

Object perception by ear and eye: fMRI studies on multisensory processes in human cerebral cortex

Citation for published version (APA):

Naumer, M. J. (2006). *Object perception by ear and eye: fMRI studies on multisensory processes in human cerebral cortex*. [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20061219mn>

Document status and date:

Published: 01/01/2006

DOI:

[10.26481/dis.20061219mn](https://doi.org/10.26481/dis.20061219mn)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 23 Jul. 2024

Summary

Invasive electrophysiological studies in animals and non-invasive investigations in human subjects provide impressive evidence that even single cognitive functions involve widely distributed regions of the brain. This is true not only for crossmodal and sensory-motor integration that, per definition, require coordination between diverse sensory and executive structures. It also holds for unimodal perception because sensory systems also consist of multiple subsystems that operate in parallel and are specialized for the analysis of particular features of objects. A core question of cognitive neuroscience is, how these multiple, widely distributed processes are coordinated and how the respective computational results are bound together in order to generate coherent unisensory and multisensory percepts. The functional magnetic resonance imaging (fMRI) studies reported in this dissertation had the common goal to contribute to a better understanding of the mechanisms that underlie crossmodal audio-visual integration. Particular emphasis has been laid on the investigation of two dimensions that are of prime importance for audio-visual integration: spatial (**chapter 2**) and semantic congruency (**chapters 3 and 4**).

Chapter 1 provides an introduction to both unisensory (auditory, visual, and tactile) as well as multisensory (audio-visual and visuo-tactile) object processing. We reviewed human fMRI studies that showed that visual, tactile, and auditory information about objects can activate cortical association areas that were once believed to be modality-specific. Processing converges either in multisensory zones or via direct crossmodal interaction of modality-specific cortices without relay through multisensory regions. We integrate these findings with existing theories about semantic processing and propose a general mechanism for crossmodal object recognition: The recruitment and location of multisensory convergence zones varies depending on the information content and the modality that provides the most reliable sensory input.

For the successful integration of visual and auditory information it is important to identify whether visual and auditory signals originate from corresponding places in the external world. In **chapter 2**, we report crossmodal effects of spatially congruent and incongruent audio-visual stimulation. Visual and auditory stimuli were matched to one of four horizontal locations in external space. Subjects had to assess the spatial fit of a visual stimulus (i.e. a gray-scaled picture of a comic dog) and a

simultaneously presented auditory stimulus (i.e. a barking sound). We were able to reveal two distinct networks of cortical regions preferentially processing either spatially congruent or incongruent audio-visual stimulation. Whereas low-level visual areas responded preferentially to incongruent stimulation, higher-level visual areas of the temporal and parietal cortex and prefrontal regions responded preferentially to congruent audio-visual stimulation. A position-resolved analysis revealed three robust cortical representations for each of our four visual stimulus locations in retinotopic visual regions corresponding to the representation of the horizontal meridian in area V1, and at the dorsal and ventral borders between areas V2 and V3. While these regions of interest (ROIs) did not show any significant effect of spatial congruency, we found subregions within ROIs in the right hemisphere that showed an incongruency effect (i.e. an increased fMRI signal during spatially incongruent as compared to congruent audio-visual stimulation). We interpret this finding as a correlate of spatially dispersed recurrent feedback during mismatch processing: whenever a spatial mismatch is detected in multisensory regions (such as the IPS), processing resources are re-directed to low-level visual areas.

In **chapter 3**, we investigated audio-visual object recognition using fMRI. We found distinct peak activations for diverse object categories (sportsmen, animals, cars, and tools) in ventral occipito-temporal cortex. These regions showed a category preference not only during unimodal visual stimulation but also during unimodal auditory stimulation. As we could rule out visual imagery as a potential confound, this might reflect learned associations between frequently co-occurring visual and auditory object features.

In **chapter 4**, we addressed the question whether neuronal correlates of audio-visual object recognition are modulated by semantic congruency and familiarity of the object features being integrated. We compared audio-visual integration of congruent and incongruent animal sounds and pictures with arbitrary pairings of complex artificial sounds and object images. We revealed audio-visual integration effects in different regions, including inferior frontal cortex (IFC), superior temporal sulcus/middle temporal gyrus (STS/MTG), and superior temporal gyrus (STG). For natural objects, activations in superior and lateral temporal areas were modulated by semantic congruency. While posterior STG was stronger activated by congruent than incongruent audio-visual pairings, posterior MTG preferably integrated incongruent animal sounds and images. Most importantly, we revealed overlapping audio-visual

integration effects in posterior STS and IFC for natural and artificial material. We conclude that AV integration in these areas is independent of stimulus familiarity, which indicates a core role of posterior STS and IFC in object-related audio-visual integration.

