

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
TRABALHO DE CONCLUSÃO DE CURSO  
BIOMEDICINA

Gabriela Mueller De Melo

**SENTIDOS EM SINERGIA:  
EFEITO INTERMODAL DO QUE SE VÊ SOBRE O QUE SE ESCUTA**

Porto Alegre

2016

Gabriela Mueller de Melo

**SENTIDOS EM SINERGIA:  
EFEITO INTERMODAL DO QUE SE VÊ SOBRE O QUE SE ESCUTA**

Trabalho de conclusão de curso de graduação apresentado ao Instituto de Ciências Básicas da Saúde da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de Bacharel(a) em Biomedicina.

Área de habilitação: Psicobiologia

Orientador: Prof. Dr. Gustavo Gauer

Co-orientador: Dr. Kingson Man

Porto Alegre

2016

Gabriela Mueller de Melo

**SENTIDOS EM SINERGIA:  
EFEITO INTERMODAL DO QUE SE VÊ SOBRE O QUE SE ESCUTA**

Trabalho de conclusão de curso de graduação apresentado ao Instituto de Ciências Básicas da Saúde da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de Bacharel em Biomedicina.

Aprovado em 07/12/2016.

**BANCA EXAMINADORA**

Profa. Dra. Denise Zancan – UFRGS

Me. Alexandre Nobre – UFRGS

Prof. Dr. Gustavo Gauer – UFRGS (orientador)

#### CIP - Catalogação na Publicação

de Melo, Gabriela Mueller  
Sentidos em sinergia: efeito intermodal do que se  
vê sobre o que se escuta / Gabriela Mueller de Melo.  
-- 2016.  
87 f.

Orientador: Gustavo Gauer.  
Coorientador: Kingson Man.

Trabalho de conclusão de curso (Graduação) --  
Universidade Federal do Rio Grande do Sul, Instituto  
de Ciências Básicas da Saúde, Curso de Biomedicina,  
Porto Alegre, BR-RS, 2016.

1. Percepção. 2. Psicofísica. 3. Efeito intermodal.  
4. Integração multissensorial. I. Gauer, Gustavo,  
orient. II. Man, Kingson, coorient. III. Título.

## **AGRADECIMENTOS**

Meus pais sempre me permitiram a liberdade de exercer minha própria ausência de livre arbítrio, sempre me respeitando, me apoiando e confiando em mim, com muito carinho e dedicação, especialmente nas coisas mais importantes da vida. Sou infinitamente grata por isso.

Também agradeço...

Ao Antonio Damasio, ao Kingson Man e ao Max Henning.

Ao Douglas Adams, ao Richard Dawkins ao Friederich Nietzsche.

Ao punk rock, especialmente ao Bad Religion, especialmente ao Greg.

Aos meus melhores amigos: Rafa, Rafa, Bernardo, Rada e Matheus.

À minha gata Zahra.

Ao meu orientador Gustavo Gauer e ao grupo do Laboratório de Biossinais Cognitivos.

Ao programa Ciências sem Fronteiras. Fora Temer.

*Well am I making haste or could it be haste is making me  
What's time but a thing to kill or keep or buy or lose or live in*

*I gotta go faster  
Keep up the pace  
Just to stay in the human race*

*I could go supersonic  
the problem's chronic  
Tell me does life exist beyond it  
When I need to sate  
I just accelerate into oblivion  
Into oblivion*

*Now here I go again  
everything is alien  
How does it feel to be outstripped by the pace of cultural change  
My deeds are senseless  
and rendered meaningless  
When measured in that vein*

Brett Gurewitz, 2002

## RESUMO

A interação sinérgica entre os sentidos é um aspecto fundamental para a nossa interação com o mundo. A natureza multimodal do ambiente ao nosso redor se manifesta através dos variados tipos de estímulos que se fazem presentes em mais de uma modalidade sensorial. A habilidade de combinar pistas provenientes de múltiplas modalidades sensoriais é uma função primordial do sistema nervoso, impactando profundamente o comportamento e a percepção, ainda que geralmente não tomemos consciência desses processos. Considerando a natureza ubíqua dos eventos multissensoriais na percepção, o estudo científico das interações entre diferentes modalidades é imprescindível para uma melhor compreensão sobre o funcionamento do sistema nervoso como um todo. Nesse sentido, o presente trabalho foi dedicado à revisão da literatura na pesquisa multissensorial e ao desenvolvimento e execução de um experimento sobre a relação intermodal entre estímulos audiovisuais em humanos.

Na introdução compreensiva, inicialmente são apresentados alguns conceitos, definições, e os principais fundamentos subjacentes à percepção multissensorial, além de uma revisão sobre resultados experimentais de estudos prévios pertinentes. Em seguida, são levantadas questões metodológicas envolvidas no estudo das relações multissensoriais com a consciência, desenvolvendo ênfase na abordagem da metacognição. Adicionalmente, como adendo à introdução, é discutido de forma mais próxima o tema da percepção integrada na mente, apresentando um arcabouço teórico para a compreensão das bases neurais da memória e da consciência, a nível de sistemas, através de uma arquitetura hierárquica organizada em zonas de convergência divergência, proposta por Antonio Damasio.

Em seguida, é apresentado um trabalho experimental na forma de manuscrito de um artigo científico, desenvolvido em colaboração com o Brain and Creativity Institute da University of Southern California, sobre as influências intermodais que estímulos visuais semanticamente congruentes são capazes de exercer sobre a percepção auditiva em condições sub-limiais. Foram elaborados três paradigmas psicofísicos: de determinação de limiar perceptivo, de detecção e de identificação de estímulos auditivos. Por fim, são feitas conclusões sobre o trabalho como um todo e são incluídas as perspectivas futuras em decorrência dele.

## RESUMO EM LÍNGUA INGLESA

The synergistic interaction between the senses is a fundamental aspect of our interaction with the world. The multimodal nature of the environment around us is a manifestation of the various types of stimuli that present themselves in more than one sensory modality. The ability to combine cues of multiple sensory modalities is a primary function of the nervous system, profoundly impacting the behavior and perception, even though generally we are not aware of processes. Considering the ubiquitous nature of the multisensory perception, the scientific study of the interactions between different modalities plays an essential role for a better understanding about the functioning of the nervous system as a whole. In this sense, the present work was dedicated to the literature review in multisensory research and to the development and execution of an experiment on an intermodal relationship between audiovisual stimuli in humans.

In the introduction, initially, some concepts, definitions and main foundations underlying the multisensory perception are reviewed, as well as experimental results from relevant previous studies. Then, methodological questions are raised involved in the study of multisensory relations with consciousness, plus developing an approach in the area of metacognition. In addition, the integrated perception of the mind is discussed, thus presenting a theoretical framework for the understanding of memory and consciousness, at systems level, by a neuronal hierarchical organization in zones of convergence divergence, as proposed by Antonio Damasio.

Next, it is presented an experimental work as a manuscript of a scientific paper, developed in collaboration with the Brain and Creativity Institute of the University of Southern California, on whether semantically congruent visual stimuli can exert cross-modal influences on auditory perception under sub-threshold conditions. Therefore, three psychophysical paradigms were elaborated: the determination of threshold perception, the detection and identification of stimuli. Finally, conclusions are drawn about the work as a whole and are included as future perspectives accordingly.



## SUMÁRIO

1. INTRODUÇÃO COMPREENSIVA.....	9
1.1 INTEGRAÇÃO MULTISSENSORIAL.....	9
1.1.1 As portas da percepção.....	9
1.1.2 Os princípios da integração multissensorial.....	11
1.1.3 Influências espaço-temporais.....	12
1.1.4 Atribuindo semântica.....	16
1.1.5 Inferência causal.....	19
1.1.6 O que Bayes inferiria sobre a integração multissensorial.....	20
1.1.7 Sobre o que, exatamente, estamos falando: o conceito de integração.....	23
1.1.8 A integração da percepção na mente consciente.....	24
1.1.9 Abordagens experimentais e problemas decorrentes.....	27
1.1.10 Como “chegar lá”?.....	28
1.1.11 Processos sobre os próprios processos.....	29
1.1.12 As possíveis relações entre consciência e integração.....	32
1.1.13 As bases biológicas do processo da mente consciente.....	34
1.1.14 Quando o cérebro cria mapas, ele informa a si mesmo.....	37
1.1.15 As zonas de convergência-divergência.....	40
1.2 JUSTIFICATIVA.....	44
1.3 OBJETIVOS.....	44
1.3.1 Objetivo geral.....	44
1.3.2 Objetivos específicos.....	45
2. ARTIGO CIENTÍFICO (em inglês).....	46
2.1 INTRODUCTION.....	47
2.2 METHODOLOGY.....	51
2.2.1 Participants.....	52
2.2.2 Apparatus and stimuli.....	52
2.2.3 Design and procedure.....	53
2.3 RESULTS.....	54
2.4 DISCUSSION.....	55

2.5 CONCLUSIONS.....	59
2.6 APPENDIX – WHAT’S NEXT?.....	60
2.6.1 A psychophysical evaluation of the influence of vision on auditory detection and identification of objects.....	62
REFERENCES.....	64
FIGURES.....	70
3. CONCLUSÕES E PERSPECTIVAS.....	75
REFERÊNCIAS.....	76
ANEXO A – NORMAS DE PUBLICAÇÃO.....	84

## 1. INTRODUÇÃO COMPREENSIVA

### 1.1 INTEGRAÇÃO MULTISSENSORIAL

*The scientist purposed that there ain't no purpose  
And the theologian told me that it's all been designed  
And I'm trying to maintain objectivity  
The world won't illuminate what really matters  
And I'm an imperfect mortal meaning extractor  
Processing the complexity  
Greg Graffin, 2002.*

#### 1.1.1 As portas da percepção

A interação da vasta maioria dos animais com o ambiente em que estão inseridos é inerentemente multimodal<sup>1</sup>. Os diferentes sentidos são empregados, tanto sequencialmente como em paralelo, para passivamente e ativamente explorar o meio externo, confirmar expectativas e adquirir novas informações (TURK, 2014). A experiência de um organismo frente ao mundo é primariamente regida pelas suas habilidades sensoriais, e a capacidade de usar pistas ambientais provenientes de múltiplos sentidos é um aspecto fundamental da função do sistema nervoso. Aparentemente sem necessitar de muito esforço, o encéfalo reconstrói o dinâmico ambiente que o permeia a partir do influxo contínuo de informações sensoriais, para então gerar interpretações perceptivo mais precisas sobre objetos e eventos<sup>2</sup>, culminando numa experiência coerente e unificada sobre o mundo. A habilidade do sistema nervoso em organizar e combinar apropriadamente informações provenientes das diferentes modalidades sensoriais é fundamental para a experiência vida como a conhecemos. Dessa forma, o entendimento dos mecanismos neurais

---

<sup>1</sup> Os termos “multissensorial”, “multimodal” e “intermodal” são empregados indistintamente no presente trabalho, se referindo à estímulos ambientais com propriedades que ativam dois ou mais sistemas sensoriais (STEIN et al., 2014).

<sup>2</sup> Nesse trabalho, os termos “objetos” e “eventos” são usados para descrever genericamente os mais diversos estímulos sensoriais presentes no ambiente. Objetos podem descrever de forma mais clara os estímulos imóveis, enquanto eventos podem ser melhor representados por estímulos em movimento. Mesmo assim, ambos os termos podem aparecer sem distinção.

subjacentes aos processos multissensoriais apresenta caráter essencial para a compreensão da função do sistema nervoso como um todo.

Os julgamentos que muitas vezes pensamos fazer com base em informações de um único sentido, tal como a visão, frequentemente também são influenciados por sinais aparentemente irrelevantes, mas informativos, provenientes de outros sentidos, como a audição e o tato. Avaliações perceptivas raramente se referem exclusivamente a um determinado sentido, pois as diferentes vias sensoriais convergem entre si e compartilham o uso dos mesmos processos neurais que medeiam a percepção e ação. Notavelmente, os sentidos também podem ser utilizados alternadamente para compensar um ao outro quando necessário, de forma a garantir o reconhecimento do evento correspondente mesmo em detrimento de um sentido. Mesmo níveis discretos de aprimoramento nas habilidades de detecção e discriminação de sinais sensoriais podem exercer um grande impacto na probabilidade de sobrevivência de organismos em situações biologicamente relevantes, por exemplo, quando a audição e o tato podem compensar a visão prejudicada em condições ambientais de escuridão (STEIN et al., 2014). Dessa forma, o funcionamento combinado dos sentidos revela-se uma estratégia amplamente adaptativa, provavelmente tendo desempenhado um papel muito importante na história evolutiva das mais variadas formas de vida.

Pressões seletivas podem ter favorecido o desenvolvimento de diferentes sistemas sensoriais de forma a maximizar a detecção de informações em variados nichos ecológicos durante os processos evolutivos (STEIN et al., 2014). Como resultado, os sistemas sensoriais de um organismo, equipados com uma especializada maquinaria de receptores periféricos, ininterruptamente reconhecem sinais sensoriais de naturezas e magnitudes específicas. Estímulos externos podem emitir diversos tipos de energia, como radiação ultravioleta, radiação eletromagnética e ondas de pressão, sendo que eventos biológicos geralmente ocorrem através de mais de uma dessas modalidades. Cada sentido apresenta seus receptores “sintonizados” a uma forma particular de energia que, conseqüentemente, fornecem impressões sobre atributos específicos dos estímulos ambientais. Considerando-se que cada sentido revela uma perspectiva única do mundo externo, e que as diferentes informações sensoriais provenientes de um mesmo estímulo ambiental geralmente são complementares entre si, a interação sinérgica entre eles é capaz de proporcionar impressões mais completas e acuradas sobre a natureza dos eventos percebidos,

que não poderiam ser obtidas a partir de um sentido isoladamente (O'HARE, 1991; STEIN et al., 2014).

### 1.1.2 Os princípios da integração multissensorial

O termo “integração multissensorial” foi originalmente cunhado a partir de estudos de neurofisiologia para descrever observações a nível celular quando estímulos eram apresentados através de duas ou mais modalidades. De forma mais geral, refere-se a uma ampla classe de processos em que a informação proveniente de diferentes sentidos é combinada, resultando num produto que é único, e não pode ser simplesmente desconstruído em seus componentes originais (STEIN et al., 2009). A integração multissensorial pode ser definida operacionalmente como uma diferença estatisticamente significativa entre a resposta evocada por uma combinação de estímulos multimodais e a resposta evocada pelo mais eficaz dos estímulos unimodais componentes. Do ponto de vista neuronal, a comparação é feita entre o número total de impulsos ou taxas de disparo evocados por estímulos sensoriais e suas combinações. Ao modificar a magnitude da resposta neuronal, a integração de sinais sensoriais pode induzir alterações na percepção que, por sua vez, podem influenciar o comportamento. Diferenças no desempenho comportamental podem ser analisadas através de variações na precisão e velocidade de detecção, localização e identificação de estímulos (ERNST; BANKS, 2002; SHAMS et al., 2005).

