental setup was identical to the one in Experiment 1. Thus, this experiment examines whether similar multimodal effects would emerge as in Experiment 1, when the task-irrelevant tactile information is replaced by task-irrelevant auditory information.

Method

Participants

Thirty individuals (23.8 ± 3.0 years of age, 9 men and 21 women) participated in this experiment. They all reported normal or corrected-to-normal vision, and no neurological problems. All procedures and experimental protocols are approved by the ethical committee board of the University of Tehran and all methods were carried out in accordance with the approved guidelines. A written informed consent was also obtained from all participants prior to data collection. They either received 8 € per hour or course credit for their participation.

Apparatus and Procedure

In the second experiment, tactile stimuli were replaced by the auditory stimuli. The auditory stimuli came through Sony MDR-XD200 stereo headphones. The leftward (rightward) auditory stimulus was a mono sound provided to the left (right) ear and the central (neutral) stimulus was a stereo sound provided to both ears. The intensity of the mono and stereo stimuli were corrected using the binaural correction method (Epstein and Florentine, 2009) where the intensity of the mono stimulus was 75 dB (SPL) and the intensity of the stereo stimulus was 63 dB (SPL). The source of the auditory stimulus was a square wave with a frequency 440 Hz and a duration of 200 ms, that is, the same duration as the visual stimulus. The onsets of the visual and the auditory stimulus were synchronous. All other experimental details were identical to those of Experiment 1.

Results

As in Experiment 1, trials with RTs $>$ 1,200 ms or $<$ 150 ms were discarded (0.44%) from the data analysis. However, statistical results were virtually identical when these trials were included in the statistical analysis. **Figure 4** shows mean RT and percentage of response errors as a function of visual and auditory congruency.

Reaction Time

RTs were again analyzed using a within-subject ANOVA with factors *visual congruency* (congruent, neutral, and incongruent) and *auditory congruency* (congruent, neutral, and incongruent). As before, there was a significant main effect on RT of visual congruency, $F_{(2, 58)} = 80.76$, $p < 0.001$, $\eta_p^2 = 0.74$, as well as of auditory congruency, $F_{(2, 58)} = 4.19$, $p = 0.019$, $\eta_p^2 = 0.13$. The interaction of visual x auditory congruency on RT was not significant, $F_{(4, 58)} = 0.25$, $p = 0.90$, $\eta_p^2 = 0.01$. A *post-hoc* Tukey test on auditory congruency illustrated that the difference between auditory neutral and auditory congruent ($p = 0.021$) was significant. However, the difference between the auditory incongruent and auditory congruent ($p = 0.19$), as well as the auditory incongruent and auditory neutral ($p = 0.53$) was not significant. Tukey tests also showed that the difference between visual incongruent and visual congruent ($p < 0.001$), as well as visual incongruent and visual neutral ($p < 0.001$) was significant. Nevertheless, and as in Experiment 1, no significant difference between visual neutral and visual congruent conditions ($p = 0.15$) was observed.

Response Error

Response errors were analyzed with the same ANOVA design as for RT. There was a significant main effect of visual congruency on response errors, $F_{(2, 58)} = 26.72$, $p < 0.001$, $\eta_p^2 = 0.48$, while the effect of auditory congruency, $F_{(2, 58)} = 0.89$, $p = 0.41$, $\eta_p^2 = 0.03$, and the interaction of visual x auditory congruency, $F_{(4, 58)} = 1.69$, $p = 0.15$, $\eta_p^2 = 0.06$, were not significant. *Post-hoc* Tukey tests on visual congruency illustrated significant differences between the visual incongruent and visual neutral ($p < 0.001$), as well as the visual incongruent and visual congruent ($p < 0.001$) conditions. No difference in terms of response errors was observed between visual neutral and visual congruent ($p = 0.18$).

Distributional Analysis of Reaction Time

RT percentiles were estimated and analyzed as in Experiment 1 by a three-way ANOVA with factors percentile, visual congruency, and auditory congruency. The main effect of percentile, $F_{(4, 116)} = 292.56$, $p < 0.001$, $\eta_p^2 = 0.91$, the effect

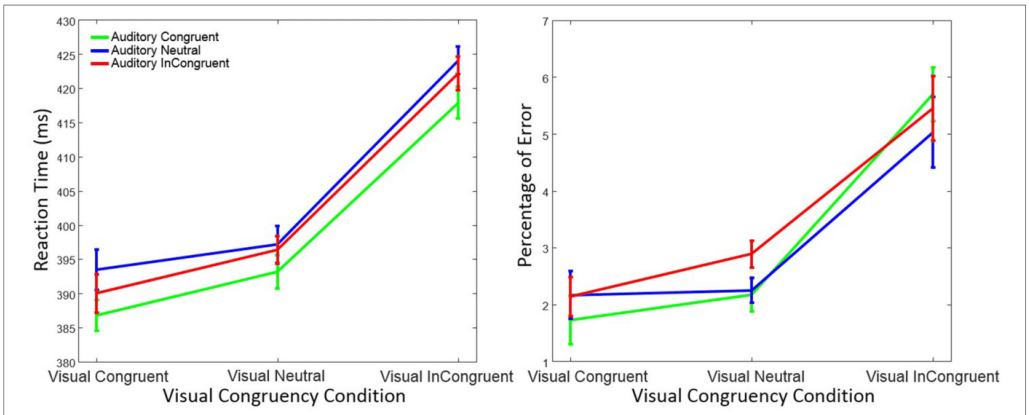

FIGURE 4 | Mean reaction time (left figure) and mean percentage of response errors (right figure) in Experiment 2 as a function of visual and auditory congruency. Error bars were computed according to Morey's method (Morey, 2008).

of auditory congruency, $F_{(2, 58)} = 5.48$, $p = 0.007$, $\eta_p^2 = 0.16$, and the effect of visual congruency, $F_{(2, 58)} = 90.36$, $p < 0.001$, $\eta_p^2 = 0.76$, were all significant. The three-way visual congruency x auditory congruency x percentile interaction, $F_{(16, 464)} = 0.38$, $p = 0.99$, $\eta_p^2 = 0.01$, and all of the two-way interactions were not significant. **Figure 5** shows the CDFs for congruent, neutral, and incongruent conditions for both the visual and auditory modality. It also illustrates how the delta function decreases with an increase of RT.

Analysis of Conditional Accuracy Functions

CAFs were calculated similarly to the first experiment. A three-way ANOVA with factors bin, visual congruency, and auditory congruency showed a significant effect of visual congruency, $F_{(2, 58)} = 8.23$, $p = 0.001$, $\eta_p^2 = 0.22$. In contrast, the effect of auditory congruency, $F_{(2, 58)} = 0.54$, $p = 0.59$, $\eta_p^2 = 0.02$, and bins, $F_{(4, 116)} = 0.48$, $p = 0.75$, $\eta_p^2 = 0.02$, were not significant. The visual congruency x auditory congruency interaction was significant, $F_{(4,116)} = 2.51$, $p = 0.046$, $\eta_p^2 = 0.08$. However, the three-way bin x visual congruency x tactile congruency interaction, $F_{(16, 464)} = 1.06$, $p = 0.40$, $\eta_p^2 = 0.04$, and all other two-way interactions were not significant.

Discussion

The second experiment investigated how simultaneous task-irrelevant visual and task-irrelevant auditory information affects visual decisions. Exactly as in Experiment 1, task-irrelevant visual information evoked pronounced congruency effects on RT and response errors. However, the effects of the task-irrelevant auditory information were less pronounced than the effects of tactile information in Experiment 1. Although there was a significant effect of auditory congruency on RT, further analysis showed that this effect was due to especially slow responses in the neutral compared to the congruent condition. No significant difference between the congruent and the incongruent auditory condition was observed. The results regarding response errors indicate that task-irrelevant auditory information did virtually not affect the accuracy of the visual decisions.

Taken together, task-irrelevant visual information significantly affected RT and response errors of the visual decisions, which is in line with the typical Simon effect and the results of the first experiment. The lack of congruency effect of the task-irrelevant auditory information presumably suggests that the influence of auditory spatial information on visual information processing is rather limited in the presence of visual spatial information. This is in line with many studies suggesting that the visual modality dominates the auditory one in processing spatial information (Howard and Templeton, 1966; Welch and Warren, 1980; Bertelson and Radeau, 1981; Slutsky and Recanzone, 2001).

Modeling

Similar to the previous models (Luce, 1986; Ulrich et al., 2015), total RT is assumed to be the sum of two parts (RT = D+R), that is, the duration of the decision process (D), and the duration of residual processes (R), which represent the duration of all processes besides the decision process. It is also assumed that the congruency of the stimuli only affects the duration of D and not of R. Within DMC, the decision process is modeled as a standard Wiener diffusion process. Specifically, the state $X(t)$ of the decision process at time is t regarded as a superimposed Wiener process, that is, $X(t) = X_c(t) + X_a(t)$, where $X_c(t)$ denotes a controlled process and $X_a(t)$ an automatic process. The superimposed process accumulates until it hits either the upper (correct) decision boundary ($b > 0$) or the lower (incorrect) decision boundary ($-b$).

According to the original version of DMC, the controlled process can be described by the following stochastic difference

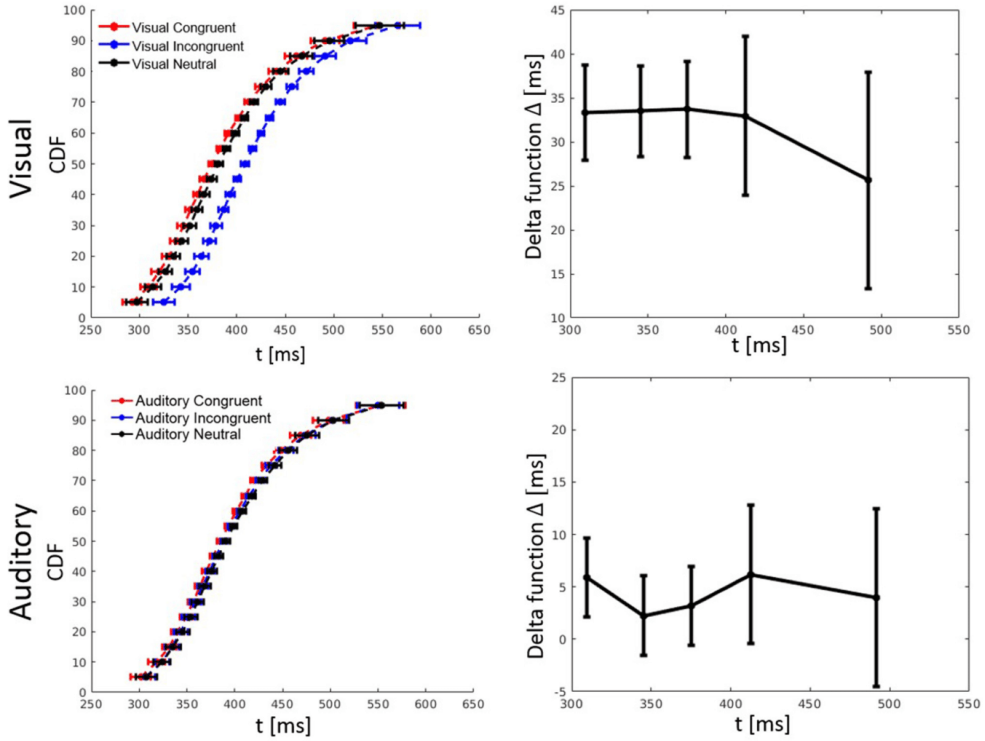

FIGURE 5 | Cumulative distribution functions (CDFs) and delta (Δ) functions for percentiles (5, 10, 15, . . ., 95%) in Experiment 2. Error bars show 95% confidence intervals and are calculated according to Morey (2008). Each of the visual (auditory) CDFs was calculated as the average over all auditory (visual) congruency conditions. For example, the visual congruent CDF is the average of CVCA, CVNA, and CVIA conditions. Delta functions show the difference between the congruent and incongruent CDFs as a function of response time.

equation

$$X_c(t+\Delta t) = X_c(t) + \mu_c(t) \cdot \Delta t + W_c(t) \cdot \sigma_c \cdot \sqrt{\Delta t} \quad (1)$$

where $X_c(t)$ denotes the state of the controlled process at time t. $W_c(t)$is the standard Wiener diffusion process (mean = 0, and variance = 1), σ_c indicates the diffusion constant, and $\mu_c(t)$ is the time-independent drift rate of the controlled process, that is, $\mu_c(t) = \mu_c$. Likewise, the automatic process is given by

$$X_a(t+\Delta t) = X_a(t) + \mu_a(t) \cdot \Delta t + W_a(t) \cdot \sigma_a \cdot \sqrt{\Delta t} \quad (2)$$

where $W_a(t)$ is a Wiener diffusion process, with diffusion constant σ_a. The drift rate of the automatic process $\mu_a(t)$ is time-dependent.