Independent of their computational function, multisensory regions presumably interact with unisensory cortices via feedback connections and thus subserve a cortical network in which semantic knowledge about objects is represented in a distributed fashion. According to this model, brain regions that process a specific type of incoming sensory information also serve as a knowledge store about the attributes of particular objects. For instance, a hammer has a number of salient attributes (such as a characteristic color, shape, sound). Ventral temporal cortex processes incoming information about visual form and color. If the incoming form and color match the stored hammer template, ventral temporal cortex identifies the object as a hammer. Lateral temporal cortex processes information about object motion. If the incoming motion information matches the stored template (e.g. up, down, impact) then the object is identified as a hammer. Frontal and parietal regions store information about the hand orientation and arm posture that would be required to grasp the hammer. The representations in these different regions are linked, so that simply seeing a photograph of a static hammer also activates the respective visual motion and motor representations. Thus, networks of neurons in auditory association cortex that respond to an auditory object are also linked to visual, tactile, and motor information about the same object represented in temporal, parietal, and frontal regions.

To better understand the rapid computations performed during object perception, future studies should combine high-resolution fMRI with methods that enable examination of the temporal sequence of information processing (for example magnetoencephalography). Such an approach should deepen our understanding of crossmodal object recognition, while the use of more natural stimulation conditions should substantially contribute to human object recognition theories with increased ecological validity.

Samenvatting

Invasieve electrofysiologische studies in dieren en non-invasieve studies in mensen leveren indrukwekkend bewijs voor de betrokkenheid van zeer verspreide hersengebieden bij zelfs eenduidige cognitieve functies. Dit is niet alleen waar voor cross-modale en sensor-motor integratie, waarbij per definitie coördinatie tussen verschillende sensorische en executieve structuren vereist is. Het is ook het geval voor unimodale perceptie, omdat sensorische systemen uit meerdere subsystemen bestaan die parallel opereren en gespecialiseerd zijn in de analyse van specifieke kenmerken van objecten. Een kernvraag binnen de cognitieve neurowetenschap is hoe deze verschillende, wijd verspreide processen gecoördineerd worden en hoe de respectievelijke computationele resultaten gebonden worden tot coherente unisensorische en multisensorische waarnemingen. De functionele magnetic resonance imaging (fMRI) studies beschreven in dit proefschrift hadden als gemeenschappelijk doel om bij te dragen aan een beter begrip van de mechanismen die ten grondslag liggen aan cross-modale audio-visuele integratie. De nadruk is gelegd op het onderzoeken van twee dimensies die van primair belang zijn voor audio-visuele integratie: spatiële (**hoofdstuk 2**) en semantische (**hoofdstukken 3 en 4**) congruentie.

Hoofdstuk 1 geeft een inleiding in zowel unisensorische (auditieve, visuele en tactiele) als multisensorische (audio-visuele en visuo-tactiele) objectverwerking. Er wordt een overzicht gegeven van humane fMRI studies waarin wordt aangetoond dat visuele, tactiele en auditieve informatie over objecten corticale associatiegebieden kunnen activeren waarvan oorspronkelijk werd aangenomen dat ze modaliteitspecifiek waren. Verwerking convergeert óf in multisensorische zones, óf via directe cross-modale interactie tussen modaliteitspecifieke cortices zonder tussenkomst van multisensorische gebieden. Deze bevindingen worden geïntegreerd met bestaande theorieën over semantische verwerking tot een algemeen mechanisme voor cross-modale objectherkenning: de rekrutering en locatie van multisensorische zones varieert afhankelijk van de informatie-inhoud, en van welke modaliteit de meest betrouwbare sensorische input levert.

Voor succesvolle integratie van visuele en auditieve informatie is het belangrijk om vast te stellen of visuele en auditieve signalen uit corresponderende locaties in de omgeving komen. In **hoofdstuk 2** rapporteren we cross-modale effecten van spatieel

congruente en incongruente audio-visuele stimulatie. Visuele en auditieve stimuli kwamen uit één van vier horizontale locaties, en proefpersonen moesten beslissen of de visuele (b.v. een zwart-wit cartoon van een hond) en auditieve (b.v. een blaf-geluid) stimuli uit dezelfde locatie kwamen of niet. We hebben hiermee aan het licht gebracht dat twee verschillende netwerken van corticale gebieden elk hun eigen voorkeur hebben voor spatieel congruente of spatieel incongruente audio-visuele stimulatie. Terwijl lagere visuele gebieden voorkeur hadden voor incongruente stimulatie, lieten hogere visuele gebieden in de temporale en pariëtale cortex, en prefrontale gebieden een voorkeur voor spatieel congruente audio-visuele stimulatie zien. Een analyse gericht op stimulus-positie onthulde drie robuuste corticale representaties voor elk van de vier visuele stimulus locaties in retinotopisch georganiseerde visuele gebieden: corresponderend met de horizontale meridiaan in gebied V1, en met de dorsale en ventrale grenzen in V2 en V3. Terwijl deze “regions-of-interest” (ROIs) geen significante effecten van spatiele congruentie lieten zien, hebben we subgebieden binnen de rechter hemisfeer ROIs gevonden die een incongruentie-effect vertoonden (d.w.z. een verhoogd fMRI signaal tijdens spatieel incongruente stimulatie t.o.v. spatieel congruente audio-visuele stimulatie). We interpreteren deze bevinding als een correlaat van spatieel verspreide terugprojecties (feedback) tijdens mismatch verwerking: steeds als een spatiele mismatch wordt gedetecteerd in multisensorische gebieden (zoals de IPS), worden verwerkingsbronnen opnieuw toegewezen aan lagere visuele gebieden.