O modelo de integração multissensorial historicamente mais bem estudado em termos neurofisiológicos é o colículo superior de gatos, uma estrutura bilateral localizada na região do mesencéfalo. Neurônios individuais do colículo superior respondem a sinais provenientes da visão, audição e somatossensação dentro de regiões espacialmente circunscritas e sobrepostas no ambiente (B E Stein & Arigbee, 1972). A principal função dessa estrutura é guiar comportamentos de detecção, localização e orientação a alvos salientes no ambiente. Estudos pioneiros da neurofisiologia do colículo superior fundamentaram três princípios gerais de funcionamento da integração multissensorial, descritos no livro “*The merging of the senses*” (STEIN; MEREDITH, 1993):

- 1) O princípio espacial prevê que a integração multissensorial é mais provável ou mais forte quando os estímulos unissensoriais constituintes são provenientes da mesma localização (ou de localizações aproximadas).
- 2) De acordo com o princípio temporal, a integração multissensorial é mais provável ou mais forte quando os estímulos unissensoriais constituintes surgem ao mesmo tempo (aproximadamente ao mesmo tempo).
- 3) O princípio da eficácia inversa descreve que a integração multissensorial é mais provável ou mais forte quando os estímulos unissensoriais constituintes evocam respostas relativamente mais fracas quando apresentados isoladamente.

O princípio da efetividade inversa garante que a integração seja potenciada em situações onde a presença dos sinais sensoriais é mais escassa ou menos confiável (STEIN; MEREDITH, 1993) como uma estratégia para maximizar a informação disponível no momento. De acordo com os outros dois princípios, sinais multimodais detectados numa janela espacial e temporal estreita geralmente intensificam as respostas dos neurônios multissensoriais, tendo como pressuposto que sinais próximos no espaço e no tempo provavelmente foram emitidos pelo mesmo evento original. Por outro lado, discrepâncias espaciais e temporais mais acentuadas tendem a causar a depressão da resposta ou o impedimento da integração dos sinais (MEREDITH; STEIN, 1986; MEREDITH et al., 1987; KADUNCE et al., 1997, 2001). Particularmente no âmbito temporal, é observada uma janela relativamente ampla para a integração multissensorial, relacionada à compensação de pequenas disparidades temporais entre os sinais sensoriais emitidos em diferentes modalidades. Essa propriedade possivelmente se deve ao fato de que neurônios demoram algum tempo para processar as informações. Enquanto que o disparo do potencial de ação dura apenas alguns milissegundos, as correntes de voltagem excitatórias ou inibitórias podem durar várias centenas de milissegundos, abrindo assim espaço para uma maior adaptabilidade na identificação de correspondência temporal entre os sinais (HOLMES; SPENCE, 2005).

### 1.1.3 Influências espaço-temporais

Um amplo conjunto de evidências baseadas majoritariamente na pesquisa psicofísica, mas também eletrofisiológica e de neuroimagem, têm demonstrado o papel das influências espaciais e

temporais na integração multissensorial sob variadas perspectivas experimentais. Por exemplo, a influência da audição na percepção visual foi ilustrada num estudo clássico envolvendo uma tarefa de classificação de intensidade luminosa (MEREDITH; STEIN, 1986). A presença de um estímulo auditivo co-localizado a um estímulo visual exerceu forte influência no aumento da percepção de luminosidade, enquanto que, quando o estímulo visual foi espacialmente deslocado, um estímulo auditivo não exerceu mais nenhum efeito. Além disso, nos casos em que os estímulos audiovisuais se encontravam co-localizados, a influência da audição foi mais pronunciada nos níveis menores de intensidade luminosa e mais amena nos níveis mais intensos, o que corrobora com o princípio da efetividade inversa.

Apesar da maioria da pesquisa multissensorial envolver efeitos de outros sentidos sobre a visão, um experimento análogo mais recente testou a influência visual na percepção auditiva (ODGAARD et al., 2004). Através de uma tarefa de classificação de intensidade sonora de ruído branco acompanhado por um flash de luz, encontrou-se que participantes julgaram o ruído como mais alto quando acompanhado pelo flash. Efeitos háptico-visuais foram demonstrados pela melhora na performance numa tarefa de detecção de contraste pela presença de um estímulo tátil congruente (GROEN, VAN DER et al., 2013). Também já foi verificada a interação entre tato e audição, pela observação de um efeito facilitador exercida por estímulos táteis de vibrações sincrônicas na detecção do volume de estímulos auditivos (GILLMEISTER; EIMER, 2007).

Outro exemplo de interação intermodal na percepção auditiva foi demonstrado que, quando a luz temporalmente precedeu o som, a sensibilidade na tarefa de detecção auditiva melhorou, mas quando a luz foi apresentada após o som, a performance mudou muito pouco (em comparação a condições sem apresentação visual). Além disso, nas condições em que a luz estava posicionada de forma mais periférica à fonte sonora observou-se um efeito maior no aumento da detecção do alvo auditivo, também em acordo com o princípio da efetividade inversa (SCHIRILLO, 2011). Estímulos visuais podem exercer influências no processamento auditivo tanto sobre potenciais de campo como na atividade de neurônios individuais, ainda nos córtices auditivos primários e secundários (KAYSER et al., 2008). Além disso, as interações audiovisuais parecem depender do perfil temporal relativo dos estímulos, com os efeitos mais robustos ocorrendo quando os estímulos visuais foram apresentados a frente dos auditivos por 20 a 80 milissegundos, em concordância com os resultados de Schirillo, 2011. Adicionalmente, o efeito facilitador intermodal pode estar

correlacionado com o restabelecimento de oscilações neurais lentas para um ângulo de fase de excitabilidade ótima.

Num paradigma desenvolvido por Sekuler, Sekuler, & Lau, 1997, a inclusão de um estímulo sonoro no momento da coincidência entre dois objetos idênticos (pequenos discos) em movimento constante promoveu a percepção de balanço. Quando o som foi apresentado em até 150ms após o estímulo visual, ainda foi capaz de aumentar a probabilidade da percepção do balanço logo após a coincidência. Notavelmente, esses resultados evidenciam a existência de uma tolerância considerável para a assincronia temporal entre os estímulos multissensoriais para que sejam integrados. Outro estudo encontrou que os participantes exibiram uma performance mais rápida e mais acurada quando um tom sonoro foi apresentado junto ao alvo visual, mesmo quando o ritmo sonoro havia sido corrompido por um *jitter* temporal (deslocando a apresentação do som no tempo em proporções sutis) (VROOMEN; GELDER, DE, 2000).

Apesar da existência de uma janela temporal relativamente flexível para a ocorrência da integração multissensorial, o efeito da sincronia entre os sinais sensoriais pode também exercer uma influência impactante na atividade neuronal e no desempenho comportamental. Relacionando a ativação encefálica medida através de imagem por ressonância magnética funcional à performance numa tarefa de detecção de alvos auditivos e visuais, um estudo (MARCHANT et al., 2012) observou que a sincronia audiovisual aumentou a resposta dependente do nível de oxigênio sanguíneo (do inglês *blood-oxygen-level dependent*, BOLD) no sulco temporal superior posterior e numa rede mais ampla de regiões, incluindo o giro temporal superior e estruturas subcorticais, sem interferência de se os estímulos fossem previsíveis ou não. Junto a isso, a performance na tarefa de detecção de alvos de alta intensidade foi também aumentada pela sincronia audiovisual.

Correspondências espaciais e temporais são fatores capazes de exercer impactos importantes sobre os mecanismos integrativos, mas enquanto essas influências podem em grande parte ser explicadas pelas propriedades dos campos receptivos de neurônios multissensoriais em estruturas subcorticais e corticais, outros mecanismos de integração, como a correspondência de conteúdo semântico entre as modalidades, operam em um nível de maior complexidade. Notavelmente, um estudo desenvolvido por Chen & Spence, 2011 demonstrou que as janelas temporais em que a simultaneidade (fator espaço-temporal) é percebida e em que a congruência semântica exerce influência podem ser dissociadas e correspondem necessariamente a mecanismos unificados. Além disso, esse estudo observou que integração de sinais congruentes ocorreu numa



janela temporal mais flexível, enquanto que propriedades incongruentes foram limitada por uma janela temporal mais reduzida para integração. O valor semântico de um objeto ou evento possivelmente envolve uma ampla rede de informações relacionadas aos estímulos e suas diversas propriedades. Essas situações podem ser representadas em abordagens experimentais através de condições experimentais "congruentes" (semanticamente correspondentes) e "incongruentes" (semanticamente não-correspondentes).

Num paradigma de rivalidade binocular<sup>3</sup>, a inclusão de um som congruente pode favorecer a dominância e reduzir a supressão de um estímulo visual correspondente (padrões visuais na forma de grade) (KANG; BLAKE, 2005). Além disso, em outro estudo (CONRAD et al., 2010) sinais auditivos em movimento foram capazes de resolver o conflito visual e impulsionar a percepção do estímulo em movimento coerente. A inclusão de estímulos táteis (objetos sinusoidais ranhurados) congruentes em frequência espacial com estímulos visuais em rivalidade binocular (na forma de padrões de grades ortogonais) promoveu a dominância e reduziu a supressão do estímulo com a mesma orientação (LUNGHI et al., 2010). Pistas olfativas congruentes também podem influenciar a dinâmica da rivalidade binocular. A influência do olfato no processamento visual também pode acontecer inconscientemente na supressão induzida por flash contínuo<sup>4</sup> (ZHOU et al., 2010). Adicionalmente, correspondências intermodais também foram capazes de exercer influência sobre a integração durante o aprendizado da associação entre formas e sons (BRUNEL et al., 2015), levando a formação de novas unidades aprendidas que tem estabilidades diferentes ao longo do tempo, com o decaimento do efeito intermodal sendo mais rápido para estímulos incongruentes do que para estímulos congruentes no aprendizado.

Apesar de fornecerem resultados interessantes de uma forma geral, os estudos referenciados no último parágrafo, assim como muitos outros, empregaram a manipulação de estímulos como formas visuais ou tons auditivos, os quais se relacionam a um grau muito fraco de variação

---

<sup>3</sup> Rivalidade binocular: É um fenômeno que ocorre quando estímulos monoculares dissimilares são apresentados a localizações correspondentes nas retinas dos dois olhos de um observador. Em vez de perceber uma mistura estável e única dos dois estímulos, o observador experimenta alternâncias na percepção ao longo do tempo à medida que os dois estímulos competem pelo domínio perceptivo. A rivalidade binocular é um exemplo de percepção multiestável em que uma estimulação fisicamente invariante leva a flutuações na percepção. É também um dos poucos fenômenos psicofísicos que têm sido explorados para estudar o processamento visual fora da consciência, outros fenômenos úteis nesse sentido incluem o mascaramento por meta-contraste, a cegueira induzida por movimento e a supressão de flash (BLAKE; TONG, 2008).

<sup>4</sup> Supressão por flash contínuo: É um fenômeno da percepção visual em que uma imagem apresentada a um olho é removida da percepção visual pela apresentação abrupta, na forma de um "flash", de outra imagem apresentada ao outro olho (TSUCHIYA, 2008).

semântica. A congruência semântica e o efeito contextual no processamento de informação que se dá pelas associações retiradas da vida cotidiana (como objetos e eventos de mais elevada complexidade) fornecem uma perspectiva abstrata diferente da simples consistência física (por exemplo, tempo e direção do movimento) entre estímulos intermodais (DOEHRMANN; NAUMER, 2008). Para acessar funções integrativas mais complexas do cérebro, é apropriada a inclusão de variações também mais complexas de estímulos semânticos, em que as associações multissensoriais foram estabelecidas no curso de aprendizagem prévia do indivíduo e mais fidedignas a situações reais.

#### 1.1.4 Atribuindo semântica

Nas últimas décadas, a pesquisa multissensorial tem investido na execução de paradigmas experimentais envolvendo de estímulos de valor semântico mais complexo. Empregando estímulos visuais de pássaros e carros num paradigma de rivalidade binocular, Y. C. Chen, Yeh, & Spence, 2011 demonstraram que a modulação semântica do contexto auditivo é aditiva com a atenção e o contraste, sendo que cada tipo de influência pode afetar aspectos diferentes da performance comportamental. O fator atencional foi capaz de modular a troca da percepção de uma figura para a outra, sendo que o contexto semântico apenas exerceu efeitos modulatórios na troca quando foi manipulado em combinação com a atenção. Já o contraste somente exerceu influência na percepção da primeira imagem dominante, sendo que essa modulação pode estar relacionada mais simplesmente ao mecanismo de inibição interocular. A influência exercida pela atenção pode ser explicada através do estabelecimento de conexões de *feedback*, enquanto que o fator da congruência semântica provavelmente atua fortalecendo a representação visual do objeto, possivelmente através de projeções *top-down*.

A modulação semântica contextual fornecida por uma trilha sonora congruente ou incongruente foi capaz de influenciar a percepção predominante da figura biestável “minha mulher ou minha sogra”, em que percepção inicial e a dominância da imagem da “sogra” foram mais frequentes quando os participantes ouviram a voz de uma mulher velha, e o inverso aconteceu para a mulher jovem. As influências da congruência e atenção seletiva na percepção inicial puderam ser dissociadas, sugerindo que a congruência semântica audiovisual é, por si só, suficiente para

modular a primeira percepção visual, porém a atenção seletiva parece ser o fator principal a modular a competição em curso (HSIAO et al., 2012).

Mesmo não fornecendo nenhum tipo de informação específica sobre a localização, sons semanticamente congruentes puderam facilitar a localização visual de objetos (IORDANESCU et al., 2008). Contudo, quando a figura do objeto foi substituída pelo seu nome escrito, a influência intermodal auditiva foi eliminada. Disso foi concluído que os sons modulam a ativação visual de forma objeto-específica, atuando em neurônios visuais responsivos as características visuais do objeto, invés de atuarem diretamente pela ativação de representações semânticas de nível mais elevado. Assim, sugeriu-se que quando as respostas visuais de ordem superior baseadas num objeto (por exemplo, um gato) são reforçadas pela co-apresentação de um som característico (por exemplo, um miado), essas respostas em troca poderiam fortificar o feedback das respostas aos estímulos-alvo em áreas visuais primárias. Ao observar diferenças entre as latências temporais necessárias para o acesso das representações audiovisuais envolvendo sons naturalísticos e palavras faladas, outro estudo (CHEN; SPENCE, 2011a) propôs que esses estímulos podem não necessariamente acessar distintos mecanismos semânticos, mas que o processamento requerido para acessar suas representações semânticas armazenadas pode ser a diferença. Dessa forma, sons e imagens podem acessar suas representações semânticas primeiro, e somente então o nome associado, enquanto que palavras faladas e escritas podem acessar suas representações léxicas primeiro e então seus significados associados.