Here we extend DMC in order to fit the data from the multimodal Simon task studied in Experiments 1 and 2. In the multimodal DMC (MDMC), two (or more) automatic processes superimpose on the controlled process to form the decision process: $X(t) = X_c(t) + X_{a1}(t) + X_{a2}(t)$. $X_c(t)$ denotes again a standard Wiener diffusion process with the constant time-independent drift $\mu_c(t) = \mu_C$. The time course of an automatic process is modeled as a pulse-like rescaled Gamma distribution, $X_a(t)$, with shape parameter $a = 2$ and the free scale parameter τ. The parameter A corresponds to the maximum of this pulse-like function. Thus, the time course of the expected mean of the automatic process is given by (cf. Ulrich et al., 2015)

$$E[X_a(t)] = A \cdot e^{-\frac{t}{\tau}} \cdot \left[\frac{t \cdot e}{(a-1) \cdot \tau}\right]^{a-1} \quad (3)$$

and thus the time-dependent drift rate $\mu_a(t)$ of the automatic process is given by the first derivative of $E[X_a(t)]$ with respect to

time,

$$\mu_a(t) = A \cdot e^{-\frac{t}{\tau}} \cdot \left[\frac{t \cdot e}{(a-1)\cdot\tau}\right]^{a-1} \cdot \left[\frac{a-1}{t} - \frac{1}{\tau}\right]. \quad (4)$$

The parameters A and τ are estimated for each of the two automatic processes. **Figure 6** exemplifies the architecture of MDMC. The expected decision process $E[X(t)]$ (blue line) is modeled as the sum of the expected controlled process $E[X_c(t)]$ (red line) and two expected automatic processes $E[X_{a1}(t)]$ and $E[X_{a2}(t)]$ (black and green lines). A congruent automatic process is represented by a positive (i.e., $A > 0$) pulse-like function (e.g., **Figure 6A**: both of the automatic processes are congruent), and an incongruent automatic process is represented by a negative (i.e., $A < 0$) pulse-like function (e.g., **Figure 6D**: both of the automatic processes are incongruent). It is assumed that the neutral automatic process does not affect the decision process (**Figure 6B**: black line). The trial-to-trial variability of the starting point is modeled by random samples from a beta distribution, with the free parameter α, supported on the bounded interval [-b, b], where b is the decision boundary.

We fitted two variants of MDMC to the data, one as is described so far, and one with a faster processing of neutral visual information. Previous studies mentioned that a visual stimulus at the center of field of view (FOV) benefits from faster retinal processing in contrast to a stimulus presented to the left or to the right of the center (fixation point) (Osaka, 1976). For example, presenting a stimulus by 5–10° degree nasal or temporal from the fovea typically increases RT by 10–20 ms (Rains, 1963). This phenomenon motivates an extension of MDMC with a separate mean residual process time for neutral visual information. This version of the MDMC model is called FN-MDMC (Faster Neutral visual-Multimodal DMC).

Fitting Criteria

The fitting procedure was similar to the method described by Hübner (2014) and also Servant et al. (2016). The MDMC was fitted to the CAFs and the CDFs for each of the nine congruency

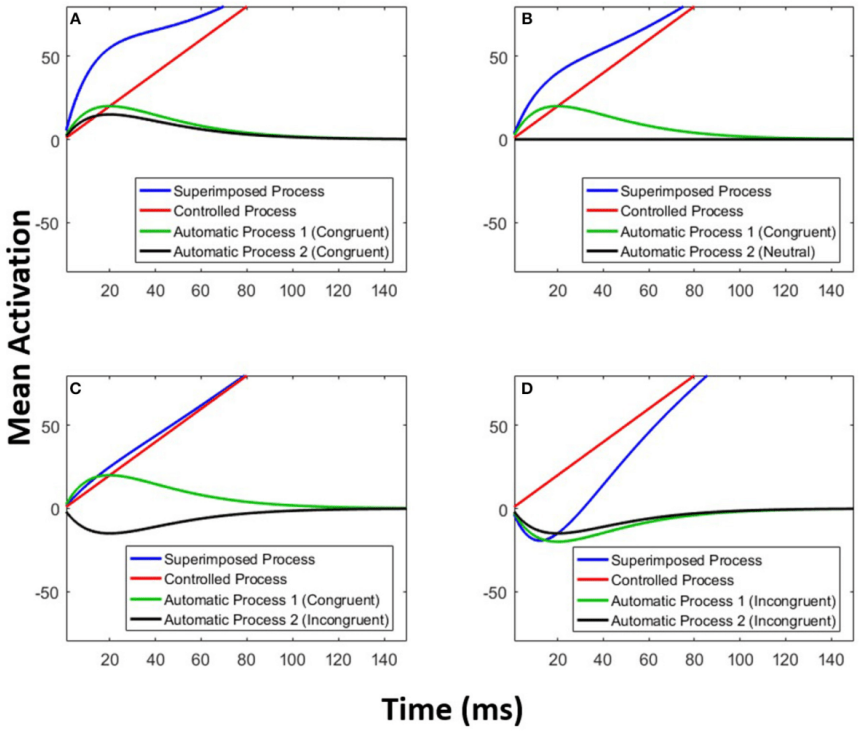

FIGURE 6 | Multimodal DMC. The decision process (blue line) is a superimposition of a controlled process (red line) and two automatic processes (green and black lines). **(A)** Both of the automatic processes are congruent. **(B)** The first automatic process is congruent and the second one is neutral. **(C)** The first automatic process is congruent and the second one is incongruent. **(D)** Both of the automatic processes are incongruent.

conditions. There were five CAF bins (0–20, 20–40, 40–60, 60–80, 80–100%), and five CDF quantiles (0.1, 0.3, 0.5, 0.7, 0.9) for each given congruency condition. MDMC predictions were generated using Monte Carlo simulations (Metropolis and Ulam, 1949) with a step size of $\Delta t = 1$ ms, and a constant diffusion constant of $\sigma = 4$ *ms* for the superimposed process, similar to Ulrich et al. (2015). The following function was employed to fit the model to the data

$$G^2 = 2\sum_{c=1}^{9} N_c \sum_{i=1}^{10} \left| p_{ci} \log\left(\frac{p_{ci}}{\pi_{ci}}\right)\right| \quad (5)$$

where p_{ci} and π_{ci} denote the observed and the predicted proportion of responses, respectively. The index c indicates the congruency condition, and the summation over the i includes both CAFs (five bins) and CDFs (five bins). N_c is the number of trials per congruency condition. Fifty thousand trials were simulated for each minimization call in each of the congruency conditions. The G^2 criterion was minimized using the MATLAB implementation of the SIMPLEX (Lagarias et al., 1998) method. Since SIMPLEX is sensitive to the choice of initial values, the fitting procedure was repeated with different sets of initial values in order to ensure the stability of the resulting estimates.[1]

Fitting MDMC

MDMC was fitted to the aggregated experimental data over all participants using the aforementioned criteria. We fitted the model to the averaged data of all participants because the data of individual participants are typically noisy and may be prone to outlier RTs. Especially if trial numbers are rather small, it is difficult to identify the best fitting model parameters for individual participant data. Even though a previous study showed a virtually negligible difference between fitting DMC to individual data and to average group data (Servant et al., 2016), it should be highlighted that fitting data to group averages neglects interindividual variability and thus may result in distortions of parameter estimates (e.g., Estes and Maddox, 2005; Cohen et al., 2008). Nevertheless, the results of model fits to individual data are available in the **Supplementary Material**. Moreover, raw data and the complete Matlab code for model fitting are available online via the Open Science Framework (Mahani, 2018).

Figures 7, **8** show the results for both CDFs and CAFs in all congruency conditions of both experiments. **Figures 9**, **10** also depict predicted delta functions of the FN-MDMC model for both experiments. In general, MDMC provides a reasonable fit of the experimental data. However, FN-MDMC fits slightly better than MDMC. Observing more errors for faster RTs is a common pattern in Simon tasks and this is especially bold in the incongruent visual conditions of the present experiments. MDMC captures this pattern relatively well, with only a few small deviations (cf. **Figure 8**). **Table 1** contains the estimated parameters for both the visual-tactile and the visual-auditory task, and for both variants of the model. This table also provides the average of G^2 for 1,000 simulations given the best parameters for each model. Similar to Servant et al. (2016), we compared MDMC and FN-MDMC by a *BIC* statistic that penalizes models based on the G^2 and number of free parameters f:

$$BIC = G^2 + f\ log \sum_{i=1}^{n} n_i \quad (6)$$

We compared the fits of MDMC and FN-MDMC using the paired-sample permutation test across 1,000 simulated G^2 and *BIC* values with 50,000 permutations. In the visual-tactile experiment, both the G^2 and *BIC* of FN-MDMC were significantly lower than G^2 and *BIC* of MDMC ($ps < 0.001$). The same result was obtained for the visual-auditory experiment, that is, G^2 and *BIC* of FN-MDMC were also significantly lower than G^2 and *BIC* of MDMC ($ps < 0.001$). **Table 1** shows the average of simulated G^2 and *BIC* values of the two experiments.

Note that μ_R and σ_R represent the mean and standard deviation of the residual process time, respectively. However, the mean of the residual process time for the neutral visual condition is given by μ_{RN} in the FN-MDMC model. α and b correspond to the shape and decision boundary of the starting point distribution, respectively. μ_C is the drift rate of the controlled process. A and τ are the parameters of the automatic process for each modality (see Equation 3). The only difference between MDMC and FN-MDMC is the addition of the parameter μ_{RN} in the latter case to enable a direct assessment of the effect of μ_{RN} on the goodness of fit. In both FN-MDMC models the shorter mean residual process time for the visual neutral condition (μ_{RN}) compared to the mean residual process time of the other congruency conditions (μ_R) results in a smaller fitting error (cf. **Table 1**). This result is consistent with the phenomenon that visual stimuli presented at the fovea benefit from faster processing and the size of this effect agrees with the typical speed benefit for foveal processing (Rains, 1963).

Table 1 also reveals that in all models the peak activation of the visual automatic process (A_V) is higher than the peak activation of the tactile/auditory automatic process ($A_{T/A}$). This result points to a relative dominance of visual stimuli over tactile/auditory stimuli (see **Figure 11**).

GENERAL DISCUSSION

Numerous studies have suggested that task-irrelevant information affects the task-relevant decision processes in speeded RT tasks. The standard Simon task assesses the influence of task-irrelevant information on the processing of task-relevant information within the visual modality. In the present study, we investigated whether additional task-irrelevant information from the tactile modality (Experiment 1) or from the auditory modality (Experiment 2) would also influence the processing of visual information. The experiments were theoretically motivated by an elaboration of DMC, which assumes that task-irrelevant information from different sense modalities superimpose. Specifically, this elaboration assumes that the contribution of task-irrelevant information from one modality does not affect the contribution of task-irrelevant information from the other modality. MDMC further assumes that this superimposed information spills over to the decision

[1]To validate this parameter estimation procedure MDMC data were simulated with the estimated parameters reported in **Table 1**. Then, the recovery of the original parameters from these simulated data was assessed.