In **hoofdstuk 3** hebben we audio-visuele objectherkenning onderzocht m.b.v. fMRI. We hebben verschillende activatiepieken gevonden voor verschillende objectcategorieën (sporters, dieren, auto’s en gereedschappen) in de ventrale occipitaal-temporale cortex. Deze gebieden lieten een categorievoorkeur niet alleen zien tijdens unimodale visuele stimulatie maar ook tijdens unimodal auditieve stimulatie. Omdat we visuele inbeelding hebben kunnen uitsluiten als mogelijke verklaring, zou het de geleerde associaties tussen veel gebruikte, en samen optredende visuele en auditieve objectkenmerken kunnen reflecteren.

In **hoofdstuk 4** hebben we ons op de vraag gericht of neuronale correlaten van audio-visuele objectherkenning beïnvloed worden door semantische congruentie en bekendheid met de objectkenmerken die geïntegreerd worden. We hebben audio-visuele integratie van congruente en incongruente dierengeluiden en -plaatjes vergeleken met arbitrair gepaarde complexe artificiële objectgeluiden en -plaatjes.

We hebben audio-visuele integratie-effecten onthuld in verschillende gebieden, waaronder de inferieure frontale cortex (IFC), superieure temporale sulcus/middelste temporale gyrus (STS/MTG), en de superieure temporale gyrus (STG). Voor natuurlijke objecten werden de activaties in superieure en laterale temporale gebieden beïnvloed door congruentie. Terwijl de posterieure STG sterker geactiveerd werd door congruente dan door incongruente audio-visuele paren, integreerde de posterieure MTG voornamelijk incongruente dierengeluiden en -plaatjes. De belangrijkste bevinding was dat audio-visuele integratie-effecten voor natuurlijke en artificiële informatie overlap vertoonden in posterieure STS en IFC. We concluderen dat AV-integratie in deze gebieden onafhankelijk is van stimulusbekendheid. Dit indiceert dat deze gebieden een kernrol spelen in object-gerelateerde audio-visuele integratie.

Het is aannemelijk dat onafhankelijk van hun computationele functie, multisensorische gebieden met unisensorische gebieden interacteren via terugprojecties, en daarmee onderdeel zijn van een corticaal netwerk waarbij semantische kennis over objecten op een verspreide manier wordt gerepresenteerd. Volgens dit model dienen hersengebieden die een bepaald type sensorische informatie verwerken ook als kennisopslag voor de kenmerken van specifieke objecten. Een hamer bijvoorbeeld heeft verschillende opvallende kenmerken (een karakteristieke kleur, vorm en geluid). De ventrale temporale cortex verwerkt inkomende informatie over visuele vorm en kleur. Wanneer de inkomende vorm en kleur overeenkomen met de opgeslagen hamersjabloon wordt het object door de ventrale temporale cortex geïdentificeerd als hamer. De laterale temporale cortex verwerkt informatie over beweging. Als de inkomende bewegingsinformatie overeenkomt met het opgeslagen sjabloon (b.v. naar boven, naar beneden, de kracht) wordt het object als hamer geïdentificeerd. In frontale en pariëtale gebieden ligt de informatie over handoriëntatie en houding van de arm opgeslagen die nodig is om een hamer vast te pakken. De representaties in deze gebieden zijn verbonden zodat alleen het kijken naar een stilstaand plaatje van een hamer de respectievelijke visuele bewegings- en motorische representaties activeert. Kort samengevat zijn dus netwerken van neuronen in auditieve associatiecortex die reageren op een auditief object ook verbonden met visuele, tactiele en motorische informatie over hetzelfde object, gerepresenteerd in temporale, pariëtale en frontale hersengebieden.

Om de snelle computaties tijdens objectherkenning beter te kunnen begrijpen, zijn toekomstige studies nodig die fMRI op hoge resolutie combineren met methoden die de temporele sequentie van informatieverwerking kunnen bestuderen, bijvoorbeeld magneto-encefalografie. Een dergelijke benadering zou ons begrip van cross-modale objectherkenning kunnen verdiepen, terwijl het gebruik van natuurlijkere stimulatieomstandigheden substantieel zou bijdragen aan meer ecologische valide theorieën over objectherkenning.