Sinais naturalísticos complexos formados pelos estímulos audiovisuais da fala são também bastante explorados na pesquisa multissensorial. A visualização dos lábios do falante pode fazer com que o ouvinte extraia melhor informações acústicas úteis entre ruído sonoro, o que resulta num aumento adicional na inteligibilidade audiovisual, que vai além da leitura labial em si (SCHWARTZ et al., 2004). Numa tarefa de detecção de frases faladas mascaradas por ruído, observou-se a influência exercida pela apresentação de uma frase escrita na forma de *priming*<sup>5</sup> (GRANT, 2001). Na condição audiovisual incongruente, os limiares de detecção reduziram por 1.6dB em relação a condição somente auditiva, demonstrando que a co-modulação intermodal pode oferecer proteção do mascaramento por ruído. Além disso, foi observada a correlação entre a área

---

<sup>5</sup> *Priming* se refere a um efeito da memória implícita em que a exposição a um estímulo apresentado muito brevemente (*prime*) exerce influência na resposta ao mesmo estímulo, ou um estímulo relacionado (alvo), apresentado em seguida, podendo alterar parâmetros como velocidade ou acurácia na identificação. Esse efeito pode ocorrer pela repetição de estímulos perceptivo, semânticos ou conceituais (SCHNEIDER et al., 2008).

da boca abrindo e as regiões espectrais de cada frase, sugerindo que a magnitude do efeito depende tanto das propriedades temporais como espectrais dos sinais-alvo da fala (Grant & Seitz, 2000).

Utilizando uma combinação de métodos eletrofisiológicos e comportamentais, um estudo (WASSENHOVE, VAN et al., 2005) caracterizou a influência do discurso visual em potenciais relacionados a eventos (ERPs) auditivos. Foram encontrados dois estágios de influência intermodal: um estágio caracterizado pelo encurtamento da latência auditiva (facilitação da latência dos ERPs), condicionado pela saliência do estímulo visual, e um estágio perceptivo de unidade, refletido pela redução da amplitude para o discurso audiovisual (redução dos potenciais evocados auditivos). Esses resultados estão de acordo com uma integração intermodal inicial e sugerem que sinais visuais carregam um código preditivo específico nos substratos neurais que medeiam o processamento da fala.

Corroborando com a hipótese de que sinais acústicos modulam a atividade cerebral num nível inicial do processamento visual, possivelmente por amplificar o sinal visual e facilitar seu processamento consciente, Alsius & Munhall, 2013 utilizaram a supressão por flash contínuo para demonstrar que um estímulo visual da fala, que é somente processado pré-conscientemente, pode ser integrado com uma informação auditiva coerente supra-limiar. Um paradigma similar demonstrou um efeito facilitador intermodal de categorização de palavras (PLASS et al., 2014), em que participantes responderam mais rapidamente para palavras correspondentes ao movimento dos lábios das faces sob mascaramento.

Tanto sinais visuais como também sinais táteis foram demonstrados capazes de influenciar a percepção do discurso e tornar a detecção auditiva mais eficiente (TJAN et al., 2014). Os limiares de detecção foram reduzidos nas condições multissensoriais, mesmo que os estímulos visuais e táteis sozinhos não fossem capazes de informar a resposta correta. Contudo, a variação nos níveis do aumento da eficiência suportam o refinamento do conhecimento sobre o estímulo auditivo ao longo dos testes. Além disso, a análise das funções psicométricas sugeriu que os estímulos visuais e táteis reduziram a incerteza sobre o sinal auditivo, dessa forma rejeitando a hipótese de integração neuronal multissensorial e processamento neural específicos ao discurso como explicações para o aumento na detecção sobre condições de ruído. Em vez disso, é sugerida uma forma de interação multissensorial onde sistemas sensoriais de outra forma irrelevantes para a tarefa informam o sistema auditivo sobre quando escutar.

### 1.1.5 Inferência causal

Na maior parte do tempo, os seres vivos se encontram envoltos por múltiplas fontes de informações sensoriais e, imerso nesse contexto amplamente multissensorial, seus respectivos sistemas nervosos são constantemente engajados na resolução de dois desafios computacionais fundamentais. Um deles é o problema da ligação para a integração dos sinais provenientes de uma fonte comum, gerando uma percepção mais confiável dos estímulos ambientais. Contudo, como nem todos os sinais existentes no ambiente são causados pelo mesmo objeto, nem todos eles devem ser associados. Outra questão que o cérebro precisa resolver é determinar quais são os sinais que se originaram a partir do mesmo evento e, portanto, devem ser integrados, para processá-los de forma diferente dos sinais que provavelmente têm origens distintas: o problema da inferência causal (KAYSER; SHAMS, 2015).

O problema de como combinar dois sinais envolve uma inferência, seja implícita ou explícita, sobre se eles são causados pelo mesmo objeto ou por objetos diferentes. O inevitável ruído presente no ambiente e no sistema nervoso faz com que as informações sejam muitas vezes inconsistentes e ambíguas, de forma que sempre existe um certo nível de incerteza sobre a informação que é obtida pelos nossos sentidos e, hipoteticamente, várias outras realidades diferentes poderiam dar origem aos mesmos sinais sensoriais (KORDING, 2014). Sob a suposição de uma fonte de sinal comum, as estimativas unissensoriais de uma propriedade devem ser combinadas e ponderadas de acordo com suas probabilidades relativas (isto é, o inverso da variância). Sob a hipótese de duas fontes diferentes, os sinais auditivos e visuais são tratados de forma independente. Uma forma pela qual o sistema nervoso pode inferir a estrutura causal do mundo é a partir dos sinais sensoriais, e múltiplos tipos de correspondências intermodais podem fornecer informações sobre a provável origem dos sinais numa fonte comum ou em fontes independentes.

Para explicar detalhadamente os mecanismos pelos quais o cérebro faz a discriminação dos casos em que os estímulos nas diferentes modalidades não estão correlacionados, várias hipóteses já foram sugeridas. Contudo, todas compartilham a ênfase na dependência das correlações espaço-temporais, em que estímulos em diferentes modalidades que ocorrem ao mesmo tempo e no mesmo lugar (ou em tempos e locais similares) devem ser tratados como referindo-se ao mesmo evento externo. De uma forma geral, a probabilidade de combinação perceptiva de dois sinais tipicamente

depende do conjunto de coerência espacial, temporal e estrutural entre eles, bem como do conhecimento prévio obtido pela experiência da combinação de sinais na natureza e das expectativas relacionadas (MURRAY; WALLACE, 2012; KAYSER; SHAMS, 2015).

Um estudo recente com fMRI investigou como acontece a inferência causal em estruturas encefálicas e o impacto de expectativas prévias. Foi encontrado que a região do sulco frontal inferior exibiu atividade ao combinar as expectativas de congruência (*top-down*) com as pistas de congruência ligadas aos estímulos físicos (*bottom-up*) para decidir entre a integração e a segregação de sinais audiovisuais. Mais especificamente, nesse experimento os participantes assistiram a filmes audiovisuais de sílabas fonologicamente congruentes, incongruentes ou tipo McGurk<sup>6</sup> que poderiam ser integradas em uma percepção ilusória. As expectativas de congruência *top-down* foram manipuladas pela apresentação de estímulos de McGurk embutidos em blocos de sílabas congruentes ou incongruentes. Como resultado, os participantes se mostraram mais propensos a integrar os sinais audiovisuais em um percepto ilusório de McGurk em contextos congruentes do que incongruentes. A nível neural, o sulco frontal inferior esquerdo apresentou ativações aumentadas para sinais incongruentes *bottom-up* em relação aos sinais congruentes. Além disso, a atividade no sulco frontal inferior esquerdo foi aumentada para estímulos McGurk fisicamente idênticos. Esse aumento na atividade para a segregação perceptiva foi amplificado quando os participantes esperavam sinais audiovisuais incongruentes, com base nas suas experiências sensoriais anteriores (GAU; NOPPENY, 2016).

### 1.1.6 O que Bayes inferiria sobre a percepção multissensorial

No século XIX, Helmholtz, 1867 descreveu a percepção como uma “inferência inconsciente” que mapeia sinais sensoriais contaminados por ruído com interpretações perceptivo e escolhas sobre a orientação das experiências prévias. De forma similar, modelos estatísticos bayesianos formalizaram a percepção como uma inferência probabilística pela qual o cérebro combina expectativas prévias com evidências sensoriais incertas para inferir o estado mais

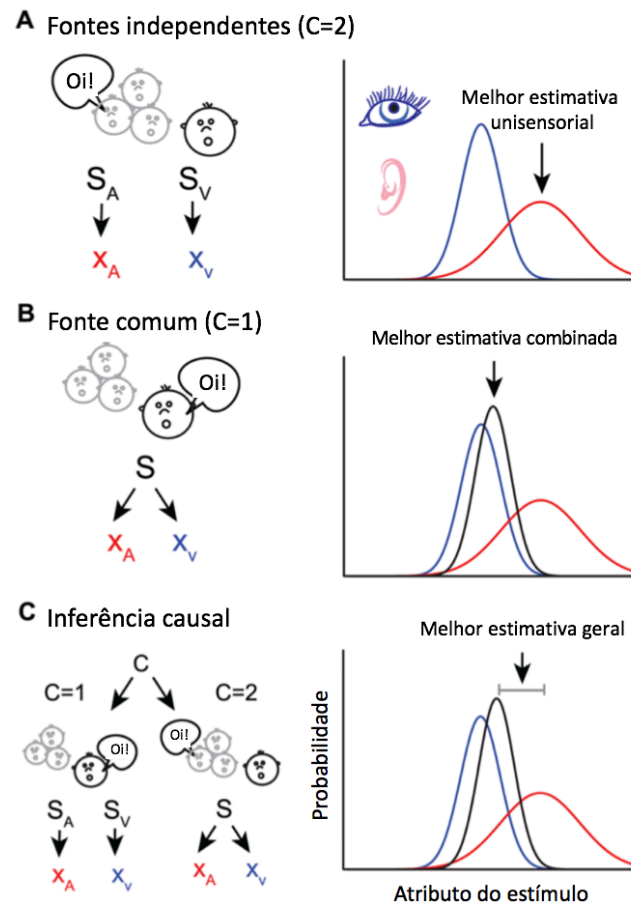
---

<sup>6</sup> O efeito McGurk é uma ilusão perceptivo descrita por McGurk & MacDonald, 1976 que demonstra a interação entre a audição e a visão na percepção da fala. Essa ilusão ocorre quando o componente auditivo é pareado com o componente visual de outro som, levando a percepção de um terceiro som.

provável do mundo. A inferência causal bayesiana modela explicitamente as possíveis estruturas causais que podem ter originado dois sinais. É um modelo normativo que descreve como um observador deve decidir entre integração ou segregação de informações para computar uma estimativa de uma propriedade ambiental. Na estatística bayesiana, a probabilidade é interpretada como um “nível de convicção” de que um evento vai ocorrer ou de que uma hipótese é verdadeira. Essa abordagem é principalmente associada com o processo de atualização de suposições prévias sobre uma hipótese na luz de novos dados. A essência da teoria de Bayes é uma forma de pensar sobre a probabilidade em si, a qual permite estimar um parâmetro estatístico ou uma propriedade do ambiente a partir de observações experimentais ou informações sensoriais (KORDING, 2014).

Aplicada ao estudo da percepção, a estatística bayesiana fornece uma estimativa da probabilidade de integração ou segregação de informações sensoriais. Cada modalidade sensorial, considerada de forma independente, pode ser compreendida como fonte de uma estimativa probabilística (contaminada por ruído) de um mesmo evento original (KAYSER; SHAMS, 2015). Por exemplo, mesmo que dois sinais sejam gerados por uma fonte comum, eles podem não ser temporariamente sincrônicos ou espacialmente co-localizados com total precisão. Além disso, diferenças nas velocidades de transmissão da luz e do som induzem variabilidade nos tempos de chegada de sinais visuais e auditivos no nível que depende da distância da fonte física do observador. Sempre existe algum grau de incerteza sobre as informações sensoriais, mas, no entanto, é possível reduzir consideravelmente essa incerteza e gerar estimativas mais confiáveis ao empregar conhecimentos adquiridos em experiências prévias e através da combinação de informações sensoriais no curso temporal (KORDING, 2014).

Pelo entendimento do modelo bayesiano, a inferência de uma causa comum é basicamente produto de dois fatores: a probabilidade da ocorrência de dois sinais sensoriais no caso de uma causa comum, chamada de *likelihood*, e a expectativa da ocorrência baseada em experiências prévias, chamada de *prior*. A medida de *likelihood* é uma função de quão semelhantes são os dois sinais sensoriais e envolve a presença de ruído sensorial. Quanto mais dissemelhantes forem os dois sinais, menor será a *likelihood*. Já o *prior* compreende as estatísticas dos eventos e envolve as expectativas a priori de uma causa comum, sendo uma função das experiências prévias, ou seja, o quão frequente dois sinais são causados pela mesma fonte no geral (MURRAY; WALLACE, 2012).



*Figura 1.* Modelos bayesianos de integração multissensorial. Esquemas de diferentes estruturas causais no ambiente dando origem a sinais auditivos e visuais (por exemplo, vendo uma face e ouvindo uma voz) que podem ou não ser originados do mesmo falante. Os painéis da esquerda mostram a estrutura causal inferida. Os painéis na direita mostram as distribuições de probabilidade de que essas representações sensoriais e a estimativa ótima do atributo do estímulo derivaram do mesmo modelo bayesiano sobre diferentes premissas sobre o ambiente. Para simplicidade dessa ilustração, é assumido que a probabilidade prior do atributo do estímulo é uniforme. (A) Assumindo fontes separadas ( $C=2$ ) leva a estimativa de localização independente dos estímulos visuais e acústicos, com o valor ótimo correspondendo a localização unissensorial mais provável. (B) Assumindo uma fonte comum ( $C=1$ ) leva a integração (fusão). A estimativa Bayesiana ótima é a combinação de estimativas visuais e acústicas, cada uma pesada pelas suas confiabilidades relativas. (C) Na inferência causal bayesiana (assumindo uma estratégia de “modelo de média”), as duas hipóteses diferentes sobre a estrutura causal (por exemplo, uma ou duas fontes) são combinadas, cada uma pesada pelas suas probabilidades inferidas dadas as sensações visuais e acústicas. A estimativa ótima do estímulo é uma mistura de estimativas unissensoriais e combinadas. Adaptado de Kayser & Shams, 2015.