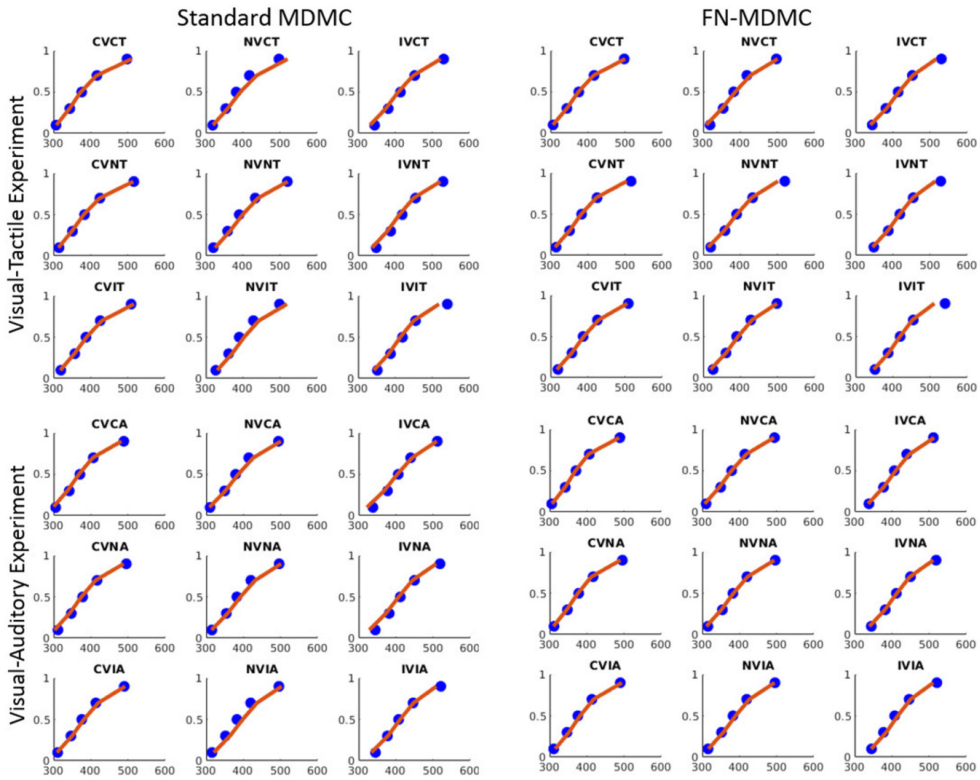

FIGURE 7 | Experimental data and model predictions of CDFs for both experiments. Blue dots show the experimental data and red lines show the model predictions. Both models provide a reasonable fit of the experimental data, however, FN-MDMC fits slightly better than MDMC.

process. The MDMC provided a reasonable account for the results of the two experiments. As expected, the results of the two experiments revealed the classical Simon effect (i.e., a task-irrelevant influence of spatial visual stimulus position on RT and response errors). In Experiment 1 we also observed the influence of task-irrelevant tactile stimulation on RT and response errors of visual decisions. In Experiment 2 the task-irrelevant auditory information affected the RT of visual decisions, but not the response errors. Furthermore, there was no difference between the auditory congruent and auditory incongruent conditions, pointing to an unreliable effect of auditory stimulus location on the RT of visual decisions. Moreover, the observed delta functions, especially for the visual congruency conditions, are negative-going, thus indicating that the congruency effect decreases with increasing reaction time. Such negative-going delta functions have been repeatedly reported for the Simon task (for an overview, see Schwarz and Miller, 2012).

Our findings also corroborate the robust phenomenon showing that task-irrelevant spatial visual information affects visual decisions. Even though there is a large number of studies on the effects of task-irrelevant information on non-visual decisions (MacLeod, 1991; Lu and Proctor, 1995; Dolk et al., 2014), so far, no one studied the effects of simultaneous task-irrelevant tactile and visual information on non-spatial visual decisions. However, several studies reported cross-modal effects of touch on visual perception (Macaluso et al., 2002; Diederich et al., 2003; Ossandón et al., 2015). Therefore, we expected to observe an influence of task-irrelevant tactile information on RT and response errors for visual decisions, and the results of Experiment 1 are consistent with these expectations.

In Experiment 2, however, the lack of a clear effect of auditory stimulation on visual decisions was rather unexpected in the light of previous studies (Simon and Craft, 1970; Donohue et al., 2013; Schupak et al., 2015). For example, Simon and

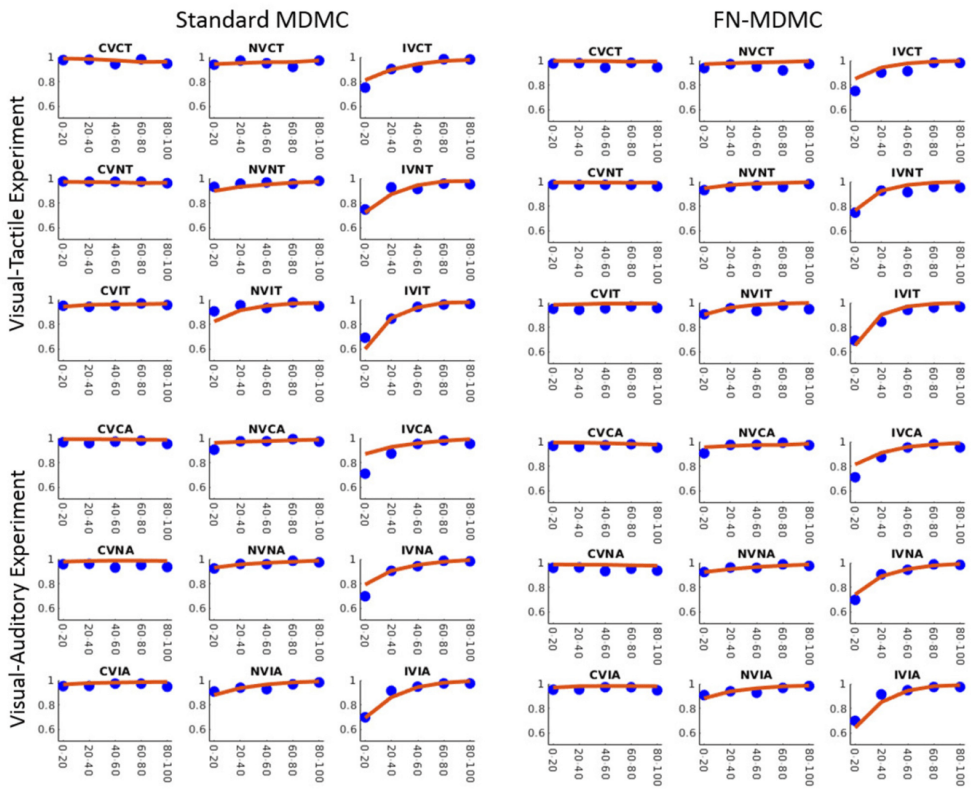

FIGURE 8 | Observed results and model predictions of CAFs for both experiments, across all congruency conditions, and for both variants of the model. Blue dots show the experimental data and red lines show the model predictions. The model appropriately predicts the experimental data except for small proportions of the incongruent visual conditions.

Craft showed that task-irrelevant auditory information can influence visual decisions in a Simon task. However, in this study, auditory stimulation was not accompanied by simultaneous task-irrelevant visual stimulation, as in the present Experiment 2. In fact, several other studies have reported the lack of or small effects of task-irrelevant auditory information on visual decisions in co-presence of both visual and auditory spatial information (Howard and Templeton, 1966; Welch and Warren, 1980; Bertelson and Radeau, 1981; Slutsky and Recanzone, 2001). These observations suggest that the effect of task-irrelevant visual information on visual decisions is much stronger than the effect of task-irrelevant auditory information when simultaneous visual and auditory stimulation is provided. Thus, the relatively small effect of task-irrelevant auditory stimulation in the present Experiment 2 might be attributed to the fact that the auditory stimulus did not carry task-relevant information and was accompanied by visual-spatial stimulation. This is also reflected in the fitted parameters of the MDMC and FN-MDMC, as a relatively small peak of the automatic activation corresponding to the auditory compared to the visual stimulation (cf. **Figure 11**, bottom row, and **Table 1**).

The performance of participants in the neutral conditions revealed a rather surprising pattern. Intuitively, one might expect that the mean RT in the neutral condition is just the average of the RTs in the congruent and incongruent conditions, if the influences of inhibition and facilitation are equally effective. Contrary to this expectation, neither RT nor response errors did significantly differ between the visual neutral and visual congruent conditions. Interestingly, MDMC assumes that the effects of inhibition and facilitation on the decision process are symmetrical (i.e., automatic activation in incongruent trials favors the wrong response to the same amount as automatic

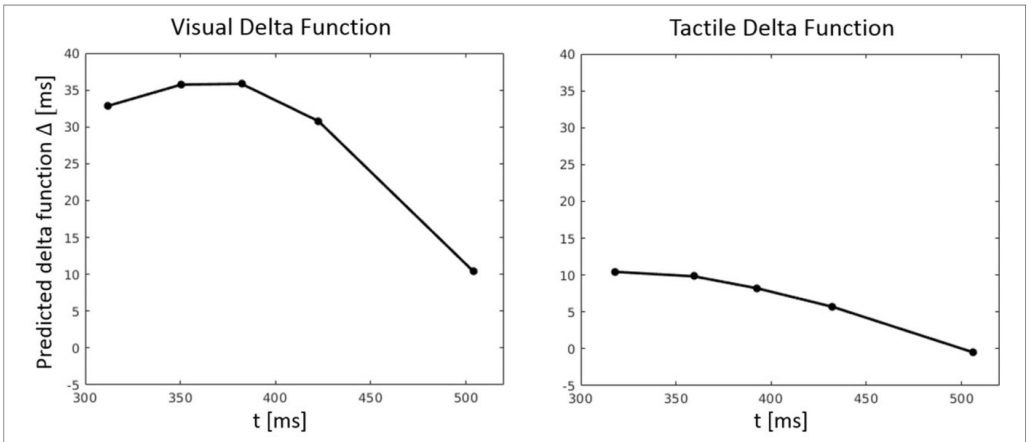

FIGURE 9 | Predicted delta (Δ) functions by FN-MDMC for the visual-tactile experiment. Delta functions show the difference between the congruent and incongruent CDFs as a function of response time.

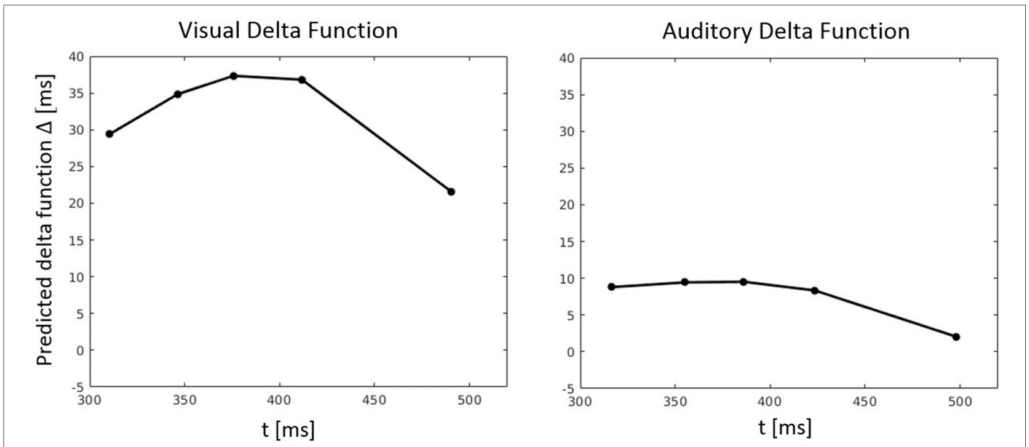

FIGURE 10 | Predicted delta (Δ) functions by FN-MDMC for the visual-auditory experiment. Delta functions show the difference between the congruent and incongruent CDFs as a function of response time.

TABLE 1 | Parameter estimates for the model fit of MDMC and FN-MDMC to the results of the visual-tactile (V-T, Experiment 1) and visual-auditory (V-A, Experiment 2) tasks.

Task	Model	μ_R	μ_{RN}	σ_R	α	b	μ_C	A_V	τ_V	$A_{T/A}$	$\tau_{T/A}$	$\overline{G^2}$	$\overline{BIC}$
V-T	MDMC	313	–	33.4	3.1	54.6	0.52	13.4	39.0	6.1	28.5	126.1	176.5
	FN-MDMC	317	303	36.8	2.9	64.1	0.66	19.8	38.5	7.1	28.2	114.5	170.5
V-A	MDMC	311	–	40.5	2.7	57.7	0.62	15.1	51.4	6.5	35.5	136.2	186.6
	FN-MDMC	315	302	35.3	3.5	55.6	0.58	16.7	46.6	5.3	32.0	98.6	154.6

Unit of measurement for μ_R, μ_{RN}, σ_R is the millisecond (ms), while the unit of μ_C is $\frac{1}{ms}$.

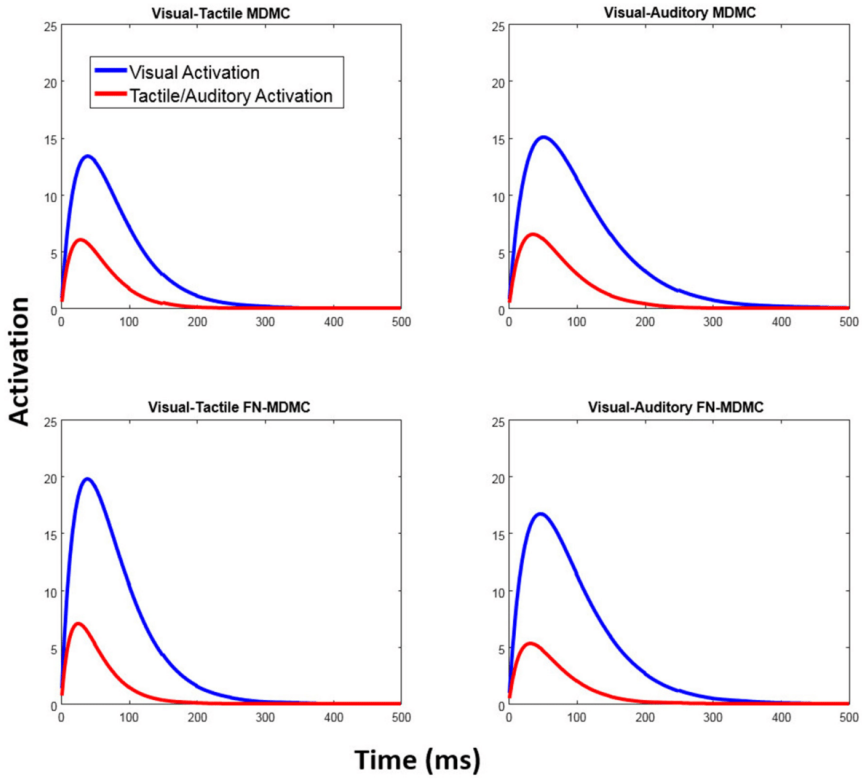

FIGURE 11 | Automatic activation processes of the fitted models. In both models, the peak activation of the visual automatic process is higher than the peak activation of the automatic tactile/auditory process and thus reflects the relatively strong influence of visual-spatial task-irrelevant information.

activation in congruent trials favors the correct response). Nonetheless, it can be shown that this symmetry need not necessarily manifest itself at the level of mean RT. It must be admitted, however, that the deviation from symmetry was so large that it cannot be captured quantitatively by MDMC.

There is at least one explanation for this asymmetry effect. One may generally refute the idea that it is possible to introduce a true neutral condition in conflict task paradigms in order to reveal the contributions of interference and facilitation, as previous studies with such baseline conditions suggest (Simon and Acosta, 1982). For example, a neutral stimulus presented at the fixation point may benefit from retinal processing in contrast to stimuli presented in the periphery, that is, to the left or to the right of the fixation point (Slater-Hammel, 1955; Osaka, 1976). Within MDMC, this would simply mean that the residual process operates faster in the visual neutral condition than in both the congruent and incongruent ones. Hence, we have investigated this proposal by extending the MDMC to FN-MDMC. The FN-MDMC indeed provides a better model fit than the standard MDMC and, as one might expect, corroborates a faster residual process for neutral visual information. Specifically, the FN-MDMC reveals that when a visual stimulus is presented at the fovea, processing time is ∼10–15 ms faster than in the periphery. This finding is consistent with simple reaction time results from a previous study (Rains, 1963).

The asymmetrical effect produced by tactile stimulation in Experiment 1 is probably more surprising than the aforementioned asymmetrical congruency effect in the visual modality. Here we observed that RTs in the neutral tactile condition were not significantly different from those in the incongruent tactile condition, although response errors were about the same as in the congruent tactile condition. One can only speculate about the reasons for this surprising pattern of results. One reason may be that tactile stimulation along the

body's median sagittal plane takes more time to process than along the body's horizontal plane. Accordingly, the residual process within MDMC should take more time for central than for peripheral tactile stimuli. Unfortunately, this account cannot address the difference in response errors. Another speculation is that there is a tradeoff between speed and accuracy within the tactile modality, which seems difficult to address within the present version of MDMC. Thus, providing a comprehensive interpretation of the tactile neutral condition is difficult. However, the results of the tactile neutral condition show that tactile stimulation cannot be ignored even if it provides task-irrelevant, modality-irrelevant, and neutral information.

In the present work, MDMC was fitted to average data. Model fits to individual data are presented in the **Supplementary Material**. The parameter estimates of both approaches are reasonably similar. Nevertheless, we preferred model fits to averaged data in the present case in order to reduce not only the computational complexity and effort, but also to minimize the influence of spurious responses that may render individual datasets noisy. Future efforts should be directed toward overcoming these limitations, for example, with the Approximate Bayesian Computation approach (Turner and Van Zandt, 2018).

In conclusion, the present study examined crossmodal congruency effects within the classical visual Simon task. In Experiments 1 and 2, the spatial position of task-irrelevant tactile and auditory stimulation, respectively, varied orthogonally with the spatial position of the relevant visual information. MDMC provided a reasonable account of the observed RT data and response errors. This model suggests that task-irrelevant activation combines additively across modalities before the summed automatic activation spills over to the processing of task-relevant information. MDMC's predictions, however, were suboptimal with regard to the neutral conditions. One reason for this suboptimal prediction is that the neutral conditions may not provide an ideal baseline for assessing the respective contributions of facilitation and inhibition through congruence and incongruence within the Simon task, a conclusion that receives support from other experimental work. In fact, the model fit was improved by an extension of MDMC, which incorporates faster residual processing time for foveally presented (neutral) visual stimuli than for peripherally presented (congruent and incongruent) stimuli. Importantly, this model extension acknowledges potential differences in processing latency according to stimulus location within the visual field, but does not change our main conclusion that superimposed automatic activation from multiple task-irrelevant information sources may overlap controlled stimulus processing. Therefore, MDMC offers a novel framework for understanding such multisensory processing in conflict tasks and thus advances our understanding of how information from different sensory modalities is processed and integrated, which is a core issue in neurocognitive sciences (e.g., Miller, 1982; Stein and Stanford, 2008).

AUTHOR'S NOTE

This study was conducted in the laboratory of RU while the first author was on sabbatical leave at the Department of Psychology, University of Tübingen. Raw data and Matlab code used for model fitting are available via Open Science Framework (Mahani, October 29, 2018, doi: 10.17605/OSF.IO/EWSJD) and via GitHub (https://github.com/manmahani/mdmc/). Correspondence concerning this article should be addressed to Mohammad-Ali Nikouei Mahani, Department of Psychology, University of Tübingen, Schleichstrasse 4, 72076 Tübingen, Germany, email address: nikouei@ut.ac.ir.

AUTHOR CONTRIBUTIONS

All authors contributed in the main manuscript text. M-AM and KMB collected the data. M-AM and RU contributed to the data analysis. RU has designed the study experiment and all authors reviewed the manuscript.

FUNDING

This research was partially supported by the German Research Foundation, DFG, BA 4110/3-2 and Open Access Publishing Fund of University of Tübingen.

ACKNOWLEDGMENTS

We acknowledge support by Deutsche Forschungsgemeinschaft (BA 4110/3-2) and Open Access Publishing Fund of University of Tübingen. We also thank Sarah-Helena Lutz for assistance in data collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2018.00507/full#supplementary-material

REFERENCES

Bertelson, P., and Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Percept. Psychophys.* 29, 578–584. doi: 10.3758/BF03207374

Cohen, A. L., Sanborn, A. N., and Shiffrin, R. M. (2008). Model evaluation using grouped or individual data. *Psychon. Bull. Rev.* 15, 692–712. doi: 10.3758/PBR.15.4.692

Cohen, G., and Martin, M. (1975). Hemisphere differences in an auditory Stroop test. *Percep. Psychophys.* 17, 79–83. doi: 10.3758/BF03204002

Cohen, J. D., Dunbar, K., and McClelland, J. L. (1990). On the control of automatic processes: a parallel distributed processing account of the Stroop effect. *Psychol. Rev.* 97, 332–361. doi: 10.1037/0033-295X.97.3.332

Coles, M. G., Gratton, G., Bashore, T. R., Eriksen, C. W., and Donchin, E. (1985). A psychophysiological investigation of the continuous flow model of human

information processing. *J. Exp. Psychol. Hum. Percept. Perform.* 11, 529–553. doi: 10.1037/0096-1523.11.5.529

Diederich, A., Colonius, H., Bockhorst, D., and Tabeling, S. (2003). Visual-tactile spatial interaction in saccade generation. *Exp. Brain Res.* 148, 328–337. doi: 10.1007/s00221-002-1302-7

Dolk, T., Hommel, B., Colzato, L. S., Schütz-Bosbach, S., Prinz, W., and Liepelt, R. (2014). The joint Simon effect: a review and theoretical integration. *Front. Psychol.* 5:974. doi: 10.3389/fpsyg.2014.00974

Donohue, S. E., Appelbaum, L. G., Park, C. J., Roberts, K. C., and Woldorff, M. G. (2013). Cross-modal stimulus conflict: the behavioral effects of stimulus input timing in a visual-auditory Stroop task. *PLoS ONE* 8:e62802. doi: 10.1371/journal.pone.0062802

Ellinghaus, R., Karlbauer, M., Bausenhart, K. M., and Ulrich, R. (2018). On the time-course of automatic response activation in the Simon task. *Psychol. Res.* 734–743. doi: 10.1007/s00426-017-0860-z

Epstein, M., and Florentine, M. (2009). Binaural loudness summation for speech and tones presented via earphones and loudspeakers. *Ear Hear.* 30, 234–237. doi: 10.1097/AUD.0b013e3181976993

Eriksen, B. A., and Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Percept. Psychophys.* 16, 143–149. doi: 10.3758/BF03203267

Estes, W. K., and Maddox, W. T. (2005). Risks of drawing inferences about cognitive processes from model fits to individual versus average performance. *Psychon. Bull. Rev.* 12, 403–408. doi: 10.3758/BF03193784

Hommel, B. (1993). The relationship between stimulus processing and response selection in the Simon task: evidence for a temporal overlap. *Psychol. Res.* 55, 280–290. doi: 10.1007/BF00419688

Howard, I. P., and Templeton, W. B. (1966). *Human Spatial Orientation.* Oxford, England: John Wiley and Sons.

Hübner, R. (2014). Does attentional selectivity in global/local processing improve discretely or gradually? *Front. Psychol.* 5:61. doi: 10.3389/fpsyg.2014.00061

Jerger, S., Martin, R. C., and Pirozzolo, F. J. (1988). A developmental study of the auditory Stroop effect. *Brain Lang.* 35, 86–104. doi: 10.1016/0093-934X(88)90102-2

Kennett, S., Eimer, M., Spence, C., and Driver, J. (2001). Tactile-visual links in exogenous spatial attention under different postures: convergent evidence from psychophysics and ERPs. *J. Cogn. Neurosci.* 13, 462–478. doi: 10.1162/08989290152001899

Lagarias, J. C., Reeds, J. A., Wright, M. H., and Wright, P. E. (1998). Convergence properties of the Nelder–Mead simplex method in low dimensions. *SIAM J. Optim.* 9, 112–147. doi: 10.1137/S1052623496303470

Logan, G. D. (1980). Attention and automaticity in Stroop and priming tasks: theory and data. *Cogn. Psychol.* 12, 523–553. doi: 10.1016/0010-0285(80)90019-5

Lu, C.-H., and Proctor, R. W. (1995). The influence of irrelevant location information on performance: a review of the Simon and spatial Stroop effects. *Psychon. Bull. Rev.* 2, 174–207. doi: 10.3758/BF03210959

Luce, R. D. (1986). *Response Times: Their role in Inferring Elementary Mental Organization.* Oxford: University Press.