Esse modelo estatístico pode ser estendido para o problema da inferência causal ao incluir inferências sobre a estrutura causal do ambiente. Dependendo da tarefa que o sistema nervoso tem de resolver, diferentes estratégias podem ser usadas para derivar estimativas de atributos sensoriais com base nas probabilidades de cada possível estrutura causal (KÖRDING et al., 2007). Para minimizar o erro numa estimativa perceptiva, a estimativa ideal é a média ponderada não-linear de dois termos: uma estimativa derivada do pressuposto de que dois sinais são provenientes de uma fonte única (fusão), e uma estimativa derivada do pressuposto de que eles têm fontes distintas (segregação), com cada uma sendo ponderada pela probabilidade da respectiva estrutura causal. Esta estratégia é conhecida como "modelo de média" (Figura 1). Alternativamente, a estimativa da estrutura causal mais provável pode ser inferida como a estimativa final, numa estratégia denominada "seleção de modelo" (KAYSER; SHAMS, 2015).

#### 1.1.7 Sobre o que, exatamente, estamos falando: o conceito de integração

Uma definição fenomenológica para integração é a combinação de diferentes características numa percepção unificada - o problema de ligação. Através de uma perspectiva mais cognitiva, a integração pode ser definida pela geração de uma representação abstrata não-perceptiva que associa sinais distintos a um novo constructo. Outra definição, mais formal e matemática, é a integração da informação contruída globalmente por um sistema, que vai além da informação fornecida pela união das partes constituintes. Na prática, pode-se tratar dos processos integrativos como os quais dois ou mais estímulos distintos são combinados em um produto único. Em todos os casos, o simples processamento de estímulos individuais não proporciona a integração, é necessária a combinação dos estímulos e o efeito mútuo que cada estímulo tem sobre o processamento do outro (MUDRIK et al., 2014).

Conforme descrito em um recente artigo de revisão, os conteúdos multissensoriais podem ser apresentados sob três diferentes formas (DEROY et al., 2016). As situações mais estudadas de integração multissensorial referem-se a situações em que em que uma única propriedade é percebida através de duas ou mais modalidades sensoriais, ou seja, quando diferentes sentidos fornecem informações redundantes sobre propriedades específicas, como a localização espacial ou o movimento de um evento. Ao integrar informações redundantes de sistemas sensoriais diferentes,

o cérebro pode formar estimativas mais confiáveis sobre as propriedades de um evento. Contudo, a integração de informações redundantes não está necessariamente vinculada à consciência multissensorial, e pode resultar em episódios de consciência unissensorial unfluenciados pela informação fornecida por outra modalidade - seja ela conscientemente percebida ou não. Dessa forma, relatos de percepção multissensorial poderiam na verdade constituir casos de um viés intermodal na consciência unissensorial, tanto quanto casos genuínos de consciência unificada multissensorial. Hipoteticamente, a primeira possibilidade poderia ser excluída experimentalmente pela demonstração de que a integração de dois sinais sensoriais leva à experiência consciente de uma nova propriedade, que não poderia ser experimentada por uma simples conjunção de episódios unissensoriais (MUDRIK et al., 2014; DERROY et al., 2014).

Além da integração de informações sobre uma mesma propriedade, dois outros tipos de situações também precisam ser consideradas quando se estuda a integração multissensorial. Numa delas, duas modalidades podem contribuir para a percepção de um objeto, mas através de propriedades diferentes ou complementares desse mesmo objeto. Assim, os dois componentes unissensoriais necessitam ser referidos ao mesmo objeto ou evento. Esses casos remetem a desafios diferentes daqueles em que os conteúdos experienciados por duas modalidades são os mesmos. Nessa situação os dois conteúdos podem permanecer unissensoriais, mas a consciência multissensorial parece atribuí-los à mesma origem. Além dos casos em que diferentes sinais sensoriais contribuem para a percepção de uma mesma propriedade ou objeto, uma terceira categoria consiste nas situações em que dois objetos em diferentes modalidades, ou mesmo dois objetos multissensoriais, são experimentados como parte da mesma cena multissensorial. Nesses casos, todos os vários objetos ou eventos unissensoriais e multissensoriais presentes num contexto ambiental são diferentes e, contudo, todos são experimentados como parte do mesmo cenário contextual (NOPPENY et al., 2016).

#### 1.1.8 A integração da percepção na mente consciente

No estudo científico da consciência, é dada grande ênfase à integração. A integração é considerada lado a lado à consciência, refletindo tanto a natureza unificada e holística da experiência consciente quanto a suposição de que a consciência é necessária para integração. Muito

se fala sobre a percepção integrada e coerente do ambiente, culminando numa experiência consciente única, mas é preciso ir muito além da metáfora para de fato compreender os mecanismos subjacentes (DEROY et al., 2014). Intuitivamente, a união entre os sentidos é geralmente considerada condição necessária para a experiência consciente. Por outro lado, a revisão de dados empíricos recentes sugere que não há dependência absoluta da integração na consciência. Nesse quadro, é atualmente foco de debate a questão da associação entre experiências perceptivas e consciência multissensorial (DEROY et al., 2016). Ainda assim, uma hipótese cautelosa e provável é que a necessidade de consciência para que ocorra a integração é maior quanto mais complexos ou desconhecidos forem os estímulos sensoriais (MUDRIK et al., 2014).

Existe uma crucial distinção conceitual entre o que as pessoas relatam e o que elas estão fenomenalmente conscientes sobre, ou ao que elas direcionam atenção. Embora não existam dúvidas sobre a ocorrência de relatos sobre a experiência de objetos e cenas multissensoriais, ainda permanece em aberto a questão se esses relatos correspondem ao que está presente na consciência em determinado momento, ou ao que é reconstruído através de outros processos de agregação de informações experienciadas em momentos diferentes (DEROY et al., 2016). Essa é a diferença crucial entre a consciência multissensorial genuína e outros processos onde a informação unissensorial consciente pode ser unida. Sem dúvida, tal questão é mais fácil de definir conceitualmente do que experimentalmente. Por exemplo, Spence & Bayne, 2015 questionam se os relatos multissensoriais devem ser tomados como um episódio consciente unificado, ou talvez sejam na verdade resultado de uma rápida mudança de atenção entre episódios unissensoriais conscientes. Nesse último caso, a atribuição de sinais sensoriais a um único objeto não dependeria da consciência do objeto multissensorial, mas de algo como um efeito de atração atencional.

Apesar de que, na prática, a investigação da consciência multissensorial não é fator central em muitos estudos que documentam interações multissensoriais, uma série de estudos recentes têm sugerido na verdade uma dissociação entre a ocorrência da integração multissensorial e a ocorrência da consciência multissensorial, em que experiências perceptivo que parecem ser modalidade-específicas também podem ser influenciadas pela atividade de outras modalidades mesmo na ausência da consciência sobre essa interação. Por exemplo, ao investigar como sinais audiovisuais são integrados, um estudo (ZMIGROD; HOMMEL, 2011) observou que a associação entre os estímulos foi totalmente independente da experiência consciente. Num paradigma apresentando estímulos audiovisuais em diferentes níveis de assincronia, participantes deveriam

julgar explicitamente se as duas características pertenceriam ao mesmo ou aos diferentes eventos. Os relatos de sincronia diminuíram com o aumento da assincronia, contudo, o mais notável foi a observação que estímulos audiovisuais puderam ser integrados quando percebidos tanto como ocorrendo juntos quanto como eventos separados. Esses resultados sugerem que a experiência consciente de unidade não é pré-requisito ou uma consequência direta da integração multissensorial. Corroborando com essa hipótese, observou-se que uma trilha sonora congruente foi capaz de facilitar o processamento inconsciente de uma cena visual complexa subliminar (sob supressão por flash) (TAN; YEH, 2015). O trabalho realizado por Faivre, Mudrik, Schwartz, & Koch, 2014 confirmou esses resultados em que sempre foi encontrado um efeito facilitador das relações semânticas audiovisuais de estímulos subliminares. Contudo, a integração multissensorial inconsciente não ocorreu quando os participantes tinham passado por treinamento consciente prévio da tarefa, sugerindo que, após aprendizado consciente, o processamento inconsciente pode ser suficiente para a integração multissensorial.

Um estudo conduzido por Palmer & Ramsey, 2012 sugeriu outro princípio fundamental sobre a função da consciência na integração multissensorial: que efeitos intermodais podem ocorrer na ausência de consciência, mas a modalidade de influência deve ser percebida conscientemente para sua informação para afetar outras modalidades. Adicionalmente, a consciência de um objeto não parece ser necessária para a ocorrência da atenção seletiva ao objeto (CHOU; YEH, 2012). Conforme demonstrado através de um paradigma de supressão por flash contínuo, condições com e sem supressão visual levaram a resultados similares, fornecendo evidências para a ocorrência da atenção baseada no objeto sobre estados inconscientes, da mesma forma com que é observado em consciência. Foi sugerido que a orientação da atenção a objetos inconscientes pode apresentar um papel fundamental em muitos processos de maior complexidade, e que mecanismos inconscientes podem implicar no reconhecimento de objetos em diferentes níveis. Objetos inconscientes na verdade podem estar sujeitos ao mesmo problema de ligação que objetos conscientes, e a orientação atencional é mecanismo candidato para explicar a ligação inconsciente.

### 1.1.9 Abordagens experimentais e problemas decorrentes

A maioria dos estudos que tem abordado a emergência da consciência da percepção em contextos multissensoriais investiga como um sinal proveniente de outra modalidade sensorial pode modular o acesso à consciência visual. Frequentemente são empregadas abordagens experimentais nas quais os estímulos visuais são suprimidos da consciência visual, através de uma variedade de paradigmas psicofísicos, como a cegueira atencional, o mascaramento e a percepção multiestável (incluindo paradigmas de percepção de figuras ambíguas, rivalidade binocular e supressão por flash contínuo) (DEROY et al., 2016). Contudo, grande parte dos métodos aplicados na pesquisa multissensorial foram na verdade estendidos a partir de métodos desenvolvidos na pesquisa unimodal. Dessa forma, muitas vezes o que é estudado consiste simplesmente em experiências unimodais sendo apresentadas juntas no estado de consciência e, por consequência, essas abordagens experimentais podem apenas demonstrar com sucesso uma modulação intermodal sobre a consciência unissensorial dos participantes (DEROY et al., 2014).

A maioria dos estudos na pesquisa multissensorial envolve a apresentação de um estímulo que é ambíguo ou difícil de identificar em uma modalidade e, ao mesmo tempo, de um estímulo claro, não-ambíguo, em outra modalidade. A lógica é que se os sinais sensoriais das duas modalidades são integrados, o estímulo mais claro deve influenciar a interpretação do estímulo ambíguo, dando origem a percepção mais nítida desse estímulo. Esses resultados são importantes ao demonstrar que a informação apresentada numa modalidade pode interferir na ocorrência de consciência perceptiva dos estímulos apresentados em outra modalidade. Contudo, nada realmente pode ser inferido sobre a natureza multissensorial da consciência através desses experimentos. Ao mesmo tempo em que é inquestionável que indivíduos reportem a experiência de objetos a partir de duas ou mais modalidades, situações multissensoriais levantam um desafio particular, em que o que deve ser evidenciado não é a ocorrência de um episódio consciente, mas a ocorrência de um certo tipo de consciência dada pelo percepto multissensorial unificado (DEROY et al., 2014).

A abordagem baseada em procurar casos de contraste pelo relato subjetivo é amplamente utilizada no campo da pesquisa multissensorial, e é o método que atualmente mais se aproxima de investigar a ocorrência da consciência multissensorial. O princípio é estabelecer quando os relatos subjetivos são consistentes com a experiência de um único objeto integrando as características processadas em duas (ou mais) modalidades e contrastá-las com relatos de duas características

separadas ou eventos em duas modalidades. No entanto, os relatos perceptivos podem indicar a ocorrência da consciência audiovisual, mas ainda assim também são compatíveis com a própria experiência em uma modalidade sensorial, em que a experiência em uma modalidade poderia muito bem ter sido influenciada pela segunda fonte de informação apresentada em outra modalidade que, no entanto, não forma um episódio consciente multissensorial unificado (DEROY et al., 2014). Em outras palavras, que os relatos de experiência multissensorial correspondem a uma consciência que é genuinamente multissensorial em natureza, mediada por processos em ligação ou integração, não é uma associação necessariamente decorrente.

Na avaliação de respostas comportamentais, frequentemente observa-se que participantes apresentam maior velocidade e maior acurácia de performance em tarefas multissensoriais, com relação a condições unissensoriais. Por outro lado, no nível neuronal, a ocorrência da integração multissensorial é evidenciada pelo fato de que respostas de neurônios únicos, populações neuronais ou *voxels* específicos (em estudos de neuroimagem), diferem entre uma configuração multissensorial e a soma das respostas unissensoriais observadas (STEIN; MEREDITH, 1993). Contudo, ainda assim todas essas evidências falham ao explicar algo sobre a natureza da experiência consciente do observado. O fato de que os participantes relatam ter a experiência de apenas um objeto multissensorial quando os estímulos componentes são apresentados dentro de uma determinada janela temporal e/ou espacial, pode significar basicamente que a proximidade espaço-temporal é necessária para os participantes julgarem que as duas informações se referem a um mesmo objeto (DEROY et al., 2014).

#### 1.1.10 Como “chegar lá”?

Apesar dos recentes esforços para elucidar a natureza consciência perceptiva, ainda permanece uma lacuna crítica entre o contexto multissensorial da consciência e o seu estudo através de abordagens majoritariamente unissensoriais baseadas nos relatos perceptivos dos participantes. É necessário ter cautela e fazer uma análise crítica mais profunda ao considerar que casos de contraste sejam capazes distinguir casos em que as informações provenientes de duas modalidades sensoriais são genuinamente experienciadas em conjunto e casos em que são experimentadas uma após a outra e meramente julgadas provenientes da mesma fonte e pertencerem à mesmo objeto.

Para obter de fato algum progresso, o estudo da consciência multissensorial não deve se limitar ao estudo de relatos qualitativos perceptivos sobre objetos e eventos (DEROY et al., 2014).

Na ausência de um bom modelo experimental estabelecido, é importante considerar a possibilidade hipotética de que a co-atribuição da percepção a um único objeto seja uma questão de representações não-conscientes, que mantém o controle de, e preveem uma relação de coocorrência e co-localização entre duas propriedades. Assim, não se deve deixar de considerar que os relatos multissensoriais, a priori, requerem acesso multissensorial e, nesse contexto, é alvo de interesse científico compreender como conteúdo e acesso podem se separar. Enquanto existem boas evidências de que observadores são capazes de monitorar conteúdos multissensoriais, uma questão-chave é explorar como esses conteúdos se manifestam na consciência ou são mantidos fora da consciência (DEROY et al., 2016).