Macaluso, E., Frith, C., and Driver, J. (2002). Crossmodal spatial influences of touch on extrastriate visual areas take current gaze direction into account. *Neuron* 34, 647–658. doi: 10.1016/S0896-6273(02)00678-5

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychol. Bull.* 109, 163–203. doi: 10.1037/0033-2909.109.2.163

Mahani, M.-A. N. (2018). *Multimodal Extension of the Diffusion Model for Conflict Tasks [data and code, 2018, October 29].* doi: 10.17605/OSF.IO/EWSJD

Mahani, M.-A. N., Sheybani, S., Bausenhart, K. M., Ulrich, R., and Ahmadabadi, M. N. (2017). Multisensory perception of contradictory information in an environment of varying reliability: evidence for conscious perception and optimal causal inference. *Sci. Rep.* 7:3167. doi: 10.1038/s41598-017-03521-2

McClain, L. (1983). Stimulus-response compatibility affects auditory Stroop interference. *Percept. Psychophys.* 33, 266–270. doi: 10.3758/BF03202864

Medina, J., McCloskey, M., Coslett, H. B., and Rapp, B. (2014). Somatotopic representation of location: evidence from the Simon effect. *J. Exp. Psychol. Hum. Percept. Perform.* 40, 2131–2142. doi: 10.1037/a0037975

Metropolis, N., and Ulam, S. (1949). The Monte Carlo method. *J. Am. Stat. Assoc.* 44, 335–341. doi: 10.1080/01621459.1949.10483310

Miller, J. (1982). Divided attention: evidence for coactivation with redundant signals. *Cogn. Psychol.* 14, 247–279. doi: 10.1016/0010-0285(82)90010-X

Morey, R. D. (2008). Confidence intervals from normalized data: a correction to Cousineau (2005). *Tutor. Q. Methods Psychol.* 4, 61–64. doi: 10.20982/tqmp.04.2.p061

Osaka, N. (1976). Reaction time as a function of peripheral retinal locus around fovea: effect of stimulus size. *Percept. Mot. Skills* 42, 603–606. doi: 10.2466/pms.1976.43.2.603

Ossandón, J. P., König, P., and Heed, T. (2015). Irrelevant tactile stimulation biases visual exploration in external coordinates. *Sci. Rep.* 5:10664. doi: 10.1038/srep10664

Poole, D., Couth, S., Gowen, E., Warren, P. A., and Poliakoff, E. (2015). Adapting the crossmodal congruency task for measuring the limits of visual–tactile interactions within and between groups. *Multisens. Res.* 28, 227–244. doi: 10.1163/22134808-00002475

Rains, J. D. (1963). Signal luminance and position effects in human reaction time. *Vision Res.* 3, 239–251. doi: 10.1016/0042-6989(63)90057-9

Ratcliff, R. (1978). A theory of memory retrieval. *Psychol. Rev.* 85, 59–108. doi: 10.1037/0033-295X.85.2.59

Ratcliff, R., and Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychol. Rev.* 111, 333–367. doi: 10.1037/0033-295X.111.2.333

Ridderinkhof, K. R. (2002). "Activation and suppression in conflict tasks: empirical clarification through distributional analyses," in *Common Mechanisms in Perception and Action: Attention and Performance XIX*, eds W. Prinz and B. Hommel (New York, NY: Oxford University Press), 494–519.

Salzer, Y., Aisenberg, D., Oron-Gilad, T., and Henik, A. (2014). In touch with the Simon effect. *Exp. Psychol.* 61, 165–179. doi: 10.1027/1618-3169/a000236

Schupak, A., Caspi, A., and Chajut, E. (2015). Coping with conflicts improves under threat: evidence from a Simon and a visuo-auditory stroop tasks. *J. Vis.* 15:459. doi: 10.1167/15.12.459

Schwarz, W., and Miller, J. (2012). Response time models of delta plots with negative-going slopes. *Psychon. Bull. Rev.* 19, 555–574. doi: 10.3758/s13423-012-0254-6

Servant, M., White, C., Montagnini, A., and Burle, B. (2016). Linking theoretical decision-making mechanisms in the Simon task with electrophysiological data: a model-based neuroscience study in humans. *J. Cogn. Neurosci.* 28, 1501–1521. doi: 10.1162/jocn_a_00989

Simon, J. R., and Acosta, E. (1982). Effect of irrelevant information on the processing of relevant information: facilitation and/or interference? The influence of experimental design. *Percepti. Psychophys.* 31, 383–388. doi: 10.3758/BF03202663

Simon, J. R., and Craft, J. L. (1970). Effects of an irrelevant auditory stimulus on visual choice reaction time. *J. Exp. Psychol.* 86, 272–274. doi: 10.1037/h0029961

Simon, J. R., Hinrichs, J. V., and Craft, J. L. (1970). Auditory S-R compatibility: reaction time as a function of ear-hand correspondence and ear-response-location correspondence. *J. Exp. Psychol.* 86, 97–102. doi: 10.1037/h0029783

Simon, J. R., and Rudell, A. P. (1967). Auditory S-R compatibility: the effect of an irrelevant cue on information processing. *J. Appl. Psychol.* 51, 300–304. doi: 10.1037/h0020586

Simon, J. R., and Wolf, J. D. (1963). Choice reaction time as a function of angular stimulus-response correspondence and age. *Ergonomics* 6, 99–105. doi: 10.1080/00140136308930679

Slater-Hammel, A. (1955). Reaction time to light stimuli in the peripheral visual field. *Res. Q. Am. Assoc. Health Phys. Edu. Recreat.* 26, 82–87. doi: 10.1080/10671188.1955.10612805

Slutsky, D. A., and Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport* 12, 7–10. doi: 10.1097/00001756-200101220-00009

Spence, C., Pavani, F., and Driver, J. (2004). Spatial constraints on visual-tactile cross-modal distractor congruency effects. *Cogn. Affect. Behav. Neurosci.* 4, 148–169. doi: 10.3758/CABN.4.2.148

Stein, B. E., and Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Rev. Neurosci.* 9, 255–266. doi: 10.1038/nrn2331

Stone, M. (1960). Models for choice-reaction time. *Psychometrika* 25, 251–260. doi: 10.1007/BF02289729

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18, 643–662. doi: 10.1037/h0054651

Turner, B. M., and Van Zandt, T. (2018). Approximating Bayesian inference through model simulation. *Trends Cogn. Sci.* 22, 826–840. doi: 10.1016/j.tics.2018.06.003

Ulrich, R., Schröter, H., Leuthold, H., and Birngruber, T. (2015). Automatic and controlled stimulus processing in conflict tasks: superimposed diffusion processes and delta functions. *Cogn. Psychol.* 78, 148–174. doi: 10.1016/j.cogpsych.2015.02.005

Welch, R. B., and Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychol. Bull.* 88, 638–667. doi: 10.1037/0033-2909.88.3.638

Wesslein, A.-K., Spence, C., and Frings, C. (2014). Vision affects tactile target and distractor processing even when space is task-irrelevant. *Front. Psychol.* 5:84. doi: 10.3389/fpsyg.2014.00084

White, C. N., Servant, M., and Logan, G. D. (2018). Testing the validity of conflict drift-diffusion models for use in estimating cognitive processes: a parameter-recovery study. *Psychon. Bull. Rev.* 25, 286–301. doi: 10.3758/s13423-017-1271-2

Yue, Z., Bischof, G. N., Zhou, X., Spence, C., and Röder, B. (2009). Spatial attention affects the processing of tactile and visual stimuli presented at the tip of a tool: an event-related potential study. *Exp. Brain Res.* 193, 119–128. doi: 10.1007/s00221-008-1599-y

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Study 3

The official citation that should be used in referencing this material is:

Mahani, Mohammad-Ali Nikouei, Karin Maria Bausenhart, Rolf Ulrich, and Majid Nili Ahmadabadi. "Learning to integrate an artificial sensory device: from decision fusion to sensory integration." bioRxiv (2020), *Under review*.

https://doi.org/10.1101/2020.01.29.924662

The final publication is available here:

https://www.biorxiv.org/content/10.1101/2020.01.29.924662v1

Author contributions

All authors contributed in the main manuscript text. M-AM and KMB collected the data. M-AM, MNA, and RU contributed to the data analysis. M-AM, MNA, and RU has designed the study experiment and all authors reviewed the manuscript.

Learning to Integrate an Artificial Sensory Device: From Decision Fusion to Sensory Integration

Mohammad-Ali Nikouei Mahani[1,2], Karin Maria Bausenhart[2], Rolf Ulrich[2], Majid Nili Ahmadabadi[1]

1 Cognitive Systems Lab, School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran

2 Cognition and Perception, Department of Psychology, University of Tübingen, Tübingen, Germany

*Corresponding Author:

Email: nikouei@ut.ac.ir

Abstract

The present study examines how artificial tactile stimulation from a novel non-invasive sensory device is learned and integrated with information from another sensory system. Participants were trained to identify the direction of visual dot motion stimuli with a low, medium, and high signal-to-noise ratio. In bimodal trials, this visual direction information was paired with reliable symbolic tactile information. Over several blocks of training, discrimination performance in unimodal tactile test trials improved, indicating that participants were able to associate the visual and tactile information and thus learned the meaning of the symbolic tactile cues. Formal analysis of the results in bimodal trials showed that the information from both modalities was integrated according to two different integration policies. Initially, participants seemed to rely on a linear decision integration policy based on the metacognitive experience of confidence. In later learning phases, however, our results are consistent with a Bayesian integration policy, that is, optimal integration of sensory information. Thus, the present study demonstrates that humans are capable of learning and integrating an artificial sensory device delivering symbolic tactile information. This finding connects the field of multisensory integration research to the development of sensory substitution systems.

1. Introduction

Humans perceive the environment through multiple sensory inputs. Lacking or unreliable sensory inputs can cause a significant drop in the accuracy of perception. Some artificial sensory devices, such as substitution systems, can partially compensate the loss of a sensory modality [1,2]. Other artificial sensory devices are designed to improve perception by providing complementary or processed information[3,4]. Invasive artificial sensory systems exert a direct effect on the neuronal system[5] that can even lead to optimal integration of multisensory information[6]. However, invasive techniques can still gain from further development. Therefore, non-invasive sensory feedback devices such as wearable systems are potentially the best alternatives for invasive techniques because they are already developed for realistic applications. Although there is a large number of studies that addressed the technical aspects of non-invasive and wearable devices [7-9], the cognitive aspects of them are less well studied. In particular, it is still unclear whether and how the input from a non-invasive artificial sensory device can be integrated into the multisensory perceptual system.

Many studies reported that adult humans integrate multiple sensory modalities in an optimal Bayesian fashion. Bayesian integration, in comparison to unimodal perception, leads to a significant decrease in response time [10-12] and an increase in accuracy and reliability of perception [12-16]. However, most of the previous studies have only examined the integration of well-experienced sensory inputs. There are only a few studies that addressed whether the learning of novel sensory devices leads also to an optimal integration or the selection of a sensory modality. Dadarlat et al. showed that monkeys could optimally integrate unfamiliar multichannel intracortical microstimulation (ICMS) signals and proprioceptive input[6]. Nevertheless, this issue has not been addressed so far for non-invasive artificial devices.

In the present study, we thus investigated the integration of visual motion information with symbolic input from an unfamiliar wearable vibrotactile device. In several consecutive training phases, participants received synchronous static vibrotactile spatial patterns and visual random dot motion stimuli. They were asked to learn these visual-tactile associations. Following each training phase, the

trained stimuli were presented either unimodally or bimodally and participants were asked to report the associated direction of motion. Accuracy and self-reported confidence were assessed to examine whether or not participants can learn the symbolic meaning conveyed by the artificial wearable device and whether the information form the two inputs could be integrated.

To this end, participants performed a multisensory learning task which involves a novel artificial vibro-tactile device (see Figure 1**Error! Reference source not found.**). The experiment consisted of seven blocks, each involving a training phase followed by an evaluation phase. During the training phase, participants simultaneously received a dot motion and a novel vibro-tactile stimulation pattern, which did not involve any directional movement. Throughout the whole experiment, each motion direction was paired with a specific symbolic vibro-tactile pattern. Participants were asked to learn the associations between each specific motion direction and the corresponding vibro-tactile stimulation pattern. In the evaluation phase, participants received a unimodal visual motion dot stimulus, a unimodal vibro-tactile stimulus, or a multimodal stimulus combining both inputs (see materials and methods for more details). Participants were then asked to report motion direction, either directly from the motion of the dots, from the associated vibro-tactile pattern, or from both. They also rated the confidence of their decisions. The visual stimulus in both unimodal conditions and the multimodal visual-tactile condition had three levels of reliability: low, medium and high. The reliability of tactile stimuli was constant during the whole experiment.

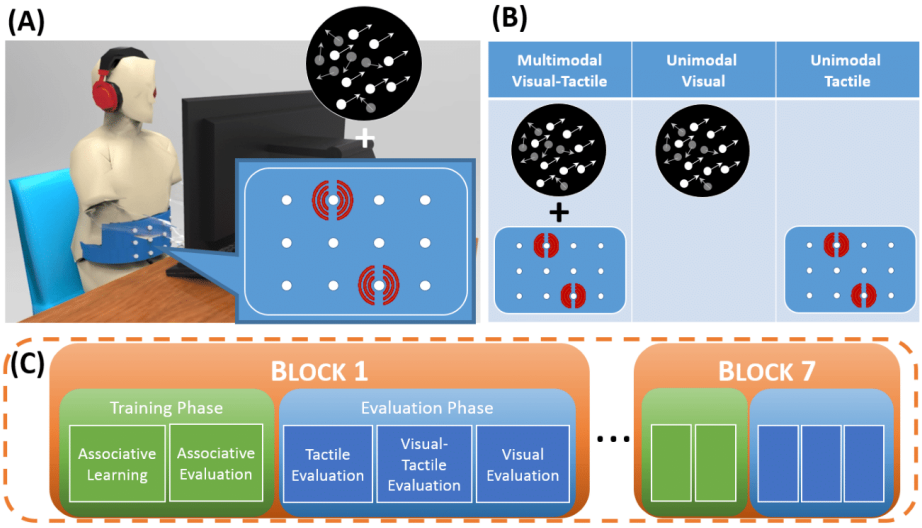

Figure 1. (A) Experimental setup. Participants received synchronous static vibrotactile spatial patterns and visual random dot motion stimuli. Vibroractile patterns were provided by a custom designed belt through a matrix of 3x4 tiny vibration motors. (B) Visual stimulus, tactile stimulus and synchronous visual-tactile stimulus. The reliability of visual dot motion stimulus was manipulated through three coherence levels, while the reliability of the vibrotactile stimulus was fixed. (C) The experiment consisted of seven blocks, each involving a training phase followed by an evaluation phase.

The impact of the artificial sensory stimulation on multisensory perception was assessed over the course of learning. We analyzed the accuracy of perception as well as the confidence in perceptual decisions in low, medium, and high reliability conditions.

2 Results

Figure 2 depicts the accuracy of perception in low, medium, and high reliability conditions. The accuracy of tactile perception increases over the course of the experiment, thus demonstrating that the visuo-tactile patterns provided by our artificial sensory device were efficiently learned and associated with the corresponding dot-motion patterns.

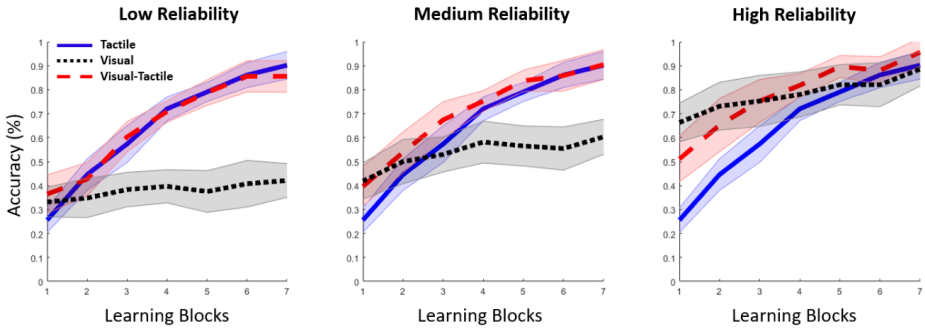

Figure 2. Accuracy of perception in low, medium, and high reliability conditions over the course of the learning procedure. For the sake of clarity, data from the tactile condition are re-plotted in each panel.

Separate two-way within-subject ANOVA with factors reliability condition (low, medium, and high) and block (block 1 to block 7) were performed on accuracy of visual and visual-tactile perceptions.

The accuracy of visual perception was significantly affected by reliability condition, $F(2,44) = 457.22$, $p < .001$, as well as by block, $F(6,132) = 9.48$, $p < .001$. The interaction of reliability condition and block was not significant, $F(12,264) = 1.47$, $p = .13$. Post-hoc Tukey tests on reliability condition showed significant differences between all reliability conditions ($ps < .001$). This result points to a successful reliability control of the visual motion stimuli.

Analysis of the visual-tactile perception also showed a significant effect of reliability condition on accuracy, $F(2,44) = 30.73$, $p < .001$. The effect of block on accuracy was also significant, $F(6,132) = 44.93$, $p < .001$. Post-hoc Tukey tests on reliability condition showed the same result as before, that is, accuracy increased from low to medium, and from medium to high reliability (all $ps < .001$). This effect of reliability condition was especially pronounced in the initial blocks of learning, as indicated by an interaction of both factors, $F(12,264) = 2.81$, $p = .001$. This shows that the information from both modalities is integrated differently depending on the reliability of the visual input.

Additional two-way within-subject ANOVAs with factors modality (visual, tactile, visual-tactile) and block (1 to 7) were conducted for each of the three reliability

conditions to assess learning and multisensory integration of the artificial sensory device.

In the low reliability condition, the effect of modality, $F(2,44) = 83.40$, $p < .001$, the effect of block, $F(6,132) = 52.76$, $p < .001$, and their interaction, $F(12,264) = 23.04$, $p < .001$, were significant. Post-hoc Tukey tests revealed that the visual stimuli were perceived less accurately than tactile ($p < .001$) and visual-tactile stimuli ($p < .001$). However, the difference between the accuracy of visual-tactile and tactile was not significant ($p = .81$). This result indicates a dominance of tactile stimulation on visual-tactile perception, when the reliability of the visual stimuli is low.

The same analysis in the medium reliability condition also showed effects of modality, $F(2,44) = 24.93$, $p < .001$, block, $F(6,132) = 52.42$, $p < .001$, and their interaction, $F(12,264) = 16.08$, $p < .001$. Although post-hoc Tukey tests indicated significant differences among all modalities (all ps $< .007$), the slope of accuracy across blocks was different between modalities. At the beginning of learning, the accuracy for the tactile stimuli was lower than for the visual stimuli, and the accuracy for visual-tactile stimulation was determined by the visual input. However, tactile perception got more accurate over the course of learning and as a result, its influence on the multimodal visual-tactile perception increased. In some learning blocks, the accuracy for visual-tactile stimulation was even higher than for the unimodal visual and tactile stimuli in isolation (see Figure 2). This indicates an effective integration of visual and tactile information (see the modeling section for a more elaborated account of this integration effect).

Finally, in the high reliability condition, accuracy was also affected by modality, $F(2,44) = 13.59$, $p < .001$, block, $F(6,132) = 47.80$, $p < .001$, and their interaction, $F(12,264) = 14.67$, $p < .001$. At the beginning of the experiment, the accuracy for visual-tactile stimulation was codetermined by both visual and tactile information, as indicated by intermediate accuracy, lying between the respective accuracies for visual and tactile unimodal stimulation. This was confirmed by a one-way within-subject ANOVA, which showed a significant difference of accuracy between the different modalities in the first block, $F(2,44) = 55.21$, $p < .001$. Post-hoc tests showed significant differences among all modalities (ps $< .001$). Most interestingly, this suggests that participants integrated the visual and tactile

information from the beginning of learning, even though this impaired performance in comparison to the unimodal visual condition.

3 Modeling

The potential mechanisms underlying the integration of the two modalities were investigated by computational modeling in the subsequent section. In order to shed light on the mechanisms behind the perceptual learning of this artificial sensory device, two computational models are proposed and fitted to the experimental data. The first model focuses on decision integration and the second one on multisensory integration.

3.1 Decision Integration Model

Confidence Confusion Matrix and Meta-Accuracy

The confusion matrix is a useful tool for evaluating the performance of participants in a multiple-choice task or the performance of a classifier algorithm in solving supervised multi-class problems. Various criteria like accuracy, precision, and recall can be determined from the confusion matrix for both 2AFC and multiple choice (multiple class) problems.

For the present purpose, the confusion matrix is extended to second order judgments, that is, self-reported confidence in judgments which is also associated with a given response. When a participant (a classifier) reports a confidence for each of the first-order judgments, this meta information can be used to compute meta-accuracy (type II accuracy). A confidence confusion matrix (CCM) is defined with the same structure as the confusion matrix as follows:

$$CCM = \begin{bmatrix} C_{1,1} & \cdots & C_{1,N} \\ \vdots & \ddots & \vdots \\ C_{N,1} & \cdots & C_{N,N} \end{bmatrix} \quad (1)$$

$C_{i,j}$ indicates the sum of all confidence ratings associated with response $R_{i,j}$, where a ground truth signal i (class i) has been predicted as signal j (class j). Similar to the accuracy, we define meta-accuracy as the sum of the diagonal divided by the sum of all elements:

$$MetaAccuracy = \frac{\sum_{i=j=1}^{N} C_{i,j}}{\sum_{i=1}^{N}\sum_{j=1}^{N} C_{i,j}} \qquad (2)$$

Meta-accuracy represents the performance of the participants not only by considering the decisions but also by weighting the decisions according to the confidence ratings associated with these decisions. High-confidence correct answers thus increase meta-accuracy more than low-confidence correct answers, while the low-confidence wrong answers decrease the meta-accuracy less than the high confidence wrong answers. In fact, CCM is a more general representation of performance than confusion matrix, which is a special case with just one confidence level.

Decision policy

We modeled the multi-modal integration of perceptual decisions considering an ideal observer. This observer has only access to the decisional information, i.e. first- and second-order decisions. In the present case, perceptual decisions are based on two modalities A and B, thus yielding two separate unimodal confidence confusion matrices, CCM^A and CCM^B. CCM can be interpreted as a decision-value table, for which the value of a decision is an internal estimation based on the confidence values. By this interpretation, decision values are considered to be correlated with the confidences, or, in other words, confidences represent the internal estimation of decision values. For example, given an observed signal i, the i^{th} column of the decision value table represents how valuable (and how likely) is each perceptual decision j. One rational way of taking decisions based on two modalities is to decide based on the Max_CCM table, where the highest confidence value is selected for each cell (*i,j*) from the two confidence confusion matrices,

$$Max\_CCM_{i,j} = \max\left(CCM^{A}_{i,j}, CCM^{B}_{i,j}\right). \quad (3)$$

Given the values of all decisions, the probability of selecting a decision when a specific signal is observed can be estimated by various decision making policies. We have investigated two decision making policies in the present study: linear decision making and parametrized softmax decision making.

The linear decision-making policy is defined as follows:

$$p(d|s_i) = \frac{V(s_i,d)}{\sum_{d' \epsilon D} V(s_i,d')} \quad (4)$$

Where $s_i^{'}$ denotes the state in which signal *i* is observed, *D* is the set of all possible decisions, and $V(s_i,d)$ shows the value of decision *d* when the signal *i* is observed. The Max_CCM table is used as the estimation of the decision value table (V). When an agent decides based on the linear policy, the perceptual accuracy of the agent is actually equal to the meta-accuracy of the Max_CCM.

The top row of Figure 3 depicts the result of the ideal linear decision maker. The model successfully predicts the integration behavior for the first blocks. However, it systematically underestimates the accuracy in the last blocks. We have investigated further the accounts for this systematic underestimation by analyzing the calibration of observed confidence ratings. Figure 4 reveals the confidence calibration for all reliability conditions and across all blocks. The orange line indicates the ideal calibration pattern. In an ideal situation, the confidence should represent the actual probability of making correct decisions. As an example, assume that a participant would report the confidence level of 2 (50% probability) for ten trials. If five decisions out of these ten are correct, the confidence would be well calibrated. Lower or higher numbers of correct decisions would show overestimation or underestimation of the confidence.

Figure 4 reveals that the confidence calibration was almost fixed across all blocks in the unimodal visual reliability conditions. This is probably due to none or very little perceptual learning in these conditions. However, confidence calibration varied across blocks in the tactile and all visual-tactile conditions. Specifically, participants tended to report lower confidence for correct decisions in the last three blocks compared to the first three to four blocks. This indicates an underestimation of confidence reports in the last blocks, which also explains the systematic underestimation of the predicted accuracy by the linear decision-making policy. Therefore, we proposed the parametrized softmax model to counterbalance the systematic underestimation error introduced by the linear model. The parametrized softmax model is mathematically equivalent to the temperature scaling [17], which is a well-known approach in deep neural networks to recalibrate the probability estimation.