Os sentimentos metacognitivos são uma categoria alternativa para investigar em relação à consciência que se manifesta em situações multissensoriais, mas que parecem ter desaparecido dos relatos experimentais desde os anos 70. Os pesquisadores já não parecem querer perguntar aos seus participantes o que sentem em relação a um experimento. Mesmo assim, não podemos esquecer que perguntar como o participante se sente também é fazer ciência – desde que sejam feitas perguntas metodologicamente adequadas. Muito recentemente, alguns poucos estudos sobre consciência perceptiva têm se recentrado em questões metacognitivas clássicas. Dentro dessa abordagem, pesquisadores se dedicam a investigar em que medida os participantes são capazes de reconhecer suas habilidades e desempenho perceptivo, e a perguntar o quão confiantes se sentem em relação à percepção de um objeto único e, também, a prestar atenção às diferenças individuais em relatos subjetivos. Esses são implementos desejáveis na pesquisa multissensorial atual, ainda que por enquanto pouco explorados, com potencial de fornecer perspectivas promissoras sobre a emergência da integração multissensorial, percepção e consciência (DEROY et al., 2014, 2016).

#### 1.1.11 Processos sobre os próprios processos

Metacognição se refere aos processos cognitivos sobre outros processos cognitivos: saber sobre o que se sabe, ou desenvolver crenças sobre as próprias crenças, podendo operar em inúmeros domínios, como na percepção, na memória, na tomada de decisão coletiva e no aprendizado social.

Um aspecto central da metacognição é a habilidade de um indivíduo em reconhecer o seu próprio processamento cognitivo, ou seja, monitorar suas próprias decisões e representações em termos de acurácia e incerteza. Abordagens experimentais analisam de que forma observadores são capazes de avaliar a incerteza ou precisão de suas representações e julgamentos perceptivos, reconhecendo assim seu próprio desempenho em tarefas comportamentais (FLEMING; LAU, 2014; DEROY et al., 2016). Na pesquisa multissensorial, a análise de sentimentos metacognitivos tem sido sugerida como um instrumento útil para investigar as relações da consciência que se manifesta em situações multissensoriais (DEROY et al., 2014; NOPPENY et al., 2016).

Modelos estatísticos bayesianos descrevem a percepção como uma inferência probabilística pela qual o cérebro combina expectativas prévias com evidências sensoriais, sempre acompanhadas a um determinado grau de incerteza, para integrar os sinais de uma causa comum e segregar os de causas independentes e, assim, inferir o estado mais provável do mundo (KERSTEN et al., 2004). Conforme já abordado nas seções 2.5 e 2.6, a percepção é inerentemente incerta e sujeita a erro. A percepção dos estímulos ambientais produz representações multissensoriais de primeira ordem de objetos e eventos, enquanto que a metacognição se refere à formação de representações de segunda ordem, abstratas, que permitem aos observadores monitorarem suas representações de primeira ordem referentes ao mundo real. Ao mesmo tempo que o ambiente natural complexo impõe desafios à percepção de primeira ordem, também impõe desafios as representações metacognitivas de segunda ordem (DEROY et al., 2016).

Num sentido mais amplo, a metacognição caracteriza a habilidade do observador em introspectar sobre os processos de inferências perceptivo que levaram as suas representações de primeira ordem sobre o ambiente (DEROY et al., 2016). Através das representações metacognitivas, observadores são capazes de avaliar a incerteza ou acurácia perceptiva sobre as representações perceptivas e julgamentos que são formadas diretamente com base em evidências sensoriais. A pesquisa em metacognição investiga particularmente o quão eficaz é um observador ao acessar suas próprias incertezas. A precisão metacognitiva pode ser medida experimentalmente com base na fidelidade dos julgamentos de confiança dos participantes com relação aos seus desempenhos objetivos numa tarefa. São consideradas questões como a confiança adequada dos observadores quanto à exatidão de suas escolhas perceptivas e, eventualmente, o uso dessa informação para ajuste das respostas subsequentes (DEROY et al., 2016). Alguns métodos podem



ser aplicados para acessar as capacidades metacognitivas, como a estimativa do “grau de confiança”, a “aposta pós-decisão” e o “jogo sem perda”.

Pela estimativa do grau de confiança, observadores podem avaliar a sua confiança sobre a exatidão da sua decisão numa escala numérica. Na aposta pós-decisão, é solicitado que os observadores façam apostas sobre a exatidão de suas escolhas relatadas. Valores mais elevados devem ser apostados quando se está mais confiante sobre a exatidão da sua decisão, a fim de maximizar os ganhos. No jogo sem perda (PERSAUD et al., 2007; CLIFFORD et al., 2008), os observadores precisam escolher se preferem receber uma recompensa de acordo com a exatidão de sua escolha perceptiva ou dependendo de uma loteria com probabilidades pré-especificadas. Evidências baseadas no uso desses três métodos (Massoni, Gajdos, & Vergnaud, 2014; Dienes & Seth, 2010; Clifford et al., 2008; Persaud et al., 2007) demonstram que observadores humanos, e não-humanos, podem de fato acessar a incerteza de suas representações perceptivas e ajustar sua confiança decisional de acordo. Além disso, algumas vezes os observadores calculam sua confiança sobre a correção de seu julgamento perceptivo de acordo com o modelo ótimo de Bayes, ou seja, o nível de confiança reflete a probabilidade de que as escolhas perceptivas sejam corretas, dados os sinais sensoriais a que são expostos (AITCHISON et al., 2015). Criticamente, a confiança decisional depende do nível de incerteza sobre representações perceptivas de primeira ordem que foram geradas sobre objetos e eventos ambientais. Por exemplo, quando expostos a sinais sensoriais muito fracos, observadores terão performance perto do nível de chance tanto nas tarefas de discriminar as propriedades do estímulo quanto ao julgar se sua resposta a essas propriedades foi correta. Dessa forma, pode-se dizer que a sensibilidade perceptiva de um observador restringe a sua sensibilidade metacognitiva máxima possível (MANISCALCO; LAU, 2012; FLEMING; LAU, 2014).

Na teoria de detecção do sinal clássica, do tipo 1, a probabilidade conjunta é dada por  $P(\text{resposta}, \text{estímulo})$ , e parâmetros dessa distribuição, como a sensibilidade perceptiva  $d'$ , estão relacionados a eficácia do observador ao discriminar estados objetivos do mundo. Já na teoria de detecção do sinal do tipo 2 (CLARKE et al., 1959) os índices de confiança estão condicionados às respostas do observador (corretas ou incorretas), e não ao estado objetivo do mundo. Nesse caso, as medidas de sensibilidade metacognitiva podem ser reduzidas a operações de uma distribuição conjunta  $P(\text{confiança}, \text{acurácia})$  (FLEMING; LAU, 2014). O meta- $d'$  foi proposto recentemente como um índice para quantificar a sensibilidade metacognitiva dos observadores. Um grande meta-

$d'$  indica que os observadores são capazes de discriminar de forma confiável seus julgamentos perceptivos corretos dos incorretos. Uma vez que o  $\text{meta-}d'$  depende tanto da qualidade da evidência sensorial como da avaliação metacognitiva, a comparação direta entre o  $d'$  perceptivo e o  $d'$  metacognitivo é uma forma de quantificar a eficiência metacognitiva dos observadores (MANISCALCO; LAU, 2012). Esse método pode fornecer evidências sobre a capacidade dos observadores em avaliar a incerteza de suas representações e escolhas perceptivas. Um observador “metacognitivamente ideal” é aquele para o qual o  $\text{meta-}d'$  se iguala ao  $d'$ , ou seja, aquele pode acessar todas as informações que foram usadas para o julgamento perceptivo de primeira ordem para sua avaliação metacognitiva de segunda ordem (FLEMING; LAU, 2014).

A perspectiva metacognitiva frente a complexidade da percepção multissensorial impõe questões e desafios que vão além do que é estudado a partir de tarefas unissensoriais mais simples. Um ponto crítico para investigação futura consiste em determinar até que ponto os relatos de estimativas de incerteza perceptiva reflete a incerteza causal percebida em tarefas comportamentais (DEROY et al., 2016). Alguns estudos começaram a abordar mais diretamente a questão da incerteza metacognitiva ou estimativas de confiança na percepção multissensorial. Por exemplo, um recente estudo psicofísico demonstrou que os observadores avaliaram corretamente a exatidão de seus julgamentos de ordem temporal nos índices de confiança (KEANE et al., 2015), sugerindo que a precisão das estimativas da relação temporal audiovisual é acessível à metacognição.

#### 1.1.12 As possíveis relações entre consciência e integração

Quase todos os possíveis tipos de relação já foram postulados no que diz respeito à consciência e à integração: alguns consideram a integração pré-requisito para qualquer experiência consciente (ENGEL et al., 1999; TREISMAN, 2003), outros afirmam que a consciência apresenta função integrativa (DAMASIO, 1999; BAARS, 2002; FAHRENFORT; LAMME, 2012; DEHAENE et al., 2014) enquanto outros ainda suportam uma relação de necessidade e suficiência bidirecional entre integração e consciência (EDELMAN; TONONI, 2000; BALDUZZI; TONONI, 2008).

De um lado, está a hipótese de que a integração (especialmente a perceptiva) é necessária para a consciência (ENGEL et al., 1999; KANWISHER, 2001; KOCH, 2012). Ou seja, a percepção

consciente não pode ocorrer na forma da experiência de unidades separadas de características/qualidades (*qualia*), fora de um todo unificado. O principal argumento a favor dessa ideia é retirado da conceituação intuitiva da experiência consciente como sendo holística, indivisível e resultante da combinação de múltiplas fontes de informação através do espaço, tempo e modalidades sensoriais. Contudo, atualmente, ainda não é aparente como esta afirmação poderia ser abordada empiricamente. Até que testes empíricos nesse sentido se tornem possíveis, deve-se considerar a afirmação de que a integração é necessária para a consciência como uma hipótese razoável, mas que ainda precisa de validação.

Partindo do ponto de vista oposto, existe a hipótese de que a consciência seja necessária para a integração. Inúmeros papéis funcionais podem ser atribuídos à consciência, mas todas essas funções compartilham a noção de que a consciência é necessária para a integração de vários tipos de informação, provavelmente através de conexões de *feedback* de longo alcance (Lamme, 2006). Enquanto isso, o processamento inconsciente seria mantido encapsulado, em termos informativos (de acordo com a Teoria do Espaço Neural Global, do inglês *Global Neural Workspace Theory*) (DEHAENE; CHANGEUX, 2011; DEHAENE et al., 2014). Contudo, e de acordo com o que sugerem resultados experimentais discutidos aqui anteriormente, é possível que a necessidade da consciência não se aplique a todos os tipos de integração, e que formas mais simples de integração possam ser realizados inconscientemente.

Em particular, a Teoria da Informação Integrada postula que informações conscientes são o mesmo que informações integradas, de forma que o nível de consciência de um sistema é equivalente à sua habilidade de integrar a informação (TONONI, 2008, 2010). Contudo, na sua forma atual, essa teoria não inclui predições comportamentais sobre processamentos conscientes e inconscientes. Enquanto isso, a Teoria do Espaço Neural Global (Dehaene & Nacchache, 2001; Dehaene & Changeux, 2011) sugere que a consciência ocorre quando uma amplificação atencional *top-down* mobiliza redes frontoparietais, transmitindo sinais neurais através de todo o cérebro. Assim, essa atividade faria com que os sinais neurais se tornassem disponíveis para uma variedade de processos, incluindo categorização perceptiva, memorização de longo prazo, avaliação de processos linguísticos e ação intencional. Essa teoria foi fundamentalmente baseada na Hipótese de Acesso Global de Baars (BAARS, 2002; BAARS et al., 2013), mas vai além ao sugerir mecanismos neurais específicos que auxiliam no acesso global da consciência.

Embora estudos recentes tenham evidenciado que a consciência não seja necessária para alguns tipos de integração, ainda assim ela parece desempenhar um papel importante nos processos integrativos. A consciência pode permitir a integração de informações em distâncias e durações prolongadas, viabilizando a integração em níveis semânticos mais elevados, através de múltiplas modalidades e facilitando a formação de novas associações. No caso da formação de novas associações, é possível que a consciência exerça um papel facilitador, estabelecendo mecanismos de integração que posteriormente não requerem mais a consciência do estímulo. Ademais, a abrangência dos processos integrativos inconscientes parece ser mais limitada, resultando em efeitos que são menores em comparação com os processos conscientes. Tomadas em conjunto, o conjunto de evidências obtidos até hoje não são suficientes para sustentar a afirmação de que a consciência não desenvolve papel funcional na integração. Em vez disso, a pesquisa na área da consciência e percepção multissensorial sugere que alguns processos integrativos podem ocorrer na ausência de consciência, enquanto que outros requerem a sua presença (MUDRIK et al., 2014).

#### 1.1.13 As bases biológicas do processo da mente consciente

A função cerebral em geral, e a percepção em particular, tem sido considerada como altamente modular há mais de um século. Embora a frenologia, muito oportunamente, seja considerada obsoleta hoje em dia, o conceito geral de que o cérebro é composto de compartimentos dedicados a funções únicas e independentes umas das outras segue sendo o modelo dominante, especialmente no contexto da percepção (PASCUAL-LEONE; HAMILTON, 2001). Supõe-se que as diferentes modalidades sensoriais se encontram organizadas em vias separadas e independentes no córtex cerebral, e que a informação é processada de forma autocontida até que os sinais sejam suficientemente "bem digeridos", para então convergirem em alguma área de processamento superior nos córtices de associação, onde finalmente ocorre a unificação da percepção multimodal (MURRAY; WALLACE, 2012). Contudo, enquanto que a pesquisa entre 1960 e 1980 preocupou-se principalmente com a identificação de módulos separados na mente/cérebro, hoje existe um crescente entendimento de que a interação entre os componentes de uma estrutura num âmbito global é tão importante quanto fracionar essa estrutura em suas partes componentes.