Moreover, many studies in neuroscience and psychology utilized the softmax policy to model decision making behavior [18-21]. In a softmax approach, usually a decision with a higher value is selected with an exponentially higher probability. The general form of the softmax policy with a linear objective function [18] can be defined as follows:

$$p(d|s_i) = \frac{\exp(\theta^T V(s_i,d))}{\sum_{d' \epsilon D} \exp(\theta^T V(s_i,d'))} \quad (5)$$

In the present study, θ is described by a linear equation with respect to the block number $\theta = wb + \alpha$, where b is the block number and $\{w,\alpha\}$ are free scalar parameters. We considered separate $\{w_r,\alpha_r\}$ for each reliability condition, where r denotes the reliability condition and $r\epsilon\{low, medium, high\}$. The total number of six free parameters has been fitted to the experimental data for all blocks and all three reliability conditions.

Figure 3 depicts results of both linear and softmax ideal observer decision makers. The performance predicted by the linear ideal observer model is close to the observed performance for the visual-tactile stimulation in the first blocks and this is true for all reliability conditions. However, the softmax model depicts a better performance on average in all blocks. It should be noticed that these models do not imply absolute dominance or a weighted combination of the modalities' inputs. Rather the simple decision integration mechanisms outlined above are sufficient to predict the quite complex result pattern. For example, both models are capable to predict the counterintuitive result that multimodal accuracy is partly higher than accuracies for the unimodal conditions (medium reliability condition, blocks 2 and 3). Both models can also predict multimodal accuracy in the high reliability condition, which is between the accuracies of the two unimodal conditions.

Nevertheless, the modeling results also show that the linear ideal observer cannot successfully predict the performance in all cases. Specifically, in the first half of the experiment (blocks 1 to 3/4), the model predicts the performance of the visual-tactile perception very accurately in all reliability conditions. However, the prediction is worse for the second half of the experiment, where this decision integration model systematically underestimates the observed multimodal accuracy. The discrepancy in the linear model can be attributed to a gradual

change in the mechanism that integrates information from the different sensory modalities as the participants become more familiar with the tactile stimulation, from a pure decisional process towards a genuine perceptual integration process. In contrast to the linear model, the softmax model benefits from a linear increase in model parameters across the learning blocks and therefore can capture the gradual change of perceptual mechanism. We will however address the possibility of a genuine perceptual integration process with a Bayesian multisensory integration model in the next section.

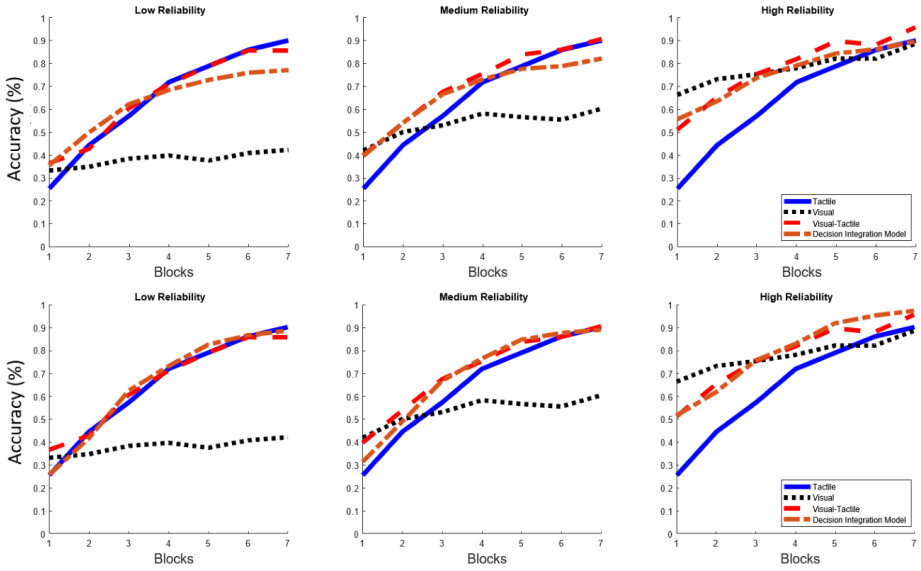

Figure 3. Linear and softmax decision integration models in low, medium and high reliability conditions. The upper row shows the linear model and the lower row depicts the softmax model. Red lines show the experimental visual-tactile integration results while orange lines indicate the modeling result. The model was fitted to the aggregated data from all subjects.

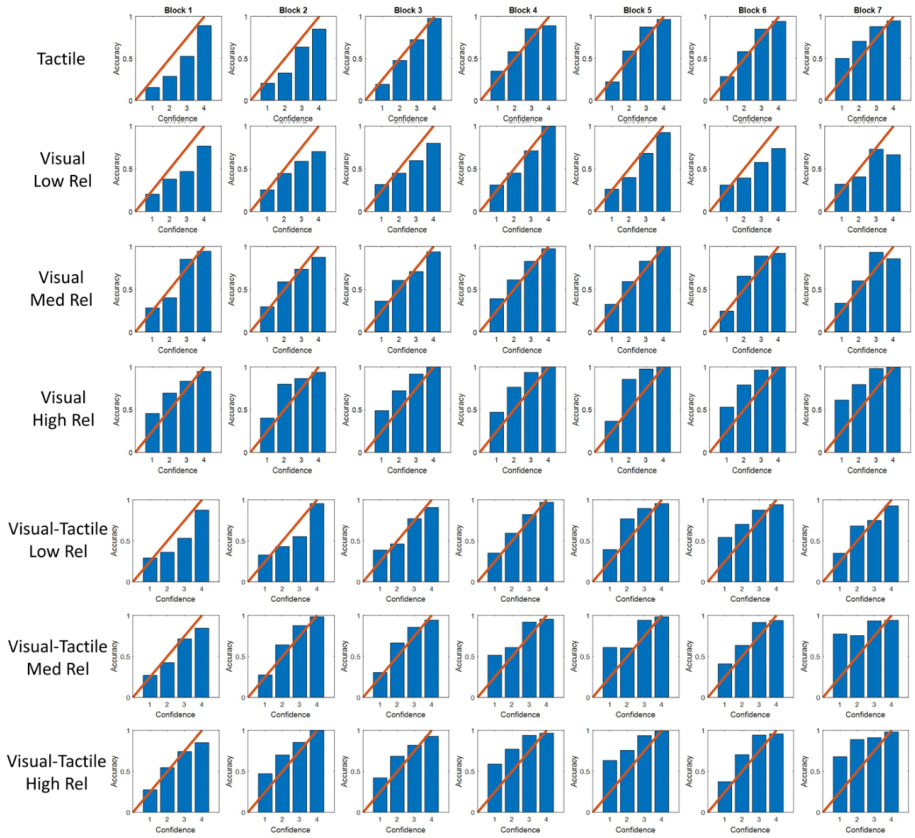

Figure 4. Confidence calibration plots of aggregated data from all participants. The orange line indicates the ideal calibration pattern. The results reveal that unimodal visual confidence calibration changed very little across blocks. However, the change of calibration is obvious in tactile and visual-tactile conditions. Specifically, the participants tended to report lower confidence for correct answers in the last blocks in comparison with the first blocks.

3.2 Multisensory Integration Model

During the previous decade, multisensory integration has become an increasingly prominent research topic and various computational models have been proposed to unravel the mechanism underlying sensory integration across multiple

modalities [10-12,16]. The well-known model of multisensory integration suggests that the sensory inputs are combined according to the principle of maximum-likelihood estimation (MLE)[13]. Even in case of small deviations the core principle of MLE is valid [22] for multisensory integration. Specifically, this model holds that the inputs from different modalities are linearly combined according to the reliability of their respective sensory inputs. This simply means that if there are two modalities, A and B, the weight of each modality in integration, w_A, and $w_B = 1 - w_A$, and the optimal integrated signal S_{AB} are calculated as follows:

$$w_A = \frac{r_A}{r_A + r_B} \qquad (6)$$

$$S_{AB} = w_A S_A + w_B S_B \qquad (7)$$

Where r_A and r_B show the reliability of signal A and reliability of signal B, respectively. We investigated how well the Bayesian model can predict the visual-tactile perception given the unimodal visual and tactile information. The angular distance between the ground truth direction (i.e. dot movement direction) and the chosen (predicted) direction by participants is assumed to represent the reliability of the perception. If in a given trial, the dot display moves in direction α while a participant selects the direction β, then the reliability of perception can be calculated as follows:

$$d_\theta(\alpha,\beta) = cos^{-1}(sin(\alpha) \cdot sin(\beta) + cos(\alpha) \cdot cos(\beta)) \qquad (8)$$

$$r_{\alpha\beta} = \pi - d_\theta(\alpha,\beta) \qquad (9)$$

$d_\theta(\alpha,\beta)$ is the angular distance between α and β, which has a maximum value of π. Therefore, $r_{\alpha\beta}$ is an estimation of the perceptual reliability that is negatively correlated with the angular distance. A higher angular distance indicates less reliability in the perception.

The confusion matrix of the Bayesian integration model, CM_{ij}^{AB}, is estimated as the weighted sum of unimodal confusion matrices, where the weights are indicated by unimodal perceptual reliabilities:

$$CM_{ij}^{AB} = r_{ij}^{A} CM_{i,j}^{A} + r_{ij}^{B} CM_{i,j}^{B} \qquad (10)$$

r_{ij}^{A} and r_{ij}^{B} are unimodal perceptual reliabilities obtained from Eq (9), CM_{ij}^{A} and CM_{ij}^{B} are unimodal confusion matrices.

Figure 5 depicts the predicted accuracy by the Bayesian integration model. As can be seen in medium and high reliability conditions, the observed accuracy for the visual-tactile condition resembles the predicted accuracy by the Bayesian model especially in the last 3 to 4 blocks. In the low reliability condition, however, the observed accuracy for the visual-tactile condition closely follows the observed unimodal tactile accuracy, suggesting selection behavior rather than integration. Thus, the Bayesian integration model failed to predict the selection behavior in the low reliability condition.

Even though the Bayesian integration model predicted the accuracy of medium and high reliability in the last blocks, it could not fit the experimental data in the first blocks, where the linear decision integration model provided a reasonable data fit. In other words, the linear decision integration model and the multisensory integration model supplement each other.

Table 1 compares BIC statistics of all three models that we mentioned in this paper. These additional results demonstrate that none of three models alone provide a satisfactory description of the experimental data. This additional analysis, however, confirms the conclusion that the linear decision model accounts for the observed accuracy for the first three blocks, whereas the Bayesian integration model accounts for the last three blocks after the initial learning phase.

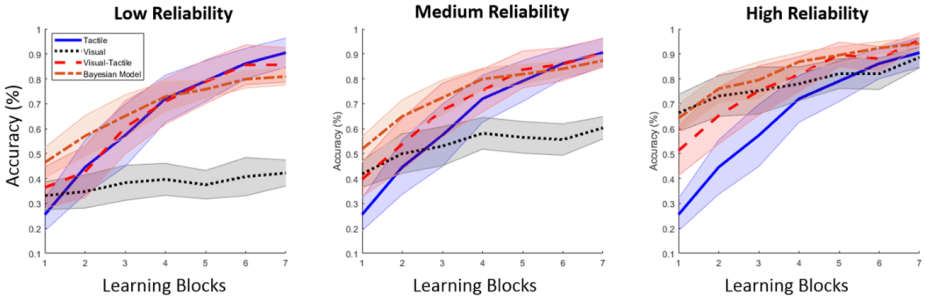

Figure 5. Predicted accuracy by the Bayesian sensory integration model in the low, medium and high reliability conditions. Accuracy is calculated according to Eq (10). Red lines depict the empirical results of the visual-tactile condition, and orange lines show the model fits of the Bayesian integration model. This model was fitted to each individual subject and the confidence intervals were computed by the method explained in [23].