A integração multissensorial é um veemente exemplo que ilustra a necessidade superar a noção da modularidade, sendo que os princípios dessa integração podem inclusive refletir uma arquitetura geral que se estende através dos diversos domínios da percepção multissensorial (DRIVER; SPENCE, 2000). Um crescente conjunto de estudos de neuroimagem e eletrofisiologia têm sugerido que as interações intermodais não são restritas a áreas de associação mais elevadas. As áreas classicamente consideradas como unissensoriais podem de fato ser ativadas, ou pelo menos moduladas, pela estimulação de vários outros sentidos (GHAZANFAR; SCHROEDER, 2006; KAYSER; LOGOTHETIS, 2007). Estudos de potenciais evocados e potenciais relacionados a eventos indicaram que a modulação da atividade visual por som ocorre em áreas corticais visuais iniciais (SHAMS et al., 2000; ARDEN et al., 2003). Utilizando fMRI, Watkins, Shams, Josephs, & Rees, 2007; Watkins, Shams, Tanaka, Haynes, & Rees, 2006 demonstraram que um sinal auditivo concordante intensificou a atividade nos córtices visuais V1, V2 e V3 durante a ilusão de flash induzida pelo som. Em concordância com esses achados, Noesselt et al., 2007, também usando fMRI, relataram efeitos da correspondência temporal audiovisual no córtex visual primário. Enquanto que fortes evidências têm sugerido a modulação pelo som na atividade do córtex como resultado da integração audiovisual, em troca estímulos visuais também podem modular o disparo de neurônios em córtices auditivos primários e secundários, tanto no nível de potenciais de campo como na atividade de unidade única (Kayser et al, 2008).

Levando em conta a frequência de situações em que diferentes sentidos são estimulados, pode ser intuitivo pensar que o que está junto na mente também está junto em um lugar do encéfalo onde diferentes aspectos sensoriais de fundem<sup>7</sup>. Essa perspectiva logicamente implica que todos os aspectos de processamento sensorial experienciados na mente aconteçam numa única estrutura do cérebro. Ou seja, que o que está junto na mente também deve estar junto no cérebro. Uma série de regiões corticais já bem estabelecidas podem responder a estímulos em diferentes modalidades: os

---

<sup>7</sup> Nota sobre um caso particular: o estudo da neurofisiologia do colículo superior de gatos é a principal fonte do que se sabe sobre o desenvolvimento da integração multissensorial a nível celular. Contudo, é necessário considerar que seus neurônios trimodais tenham ação mais específica à estrutura a que pertencem, e não podem simplesmente ter seus mecanismos fisiológicos extrapolados para a função de regiões encefálicas. O colículo superior é uma estrutura do tronco encefálico que relaciona mapas visuais, auditivos e somatossensoriais, sendo que, junto a outras estruturas, como o núcleo do trato solitário e o núcleo parabraquial, é um dos primeiros fornecedores de mapas completos do corpo para o sistema nervoso central. Mais do que estar diretamente vinculado à experiência multissensorial, o colículo superior e outros núcleos subcorticais podem funcionar como precursores dos sentimentos primordiais e consequentemente dos processos da mente e subjetividade, que depois culminam no córtex cerebral – o que não é uma função em nada menos importante. Para as informações se tornarem parte da consciência, elas precisam estar na forma de imagens, e o único lugar capaz de produzir essas imagens são os córtices sensoriomotores primários.

córtices associativos multimodais nas porções posteriores dos lobos temporal e parietal. Contudo, o principal argumento contra a ideia de que existe um lugar no cérebro responsável pela fusão das informações multissensoriais numa experiência única, é que simplesmente não existe nenhuma região no cérebro humano capaz de processar, simultaneamente, representações de todas as modalidades sensoriais ativas no momento da experiência concomitante de diferentes estímulos sensoriais num registro temporalmente e espacialmente perfeito. O tipo de integração que ocorre nas regiões associativas multimodais do córtex, tomada isoladamente, pouco provavelmente constitui a verdadeira base para a geração de uma consciência multissensorial, uma vez que lesões nessas regiões não corrompem a experiência de uma mente integrada (DAMASIO, 1994).

Talvez os efeitos intermodais possam acontecer independentemente da fusão absoluta da informação sensorial num único local, e que, em vez disso, um sentido possa exercer influência no outro através de mecanismos sinápticos paralelos, até mesmo através de regiões diferentes, possivelmente coordenados de forma hierárquica e dependentes de coincidência temporal. Os córtices associativos provavelmente exercem papel crucial na coordenação dessa estrutura hierárquica, mas, invés de serem a fonte direta da experiência multissensorial, suas contribuições se deem pela influência na atividade das regiões que podem tornar a experiência multissensorial explícita, ou seja, acessível para a consciência.

Um conjunto de evidências recentes sugere que é preciso sobrepor-se à noção tradicional de convergência unidirecional para explicar a percepção multissensorial. É possível que as áreas de processamento em que culmina a convergência multimodal também influenciem níveis tradicionalmente considerados unimodais através de mecanismos de *feedback*, “descendo” de volta na hierarquia. As áreas cerebrais tradicionalmente consideradas como unimodais talvez somente assumam essa característica em termos de suas projeções aferentes. As projeções provenientes das áreas de convergência multimodais podem permitir que respostas em áreas cerebrais unimodais sejam moduladas por estímulos provenientes de uma segunda modalidade. Assim, é possível que a própria experiência subjetiva dentro de uma modalidade seja afetada pela estimulação através de outra modalidade. De uma forma geral, os mecanismos de feedback divergentes podem culminar numa percepção determinada de forma multimodal, mas que ainda assim possui as propriedades unimodais associadas com a ativação de áreas cerebrais que recebem aferências periféricas de apenas um sistema sensorial (DRIVER; SPENCE, 2000).

Partindo da perspectiva de que o nosso forte senso de integração da mente é criado a partir da ação orquestrada de sistemas de larga-escala, contando com conjuntos de atividade neural temporalmente sincronizados em regiões cerebrais diferentes, ao final da década de 1980 Antonio Damasio propôs um abrangente mecanismo para explicar as bases neurobiológicas da mente, compreendendo a geração de mapas e imagens sensoriais e a organização da informação em zonas de convergência e divergência como os substratos neurais da memória e consciência.<sup>8</sup>

#### 1.1.14 Quando o cérebro cria mapas, ele informa a si mesmo

Talvez a habilidade mais distintiva do sistema nervoso humano seja a sua excepcional habilidade de criar mapas. Eles são construídos o tempo todo, a cada momento que interagimos com objetos, animais ou lugares, e também são (re)construídos toda vez que recordamos memórias de todosos tipos. A produção de mapas da realidade topograficamente organizados nos córtices sensoriais permite a experiência dessas representações na forma de imagens mentais, que constituem o principal elemento da mente consciente. A informação contida nos mapas pode também permanecer inconsciente, contribuindo para a geração de comportamentos e respostas motoras eficientes, e exercendo papel essencial para um gerenciamento mais sofisticado dos processos fisiológicos e manutenção da vida. A nossa interação com o complexo mundo ao nosso redor depende fundamentalmente das capacidades de aprendizado e recordação dos mais diversos objetos e eventos. Nós reconhecemos pessoas e lugares por que estabelecemos alguma forma de registro de suas características e trazemos a informação contida em alguns desses registros de volta no momento oportuno. Além disso, a habilidade de imaginar possíveis eventos futuros, que constitui a base do nosso raciocínio e planejamento, também depende do aprendizado e da recordação. Apesar da ubiquidade da formação de mapas para a função do sistema nervoso humano, explicar exatamente como o mapeamento acontece não é uma tarefa fácil, visto que mapas de forma nenhuma constituem meras cópias dos estímulos ambientais (DAMASIO, 2010).

A suposição de que o cérebro faz registros passivos de objetos e eventos do ambiente, armazenando informações sobre suas características e preservando esses registros para recordar

---

<sup>8</sup> As ideias descritas no texto daqui para frente foram fortemente e descaradamente influenciadas pelo trabalho de Antonio Damasio. Fundamentalmente, ele é o autor das ideias propostas, eu apenas concordo com ele e eventualmente faço reflexões relacionadas. Para escrever essa parte, eu me inspirei em seus livros e artigos científicos referenciados.

mais tarde, não passa de ficção. O organismo interage com objetos, e o cérebro reage a essa interação. Mais do que fazer registros da estrutura de um objeto/evento, o encéfalo na verdade registra as múltiplas consequências da interação do organismo dessa objeto/evento. Antes de mais nada, os sinais do ambiente precisam percorrer as estruturas sensoriais na periferia do corpo, eles não podem simplesmente entrar no cérebro diretamente, ou seja, as informações do meio externo ao corpo só podem chegar ao cérebro via o próprio corpo. Ao passo que o corpo e o ambiente interagem um com o outro, as alterações que essas interações causam no corpo são mapeadas em regiões sensoriomotoras primárias do encéfalo. Dessa forma, a memória sobre um determinado objeto tipicamente é composta pelas atividades sensoriais e motoras relacionadas à interação entre o organismo e esse objeto durante um certo período de tempo. Além disso, as nossas memórias sobre objetos e eventos também são influenciadas pelos nossos conhecimentos prévios acerca de estímulos e situações similares (DAMASIO, 2010).

Depois de formadas, as imagens também não são estocadas permanentemente como cópias completas de objetos, eventos, palavras ou frases. Considerando a grande quantidade de conhecimento que é geralmente adquirido ao longo da vida de um indivíduo, uma forma de estoque permanente provavelmente levaria a problemas intransponíveis de capacidade de armazenamento. É impossível que mapas sensoriais subjacentes às imagens sejam armazenados em seu formato completo original, primariamente pois os córtices primários estão ocupados demais para isso, constantemente construindo novos mapas atualizados sobre o ambiente atual. Essa forma de armazenamento também provavelmente levaria a problemas na eficiência da recordação. Sempre que se recorda de um evento, não se obtém uma reprodução exata, mas uma interpretação, como uma nova versão reconstruída do evento original. Ao longo da vida, com a aquisição de novas experiências, as versões sobre as mesmas coisas são transformadas (DAMASIO, 2010).

A consequência mais espetacular do incessante mapeamento feito pelo encéfalo é a emergência da mente consciente na forma de imagens mentais. As imagens da nossa mente são constituídas pelo conjunto de mapas cerebrais momentâneos sobre, literalmente, tudo (ou pelo menos sobre tudo que de fato importa para nós). São incluídos mapas das condições internas do nosso corpo (interocepção), mapas dos estímulos nos arredores (exterocepção), mapas concretos, mapas abstratos, mapas atuais e mapas previamente registrados na memória. As imagens representam as propriedades físicas de entidades e suas relações espaciais e temporais entre elas e com o organismo, assim como as ações decorrentes dessas relações (DAMASIO, 2010). O processo



da mente constitui-se de um fluxo contínuo dessas imagens, sendo que todo tipo de conhecimento factual emerge na mente na forma de imagens.

Contudo, a formação de representações sensoriais no córtex não é suficiente para a consciência das imagens mentais por si só. Para correlacionar as imagens ao próprio “eu”, é necessária a inclusão do componente “eu”, ou seja, a noção de subjetividade. As imagens perceptivas precisam ser associadas a outras imagens, que constituem as bases neurais do *self*, e estão fundamentalmente relacionadas ao mapeamento dos estados internos do corpo (DAMASIO, 2010). Os mapas interoceptivos começam a ser montados nos núcleos do tronco encefálico, que recebem informações sobre o meio interno do corpo (estado visceral) através do nervo vago e pela lâmina I da via espinotalâmica. Por carregarem informações relacionadas à regulação da homeostase do organismo, esses mapas têm papel essencial na geração dos sentimentos primordiais (veja A. Damasio & Carvalho, 2013).

As imagens da mente são construções internas, e tudo o que se pode saber é que essas imagens são reais para quem tem experiência delas. É muito (muito) improvável que algum dia poderemos saber o que a realidade é em absoluto. Nós tomamos consciência da realidade com base nas imagens construídas no nosso próprio cérebro. Essas imagens são representações da nossa forma de interagir e perceber o mundo. Se ao longo da evolução nosso organismo tivesse desenvolvido sistemas sensoriais com propriedades diferentes, a nossa percepção, as nossas imagens mentais e a nossa noção de realidade também seriam diferentes. Jamais saberemos como. Contudo, existe uma notável consistência na habilidade de construção de imagens entre os seres humanos, e muitos outros animais. Sendo a experiência diretamente relacionada à emergência de imagens na mente, é coerente inferir que a forma com que experienciamos o mundo é comparável entre nossos coespecíficos e, em maior ou menos grau, similar a organismos de outras espécies.

Talvez não será uma possível conquista para a espécie humana algum dia desvendar o que a realidade realmente é, como vista “do lado de fora”. Mas, talvez, tal ambição seja na verdade irrelevante. No final das contas, o que importa é o que é real para nós mesmos e para os demais organismos que provavelmente compartilham essa mesma perspectiva. E talvez não seja necessário, nem justificável, tentar ir além disso. Ainda assim podemos tentar entender como nós mesmos funcionamos em termos biológicos, investigar como as imagens da nossa mente são construídas e buscar compreender os substratos neurais do processo subjacentes a nossa

perspectiva sobre a realidade. Talvez o *hard problem* da consciência não seja tão duro assim, ou pelo menos, talvez não seja impossível desconstruí-lo aos poucos - de baixo para cima.

#### 1.1.15 As Zonas de Convergência-Divergência

Ao final da década de 80, Antonio Damasio propôs um modelo de arquitetura neural para explicar o reconhecimento e a recordação, cujo componente essencial são as zonas de convergência-divergência (CDZs). Sobre o panorama das CDZs, redes neurais similares são ativadas quando objetos ou eventos são processados em termos perceptivo e quando são recordados na memória. Mas, antes de mais nada, o que, exatamente, são as CDZs? As CDZs não devem ser entendidas como um conceito simbólico ou metafórico, elas consistem em grupos de neurônios. Nos níveis mais altos da hierarquia de processamento, esses conjuntos de neurônios, ao invés de armazenarem representações explícitas de mapas sobre objetos e eventos, registram a coincidência de atividade neuronal proveniente de diferentes lugares do cérebro e contêm instruções para a reconstrução desses registros quando necessário. É importante enfatizar que nenhuma parte do mapa sobre um objeto precisa ser permanentemente estocada. Para que um objeto seja memorizado, apenas a coincidência dos sinais neuronais ligados ao mapa sensorial provocado por esse objeto precisa ser representada nas CDZs (DAMASIO, 1989; MEYER; DAMASIO, 2009).

As imagens construídas durante a percepção devem ser reconstruídas durante a recordação e a imaginação, não como réplicas perfeitas e não necessariamente tão vívidas, mas como aproximações mais tênues. Para explicar o mecanismo de reconstrução do mapa original e geração da recordação, Damasio propôs um mecanismo de retro-ativação, em que é necessário que os componentes de um mapa sejam retro-ativados aproximadamente dentro do mesmo intervalo de tempo, assim o que ocorreu junto (ou temporalmente próximo) na percepção possa ser reinstaurado também junto na recordação. Antes de entrar em mais detalhes sobre os mecanismos dessa retro-ativação, o modelo proposto postula dois ambientes diferentes no encéfalo, mas amplamente interconectados: mapas/imagens e disposições (DAMASIO, 1989, 2010).

O ambiente das imagens é localizado nas áreas produtoras de mapas: o conjunto de córtices sensoriomotores primários, além do núcleo do trato solitário, do núcleo parabraquial e do colículo superior. Nessas estruturas ocorrem as imagens sensoriais de todos os tipos, incluindo as imagens

que se tornam conscientes e aquelas que permanecem inconscientes. Os mapas podem ser registrados em formato de disposições e, por sua vez, essas mesmas disposições podem ser usadas para reconstruir os mapas nos córtices sensoriais primários, no formato da experiência original, para que assim tomem forma de imagens mentais. Os mapas topograficamente organizados presentes nos córtices sensoriais primários constituem o único formato que a informação pode assumir para tornar-se parte da nossa mente. Notavelmente, o mecanismo gerenciado pelas disposições apresentam como grande vantagem a economia de espaço, de forma a maximizar a capacidade de armazenamento de informações no encéfalo. Mais especificamente, o ambiente das disposições é constituído pelo conjunto de córtices de associação nos lobos temporais, parietais e frontais, parte dos córtices límbicos, além das estruturas subencefálicas do prosencéfalo basal, gânglios da base, tálamo, hipotálamo e tronco encefálico. Quando os circuitos das disposições são ativados nessas regiões, eles sinalizam aos circuitos ligados a geração de imagens (MEYER; DAMASIO, 2009; DAMASIO, 2010).

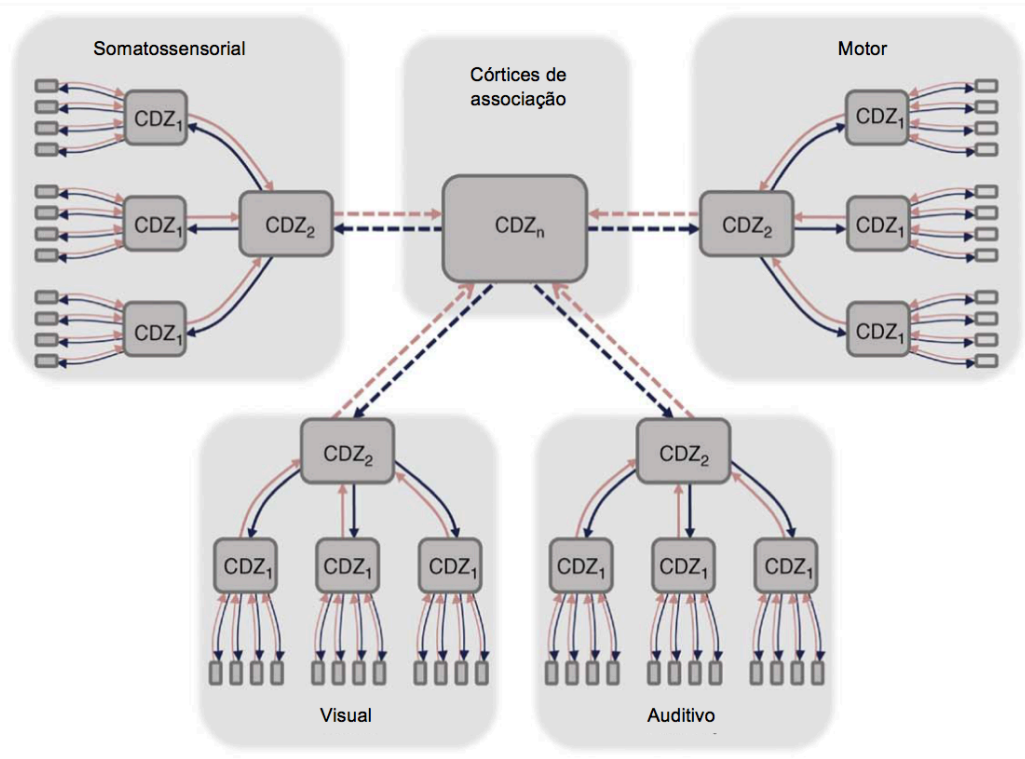
O conteúdo armazenado no ambiente das imagens é explícito, enquanto que os conteúdos do espaço das disposições são implícitos, ou seja, eles existem em forma encriptada. Pois é, Freud, a base mais primordial do conhecimento é enfim inconsciente (mas num sentido diferente). Como reflexo da natureza implícita das disposições, não é possível que se tome consciência sobre o conhecimento necessário para realizar uma determinada tarefa, por exemplo, nem sobre os passos intermediários a serem executados, apenas é possível ter consciência dos resultados finais (geralmente expressos através de alguma forma de movimento). Todas as memórias de um indivíduo, sejam elas provenientes da herança genética ou adquiridas pelo aprendizado durante a vida, existem na forma de disposições que, uma vez ativadas, podem tomar forma de imagens ou ações explícitas.

O processo da percepção sensorial é tradicionalmente compreendido por uma cascata de processamento em que os córtices sensoriais primários se projetam para córtices associativos multimodais numa forma gradual. Esse processamento culminaria nas porções anteriores dos lobos frontais e temporais, onde as imagens mentais integradas da realidade corrente seriam geradas. A diferença mais crucial entre a perspectiva tradicional e a perspectiva das CDZs é que nessa última o conhecimento não seria de alguma forma transferido das áreas primárias para as áreas de ordem superior do córtex. Os fragmentos de conhecimento mantidos nos córtices sensoriomotores primários não precisam se transportados, “subindo” numa hierarquia de processamento, e não

precisam ser projetados a uma única região para que se obtenha uma percepção integrada e multissensorial. Os grupos de neurônios que compõem CDZs nos córtices associativos não guardam uma versão processada ou composta do que existe nos córtices primários. Invés disso, as CDZs retêm as ligações e correspondências entre os fragmentos de conhecimento, ou seja, elas contêm um registro do arranjo combinatório das atividades em múltiplas localizações que descrevem um objeto durante a percepção (MEYER; DAMASIO, 2009).

A recordação, ou o reconhecimento baseado numa parte do evento original, podem ser alcançados através do mecanismo de retro-ativação, onde as CDZs reinstauram uma aproximação do padrão original em diferentes áreas sensoriomotoras através de projeções divergentes dos córtices associativos. Assim, tanto durante a percepção como durante a recordação, as imagens mentais da experiência surgem a partir da atividade neuronal em áreas sensoriomotoras primárias. Neurônios em córtices de associação coordenam essa atividade, mas não fornecem por eles mesmos os correlatos neurais das imagens mentais (Figura 2). Quanto mais complexos forem os objetos e eventos, a sua representação também se torna mais complexa. Presumivelmente, a reinstauração de um maior número de fragmentos de conhecimento, capazes definir de forma compreensiva e sem ambiguidades um objeto ou evento complexo, requer a atividade de CDZs nos níveis mais altos da hierarquia de processamento de informação (MEYER; DAMASIO, 2009).

O sulco temporal superior posterior no córtex temporoparietal pode representar objetos de forma similar através das modalidades visual e auditiva, como indicado por um estudo de fMRI em humanos (MAN et al., 2012). Essa região parece conter representações neurais de objetos que são específicas ao conteúdo e invariantes para as modalidades sensoriais, possivelmente exercendo um papel importante no reconhecimento de objetos e no seu processamento num nível conceitual, configurando uma região candidata a abrigar CDZs. Experimentos subsequentes avaliaram a generalização intermodal incluindo ainda o sentido do tato, além da visão e audição (MAN et al., 2015). Foram encontradas propriedades de invariância audiovisual na junção temporo-ociptal, de invariância audiotátil no giro pós-central e no opérculo parietal, e de invariância visuotátil no giro pós-central e no giro supramarginal.



*Figura 2.* Esquema da hierarquia das CDZs. Grupos de neurônios nos córtices sensoriais primários de diferentes modalidades (representados pelos retângulos menores) enviam projeções (flechas vermelhas) para CDZs de primeira ordem (CDZ1s), as quais, por sua vez, enviam projeções divergentes (flechas azuis) de volta para os córtices primários. Padrões de conectividades convergente-divergentes interligam CDZ1s com CDZ2s, e CDZ2s com CDZns em córtices de associação de elevada ordem (através de vários passos não representados aqui; flechas tracejadas). Adaptado de Meyer & Damasio, 2009.

Ao mesmo tempo, outros estudos de fMRI têm apoiado a ideia de que córtices sensoriais primários representam a informação relevante para suas respectivas modalidades mesmo quando essa informação chega ao encéfalo através de outro sistema sensorial. No estudo de Meyer et al., 2010, a experiência subjetiva de vários sons (de animais, instrumentos musicais e objetos), através de estímulos visuais indicativos sobre esses sons, mas na ausência de estímulos auditivos propriamente ditos, pode ser associada com uma atividade específica ao conteúdo nos córtices auditivos primários. De forma similar, a visualização do toque pode ser correlacionada com a atividade no córtex somatossensorial primário (MEYER et al., 2011). Esses resultados suportam a ideia de que a atividade nos córtices primários reflete a experiência perceptivo, mais do que a estimulação sensorial isoladamente.

Quando assistimos um músico tocar um instrumento, digamos, um *triângulo*, ou então quem sabe um *pandeiro*, simultaneamente ouvimos o som e visualizamos o movimento do instrumento em questão. Conforme postulado pelo modelo das CDZs, se um movimento específico do triângulo repetidamente co-ocorre com um som específico (em termos de frequência e amplitude), as respectivas representações neurais dos dois eventos nos córtices visual e auditivo primários serão associadas através do registro da ligação entre os dois eventos sensoriais numa CDZ compartilhada. Numa situação hipotética em que só se tenha acesso ao evento visual, o padrão de atividade produzido nos córtices visuais primários iriam disparar a CDZ relacionada, que por sua vez iria retroativar, nos córtices auditivos primários, a representação do som que geralmente acompanha o movimento do triângulo, mesmo que o som não seja parte do ambiente sensorial naquele momento.

## 1.2 JUSTIFICATIVA

Os temas da integração multissensorial e as influências intermodais apresentam grande relevância no contexto experimental da psicologia cognitiva e consciência perceptual. Nesse sentido, estudos empregando estímulos sensoriais naturalísticos mais complexos são especialmente promissores ao mimetizar de forma mais fidedigna o contexto ambiental na vida real. Contudo, enquanto que já existe literatura científica relativamente ampla sobre os efeitos que a visão pode exercer sobre outros sentidos, especialmente a audição, as influências que outros sentidos podem sobre a visão ainda não são tão exploradas.

## 1.3 OBJETIVOS

### 1.3.1 Objetivo geral

O presente trabalho dedica-se a explorar a temática de integração multissensorial no contexto da psicologia experimental, focando nos efeitos intermodais do sentido da audição sobre a percepção de estímulos visuais.

### 1.3.2 Objetivos específicos

São objetivos desse trabalho o desenvolvimento e a execução de um trabalho experimental envolvendo um paradigma psicofísico audiovisual, com a redação de um artigo científico em inglês sobre esse experimento.

## 2. ARTIGO CIENTÍFICO (em inglês)

### **Congruent visual stimulation facilitates auditory detection under masking noise**

Gabriela Mueller de Melo<sup>1</sup>, Kingson Man<sup>2</sup>, Jonas Kaplan<sup>2</sup>, Gustavo Gauer<sup>1</sup>

<sup>1</sup> Cognitive Biosignals Lab, Federal University of Rio Grande do Sul, Porto Alegre, Brazil.

<sup>2</sup> Brain and Creativity Institute, University of Southern California, Los Angeles, USA.

#### **Abstract**

Our interaction with the environment on our surroundings is inherently multimodal. The ability to synergistically combine information provided by multiple senses constitutes a fundamental aspect of brain function. Using semantically complex, realistic audiovisual stimuli of two musical instruments, we developed a 2-interval forced-choice paradigm for auditory detection under white noise masking. Following an adaptive threshold procedure, we tested whether a sound may become more easily detectable in lower intensities, i.e. lower detection thresholds, when co-presented with a synchronized congruent video of the same instrument being played, in comparison to when no visual cues of the object were presented. Indeed, our results revealed that thresholds for detecting the masked auditory signals gradually decreased from conditions without any visual cue, with a temporal visual cue and with an object visual cue. Pairwise analysis revealed a significant reduction on detection thresholds of both tambourine and triangle sounds at the object cue condition in relation to the temporal cue condition of 2.27 dB and 1.84 dB, respectively. In other words, when presented to a semantically congruent, time-synchronized, visual stimulus, participants were capable to detect the auditory signal under increased proportions of noise, in comparison to when no object cue accompanied the sound. Here we provided evidence of a bias-free cross-modal effect observed when an object was visually presented synchronously to its corresponding sound, which can be plausibly explained by an early-stage sensory interaction between semantically congruent audiovisual stimuli.

Keywords: multisensory, audiovisual, noise masking, cross-modal enhancement.



## 2.1 INTRODUCTION

Most organisms currently living on Earth are found on nearly constant perceptual interaction with a vast diversity of environmental events that provide information to multiple sensory modalities. The natural environment is inherently multisensory, and the ability to use cues from multiple senses synergistically is a fundamental aspect of brain function, impacting on both behavior and perception. While it seems natural for the perceiver that information from several sensory modalities is integrated to form a coherent whole, a central question in behavioral and neurosciences concerns the mechanisms mediating this integration. Several factors have been reported to have a major impact on these integrative mechanisms, such as spatial and temporal correspondence of incoming sensory signals in different modalities, e.g. from the visual and auditory channels (Stein & Meredith, 1993).

On multisensory research, spatiotemporal constraints have been explored extensively, creating a substantial base of knowledge on how these physical relations lead to cross-modal effects, such as decreasing reaction times (Forster, Cavina-Pratesi, Aglioti, & Berlucchi, 2002; Hughes, Reuter-Lorenz, Nozawa, & Fendrich, 1994; Molholm, Ritter, Javitt, & Foxe, 2004; Schroger & Widmann, 1998), lowering thresholds for stimulus detection (Frassinetti, Bolognini, & Ladavas, 2002; Lovelace, Stein, & Wallace, 2003; Odgaard, Arieh, & Marks, 2004; Vroomen & de Gelder, 2000), enhancing visual search (Iordanescu, Guzman-Martinez, Grabowecky, & Suzuki, 2008) and improving perceptual learning (Seitz, Kim, & Shams, 2006). It is widely understood that stimuli presented through different sensory modalities provide complementary information yielding elevated performance levels, and this effect likely results from neural integrative processes instead of simple probability summation (Miller, 1982). However, while the influence of these spatiotemporal factors can be explained by the receptive-field properties of multisensory neurons in subcortical and cortical structures, more complex mechanisms of integration, like a close correspondence of content, operate on a higher level and are more difficult to accommodate in theoretical frameworks (Doehrmann & Naumer, 2008).