Table 1. Model Comparison

Model	χ^2 Low reliability	χ^2 Medium reliability	χ^2 High Reliability	χ^2 First Three Blocks	χ^2 Last Three Blocks	χ^2 Total	BIC First Three Blocks	BIC Last Three Blocks	BIC Total
Linear Decision Fusion	3.86	1.86	1.29	1.74	5.26	7.00	***1.74***	5.26	7.00
Softmax Decision Fusion	0.86	2.39	3.66	5.82	1.08	6.90	11.54	6.80	14.83
Sensory Integration	8.80	6.86	5.87	19.70	1.83	21.54	19.70	***1.83***	21.54

4. Discussion

With the advance of medical and wearable devices, understanding the mechanisms of perception for novel artificial devices is of great interest to scientists and engineers. This study addresses whether and how symbolic information from a novel non-invasive artificial sensory device is learned and becomes integrated within the human sensory system. We have introduced a custom designed vibrotactile belt that generates novel and unexperienced tactile patterns. Participants were asked to learn the associations between these novel

tactile stimuli and the direction of visual dot motion stimuli across seven consecutive training phases. At the end of each training phase, there was an evaluation phase in which the performance in unimodal visual, unimodal tactile, and visual-tactile conditions was assessed. Thus, perceptual accuracy for the different modality conditions could be tracked over the time-course of the whole experiment.

The results show that accuracy in the tactile condition increased throughout the experiment, and thus, participants could learn the novel tactile patterns. Results also revealed that symbolic information from the novel artificial device can be integrated into the visual-tactile perception from the very beginning of the training. This integration process was evident even though accuracy for the tactile modality was initially lower than for the visual modality. This phenomenon was evident in the medium and high visual reliability conditions. However, in the low visual reliability condition, the tactile information dominated over the visual information, resulting in selection behavior rather than sensory integration.

To better understand the mechanisms underlying the perception and integration of novel symbolic information, two computational models were proposed and evaluated with the data of this study. The decision integration model predicts accuracy in the visual-tactile condition based on first- and second-order perceptual decisions in the two modalities. The multisensory integration model assumes a genuine perceptual information integration, which combines the sensory inputs based on their respective sensory variances on a pre-decisional processing level. This model thus assumes that decision processes operate on the combined sensory multisensory input.

The decision integration model treats confidences as estimates of decision values. Higher confidences are associated with greater values. Assuming that these values are known, the modeling problem is reduced to finding a proper decision-making policy that fits the experimental data. Two decision making policies were evaluated: A linear policy and a parametrized softmax policy. The linear policy explained well the integration behavior in the first half of the experiment. However, it did not fit well the data in the second half of the experiment. Specifically, it underestimated accuracy in the visual-tactile condition. To overcome this problem, we exploited the parametrized softmax policy with a

linear equation parameter (see the modeling section for more details). Since the policy parameters were linearly updated w.r.t. training blocks, the model now fitted reasonably the experimental data in all training blocks. In general, the decision-making model revealed that confidences in decisions properly represent the decision values. This finding together with the notion of confidence confusion matrix provides a new perspective for future research on decision-making models employing self-reported confidence ratings. Moreover, the comparison between the linear policy and the parametrized softmax policy provides evidence for a gradual change in the integration mechanism over the course of training.

We have hypothesized that this gradual change might reflect a switch from a pure decisional integration towards perceptual integration. We addressed this hypothesis by fitting a Bayesian integration model to the experimental data. In contrast to the linear decision-making policy, the Bayesian model fitted well the experimental data only in the second half of the experiment. The complementary fits of the linear policy decision integration model and the Bayesian sensory integration model are consistent with the proposed hypothesis.

In conclusion, the present data show that participants can utilize symbolic tactile information to improve the processing of visual information, thereby improving the accuracy of decision making. We have suggested two models of how this multimodal integration might proceed. In the initial learning phase, participants seem to rely on a linear decision integration process, whereas in a later phase the integration process operates in a Bayesian fashion.

5. Methods

Participants

18 women and 11 men (23.72 ± 3.42 years old) recruited from student population of Tübingen University participated in the experiment. They all reported normal or corrected-to-normal vision and no neurological or psychiatric disorder. The experiment was conducted in accordance with the Helsinki Declaration and the guidelines for scientific work of the University of Tübingen. Written informed consent was obtained from all participants prior to data collection. Participants were compensated with 8 Euros per hour or course credit for their participation.

They were also told that the best performer would get a 50 Euro Amazon gift card. The results of six participants were excluded from all analyses, since their performance in tactile perception was always below 50% in all blocks.

Stimuli and Apparatus

Participants were seated in a sound-attenuated room in front of a 19" screen with 100 Hz refresh rate and 1024x768 pixels resolution, on which the visual stimuli were presented. The distance between participants' eyes and the monitor screen was about 50 cm. The experiment was implemented in C++.

The visual stimuli consisted of a random dot motion display, made up of 100 white dots against a black background. The diameter of each dot was 10 pixels, approximately 0.33° visual angle. The dots were randomly distributed within a circle with a diameter of 400 pixels (approximately 13.3° visual angle) called excircle. The initial x/y position of each dot came from two Gaussian distributions with mean zero (center of the screen) and variances of 0.1 of the width/height of the screen. Each dot started to move with the speed of 11.25°/s in a specific direction. Whenever a dot hit the excircle, it was replaced by a new dot. There were eight possible directions (0, 45, 90, 135, 180, 225, 270, and 315°). The percentage of the dots moving in the same direction, called coherence level, determined the reliability of the visual stimulus. For example, a coherence level of 75% means 75 out of 100 dots moved in the same direction, while 25 dots moved in a random direction. The duration of the visual stimulus was one second.

Tactile stimuli were provided by a custom designed belt fastened around the subjects' waist. The vibrotactile belt was designed in the Cognitive Robotics Lab at University of Tehran. It consists of 12 vibration motors located in a 3x4 matrix formation on a cotton canvas tape. All motors are controlled by a custom embedded device similar to the one employed in a previous study. [16] Eight different tactile stimulation patterns were produced, each consisting of two simultaneous vibrations at pre-defined locations on the belt. These patterns were defined to be maximally distinguishable. The vibration of all tactile stimuli lasted one second, that is, the same duration as the visual stimuli. White-noise auditory stimuli were continuously delivered to the subjects through a pair of Sony MDR-XD200 stereo headphones to override the sound produced by the vibration motors.

Procedure

The experiment consisted of seven consecutive blocks, each of which was divided into a *training phase* and an *evaluation phase (*see Figure 1**Error! Reference source not found.**).

Training phase. Each block started with a training phase in which participants were asked to learn the associations between the tactile patterns and the visual motion directions. A random, subject-specific, and one-by-one mapping associated each of the tactile patterns with one of the visual motion directions. The mappings were randomly assigned prior to the experiment for each participant. The training phase included an associative learning section followed by an associative test section. In the learning section, all visual-tactile association pairs were presented four times in a random order. In the associative test section, participants were asked to decide if the presented pair of visual-tactile stimuli were associated or not, by selecting a green/red circle with a mouse click. Subsequently, they reported their confidence in these decisions on a scale of 1 (lowest confidence level) to 4 (highest confidence level) by selecting one of four adjacent circles by mouse. The cursor position was set to the screen center prior to each judgement to avoid response bias. Participants received feedback about the correctness of their decision at the end of each trial. A splash screen, which showed the overall accuracy achieved during the test section, concluded each training phase. The coherence of the visual stimuli during the whole training phase was 75 %.

Evaluation phase. Following each training phase, performance for visual, tactile, and visual-tactile judgments was assessed. Each evaluation phase included a visual evaluation section, a tactile evaluation section, and a visual-tactile evaluation section, which were presented in random order. In each trial of the visual evaluation section, participants received a unimodal visual stimulus with a coherence level of 10%, 15%, or 25% (low, medium, and high reliability, respectively). Participants were asked to judge the direction of the dot motion by a mouse click on the corresponding sector in a circle. Then, they were asked to report their confidence level on a scale of 1 to 4, similar to the test section of the training blocks. The visual evaluation section consisted of 48 trials (8 directions x 3 coherence levels x 2 repetitions), presented in a random order. The visual-tactile evaluation section was identical to the unimodal visual evaluation section, with

the exception that the visual stimulus in each trial was accompanied by its associated tactile pattern. The tactile evaluation section consisted of 16 trials (8 patterns x 2 repetitions). Participants were asked to report the learned associated motion direction and their confidence in this decision.

References

1 Abboud, S., Hanassy, S., Levy-Tzedek, S., Maidenbaum, S. & Amedi, A. EyeMusic: Introducing a "visual" colorful experience for the blind using auditory sensory substitution. *Restorative neurology and neuroscience* **32**, 247-257 (2014).

2 Maidenbaum, S., Abboud, S. & Amedi, A. Sensory substitution: closing the gap between basic research and widespread practical visual rehabilitation. *Neuroscience & Biobehavioral Reviews* 41, 3-15 (2014).

3 Shull, P. B. & Damian, D. D. Haptic wearables as sensory replacement, sensory augmentation and trainer–a review. *Journal of neuroengineering and rehabilitation* 12, 59 (2015).

4 Spiers, A. J. & Dollar, A. M. in *Haptics Symposium (HAPTICS), 2016 IEEE.* 34-40 (IEEE).

5 Collins, K. L. *et al.* Ownership of an artificial limb induced by electrical brain stimulation. *Proceedings of the National Academy of Sciences* 114, 166-171 (2017).

6 Dadarlat, M. C., O'doherty, J. E. & Sabes, P. N. A learning-based approach to artificial sensory feedback leads to optimal integration. *Nature neuroscience* 18, 138 (2015).

7 Iqbal, M. H., Aydin, A., Brunckhorst, O., Dasgupta, P. & Ahmed, K. A review of wearable technology in medicine. *Journal of the Royal Society of Medicine* 109, 372-380 (2016).

8 Mukhopadhyay, S. C. Wearable sensors for human activity monitoring: A review. *IEEE sensors journal* 15, 1321-1330 (2015).

9 Son, D. *et al.* Multifunctional wearable devices for diagnosis and therapy of movement disorders. *Nature nanotechnology* 9, 397 (2014).

10 Diederich, A. & Colonius, H. Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Perception & psychophysics* 66, 1388-1404 (2004).

11 Drugowitsch, J., DeAngelis, G. C., Klier, E. M., Angelaki, D. E. & Pouget, A. Optimal multisensory decision-making in a reaction-time task. *Elife* 3 (2014).

12 Drugowitsch, J., DeAngelis, G. C., Angelaki, D. E. & Pouget, A. Tuning the speed-accuracy trade-off to maximize reward rate in multisensory decision-making. *Elife* 4 (2015).

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

14 Butler, J. S., Smith, S. T., Campos, J. L. & Bülthoff, H. H. Bayesian integration of visual and vestibular signals for heading. *Journal of vision* 10, 23-23 (2010).

15 Pouget, A., Beck, J. M., Ma, W. J. & Latham, P. E. Probabilistic brains: knowns and unknowns. *Nature neuroscience* 16, 1170 (2013).

16 Mahani, M.-A. N., Sheybani, S., Bausenhart, K. M., Ulrich, R. & Ahmadabadi, M. N. Multisensory Perception of Contradictory Information in an Environment of Varying Reliability: Evidence for Conscious Perception and Optimal Causal Inference. *Scientific Reports* 7 (2017).

17 Guo, C., Pleiss, G., Sun, Y. & Weinberger, K. Q. in *Proceedings of the 34th International Conference on Machine Learning-Volume 70.* 1321-1330 (JMLR. org).

18 Reverdy, P. & Leonard, N. E. Parameter estimation in softmax decision-making models with linear objective functions. *IEEE Transactions on automation science and engineering* 13, 54-67 (2016).

19 Cohen, J. D., McClure, S. M. & Yu, A. J. Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences* 362, 933-942 (2007).

20 Cooper, J. A. *et al.* Training attention improves decision making in individuals with elevated self-reported depressive symptoms. *Cognitive, Affective, & Behavioral Neuroscience* 14, 729-741 (2014).

21 Daw, N. D., O'doherty, J. P., Dayan, P., Seymour, B. & Dolan, R. J. Cortical substrates for exploratory decisions in humans. *Nature* 441, 876 (2006).

22 Meijer, D., Veselič, S., Calafiore, C. & Noppeney, U. J. C. Integration of audiovisual spatial signals is not consistent with maximum likelihood estimation. Cortex 119, 74-88 (2019).

23 Morey, R. D. Confidence intervals from normalized data: A correction to Cousineau (2005). *reason* 4, 61-64 (2008).