In nature, multisensory stimuli derived from the same event, in addition to sharing spatial and temporal concordance, share many other features that likely aid in their binding into a unitary event. Many of these features lie in the contextual or semantic domains, and are likely the product of associations that are made throughout life. On this realm, semantic aspects of multisensory

perception have been increasingly acknowledged on the past few decades and studies have systematically shown evidences that the semantic content of a multisensory stimulus plays a critical role in determining how it is processed by the nervous system (Laurienti et al, 2003; Beauchamp et al, 2004; Belardinelli et al, 2004; Taylor et al, 2006; Naghavi et al, 2007; Noppeney et al, 2007; Hein et al; 2007).

A very weak degree of “semantic” variation might already be given whenever stimulus properties other than spatial and temporal factors are varied in experimental contexts (Doehrmann & Naumer, 2008). Probably due to the advantages for stimulus control, most multisensory studies thus far have employed fairly simple stimuli with little contextual salience, such as flashes of light, geometrical shapes, pitch and sound bursts (e.g. Brunel, Carvalho, & Goldstone, 2015; Brunel, Goldstone, Vallet, Riou, & Versace, 2013; Chang, Kanai, & Seth, 2015; Conrad, Bartels, Kleiner, & Noppeney, 2010; Giard & Peronnet, 2006; Haynes & Rees, 2005; Kang & Blake, 2005; Maddox, Atilgan, Bizley, & Lee, 2015; Odgaard, Arieh, & Marks, 2004; Sekuler, Sekuler, & Lau, 1997; Stein, London, Wilkinson, & Price, 1996; Vroomen & de Gelder, 2000). However, the concept of semantics in a stronger sense possibly relates to a whole network of information regarding objects and their diverse properties and is therefore heavily connected to semantic memory (Martin, 2007; Patterson, Nestor, & Rogers, 2007).

Typical multisensory configurations involving semantic contexts are those in which a complex visual stimulus is paired with a matching auditory counterpart. Audiovisual stimuli commonly appear on the majority of experimental paradigms, since these two sensory modalities most reliably transmit meaningful content and, conversely, their integration is more likely to be influenced by semantics (Doehrmann & Naumer, 2008) – although sensory interactions may occur between virtually every modality. When studying perceptivo awareness under multisensory contexts, most experiments have been designed to investigate how a signal arising from a different sensory modality can modulate the access to visual awareness, using a variety of paradigms in which visual stimuli are suppressed from consciousness (Deroy, Chen, & Spence, 2014; Mudrik, Faivre, & Koch, 2014).

The facilitatory effect exerted by a semantically congruent sound on the perceptual awareness of a visual stimulus is well described using visual masking paradigms (Chen & Spence, 2010, 2011b), bistable figure perception (Hsiao, Chen, Spence, & Yeh, 2012), binocular rivalry (Chen, Yeh, & Spence, 2011; Conrad et al., 2010; Kang & Blake, 2005), continuous flash

suppression (Cox & Hong, 2015), motion-induced blindness (Chang et al., 2015). The other way around, however, has not received nearly as much attention. Nevertheless, if the neural representations associated with visual and auditory stimuli can interact in a shared semantic system, it is expected that congruent visual stimulation also influences auditory performance.

Previous studies approaching semantic visual influences on auditory processing have largely focused on domains closely related to language and audiovisual speech (Bernstein, Auer, & Takayanagi, 2004; Calvert et al., 1997; Grant, 2001; Grant & Seitz, 2000; Plass, Guzman-Martinez, Ortega, Grabowecky, & Suzuki, 2014; Tjan, Chao, & Bernstein, 2014; Van Wassenhove, Grant, Poeppel, & Halle, 2005; Helfer & Freyman, 2005; Schwartz, Berthommier, & Savariaux, 2004). For example, when listening to a speech in a noisy setting, watching a talker's face drastically improves speech intelligibility (J. Bernstein & Grant, 2009; Binnie, 1973). Indeed, this improvement is particularly apparent under difficult listening conditions, such as when speech is masked by spatially coincident competing speech (Helfer & Freyman, 2005). Nevertheless, these cross-modal benefits are likely to refer more specifically to speech processing.

Language itself could be regarded as a system of symbols which transmits meaning, primarily as independent from the sensory modality used for communication. Even though, this assumption does not rule out the possibility that some cortical regions involved in language processing might fulfill a similar task in the domain of audiovisual integration (for instance, the inferior frontal cortex) (Hagoort, 2005). Moreover, Chen & Spence, 2011a, 2013 showed that the time-course of the cross-modal semantic modulation of visual picture processing differs when it is induced by naturalistic sounds or spoken words. As suggested by the authors, these findings likely reflect the different processing pathways to access the relevant semantic representations by naturalistic sounds and spoken words, although not necessarily assuming they access distinct semantic systems.

On the other hand, results from the work of Eramudugolla, Henderson, & Mattingley, 2011, testing the effects of audiovisual integration on the detection of masked speech and non-speech sounds, showed that congruent visual stimuli improved detection equally for both speech and non-speech stimuli. In turn, these findings suggest that speech stimuli are not specifically advantages by audiovisual integration for detection at threshold when compared with other naturalistic sounds. Additionally, the study performed by Calvert et al., 1997 supported the hypothesis that seeing a speech influences the perception of hearing the speech at a pre-lexical stage. Their results showed

that the linguistic visual cues of watching the speaker's lips movements are sufficient to activate auditory cortex in normal hearing individuals in the absence of auditory speech sounds. Furthermore, the auditory cortical areas were not engaged when an individual viewed nonlinguistic facial movements whereas appeared to be activated by silent meaningless speech-like movements. Tjan et al., 2014 indeed showed that a task-irrelevant visual or tactile signal can makes auditory speech detection more efficient by reducing uncertainty, even though the visual and/or tactile stimuli alone could not inform the correct response. Multisensory neuronal integration and speech-specific neural processing where thus rejected as possible explanations for the enhanced auditory speech detection under noisy conditions.

To evaluate the influences of audiovisual temporal coherence alone on selective listening, free of linguistic confounds, Maddox et al., 2015 presented listeners with competing auditory streams with varying amplitude and a visual stimulus with varying radius, while manipulating their cross-modal temporal relationships. When the time course of the auditory target matched that of the visual stimulus, participants performance improved. Because the coherence involved task-irrelevant stimulus features, this improvement probably stemmed from the integration of auditory and visual streams into cross-modal objects, enabling listeners to better attend the target. Thus, in everyday conditions, where listeners can often see the source of a sound, visual cues may help listeners to select one sound source from a mixture.

Although the "speech is special" assumption is still subject of intense debate, the results taken from studies using linguistic stimuli likely leave open questions with regard to the general principles that govern audiovisual integration, and a more cautious approach would avoid making use of this particular type of stimuli when the interest is on investigating the general mechanisms of multisensory perception. At the same time, it is still desirable to make use of stimuli of increased complexity, closer to daily situations, as common audiovisual objects. In the present study, the sound of a musical instrument is rendered outside awareness by a white noise mask. The sound intensity may increase or decrease adaptively to participants' performance on a sound detection task, thus facilitating or impairing detection. Our proposal is to test whether a sound may become more easily detectable in lower intensities, i.e. in lower detection thresholds, when co-presented with a synchronized congruent video of the same instrument being played, compared to when no visual cues of the object were presented.

## 4.2 METHODOLOGY

This experiment tested whether sound signal detection can become more sensitive, i.e. signal thresholds will lower, with the concurrent presentation of a video of the object creating the sound. A 2-interval forced choice (2-IFC) paradigm was used, presenting audiovisual stimuli under auditory masking. The task procedure followed an adaptive up-down method for determining the auditory detection thresholds.

Forced-choice tasks are designed to give a qualitative measurement of how distinct or discriminable several stimuli are from one another. On 2-alternative forced-choice (2-AFC) tasks, two alternatives are given and the participants must choose one of them for each stimulus. The alternatives might be absolute descriptions or relative to some standard. The 2-interval forced-choice tasks are a special variant of the 2-AFC task in which, rather than presenting one stimulus and two alternatives, two stimuli are shown sequentially and the participants are required to choose the interval that most closely matches some criterion (Green & Swets, 1966).

Auditory masking occurs when the perception of one sound is affected by the presence of another sound. Many examples of masking can be explained in terms of the way sounds are processed in the inner ear (Oxenham & Moore, 1995). The masking sound produces a pattern of excitation in the cochlea that either swamps or suppresses the activity due to the target sound, so that the target is no longer accurately represented in the auditory nerve (Delgutte, 1990). Masking by suppression refers to the reduction in the response to one stimulus by the introduction of a second, implying that the neural activity due to the original signal is reduced by the masker to a level indistinguishable from spontaneous activity. Suppression is thought to be an inherent property of nonlinear cochlear mechanics (Oxenham & Plack, 1998).

An adaptive procedure is one in which the stimulus level on any one trial is determined by the preceding stimuli and responses. Up-down adaptive methods exhibit widespread application in psychoacoustics and form a subset of a broader class of testing procedures generally known as sequential experiments. A sequential experiment may be defined as one in which the course of the experiment is dependent on the experimental data. For up-down procedures, the data determine the choice of stimulus levels (Levitt, 1971). On this experiment, a two-down one-up rule was employed to target a 70.7% auditory detection threshold, i.e., two consecutive correct responses lead to a reduction of signal-to-noise ratio (SNR), increasing the difficulty of the task, whereas one incorrect

response leads to an increase of SNR, reducing the difficulty of the task. The threshold SNR level is estimated from an adaptive staircase, terminated after a given number of reversals. A reversal refers to points on the same stair case at which the direction of change of the SNR (i.e., increasing/decreasing) reverses. These points, particularly towards the end of the staircase, reflect the participants' uncertainty regarding the presence of the stimulus.

### 2.2.1 Participants

Eight volunteers took part in this experiment. All of the participants were naïve as to the specific purpose of the study. They were between 20 to 30 years old, and all had normal or corrected-to-normal vision and normal hearing by self report. The study has been approved by the ethic committee from the Department of Psychology, Federal University of Rio Grande do Sul (UFRGS). Participants were informed of their rights in accordance with the ethical standards and signed a consent form.

### 2.2.2 Apparatus and stimuli

The experimental sessions took place at UFRGS dependencies. The stimuli were presented on a Dell 14" laptop color screen and closed-ear Senheiser headphones. The participants sat at a viewing distance of 58 cm from the monitor inside of a 1.5m x 1.5m audiology testing booth. Participants gave their responses using the computer keyboard as instructed. The visual test stimuli consisted of 2-second movies in .mp4 format of a triangle and a tambourine being played. The auditory test stimuli consisted of 2-second tracks in .wav format of the sound of a triangle and the sound of a tambourine (target signal), in coherence to the visual stimuli, masked by gaussian white noise. The full experiment was developed on MATLAB Psychtoolbox.

### 2.2.3 Design and procedure

Participants were first familiarized with the stimuli by watching the intact audio-visual clips of each instrument during 1 minute. The main experiment consisted of adaptively estimating threshold stimulus intensity for detection of a signal embedded in noise, using the transformed up-down procedure of Levitt, 1971 to target a 70.7% threshold. A two-down one-up rule was employed. Two consecutive correct responses led to a reduction of SNR, increasing the difficulty of the task, whereas one incorrect response led to an increase of SNR, reducing the difficulty of the task. At the starting of testing, the SNR was 0 dB, i.e., signal and noise were presented at the same intensities. The step sizes during the adaptive testing were adjusted to that at the beginning, step changes following an adaptive rule were 3 dB. Then the step changes were reduced to 2 dB until the second and third reversals occurred, as specified by the adaptive rules. Changes in the signal step sizes then followed the schedule of 1 dB until the fifth reversal started; and 0.1 dB for the final two reversals. The threshold was determined by averaging the values of the stimulus intensity in SNR dB at the final nine reversal points, whereas the first three reversals were ignored.

The task consisted of a 2-IFC paradigm, where each trial contained two observation intervals: one interval contained only white noise and the other contained the auditory signal added to the noise, plus the visual presentation (see experimental conditions below). The experimental paradigm is schematized on Figure 1. Participants indicated whether the sound signal occurred in the first or second interval. The signal-containing interval was randomly selected on each trial. The amplitude of white noise was kept constant for all trials, whereas the signal amplitude was scaled to reach the target SNR in dB in the signal-containing interval.

The noise masker always began before and ended after the signal. In each interval, the noise lasted for 4 seconds and the signal lasted for 2 seconds. The onset of the auditory signal was jittered 1-2 seconds after noise onset, to prevent participants from relying on the timing and rhythmic relationships within the two stimulus observation intervals and anticipating their responses. Only the timing of the stimulus presentation was jittered, not the relationships between audio and video for audiovisual stimuli. The duration of onset jitter was randomly selected. Additionally, to avoid abrupt starting of the signal, which could provide an extra temporal cue, the sound was presented with a slow fade-in and fade-out effect, each taking 1/4 of total signal duration.

Three conditions for concurrent visual stimulation were applied:

- 1) White fixation cross presented throughout the auditory interval (O);
- 2) Color temporal cue for signal onset/offset (A). The white fixation cross became red at signal onset and returned to white at signal offset. The color cue behaved identically for both the signal + and signal - intervals in a trial;
- 3) Visual object cue (Ta/Tr). The white fixation cross was replaced by a time-synchronized video of one of the objects for the duration of the auditory signal. The video was presented at the same time for both signal + and signal - intervals in a trial.

The experiment consisted of separated blocks for each stimulus. Each block was formed by of three separated runs for determining threshold on each visual condition. The tambourine block was performed first, then participants could take a 10-minute break, before the block of the triangle started. For each participant, a full experimental session took about one hour and 30 minutes to two hours to be completed. Before starting the first testing block, the experimenter briefly explained the 2-IFC task: “Two intervals will be presented, choose the interval containing the auditory signal additionally to the noise. It starts easy but quickly gets harder. Do you best, and give your best guess if you need to.” The collected data was analyzed by performing a repeated-measures analysis of variance (ANOVA) and pairwise comparisons using the SPSS software. The repeated factors were the three conditions tested. Statistical hypothesis testing was used for correcting for multiple comparisons as Tukey’s test. Family wise significance was adjusted for 0.05 with 95% confidence interval.

## 2.3 RESULTS

The group mean thresholds across the three conditions for both instruments and eight participants are shown in Figure 2. Applying analysis of variance to the results, it was observed a significant main effect of condition for both stimuli (tambourine  $F = 33.96$ ,  $p < 0.001$ ; triangle  $F = 46.02$ ,  $p < 0.000$ ). The threshold means (in dB SNR) for the three tambourine conditions were: O = -3.50, A = -4.09 and Ta = -6.37. For the three triangle conditions, the threshold means were: O = -8.00, A = -9.16 and Tr = -11.00. Figure 3 shows the mean results for each participant and condition,



indicating that the group pattern of highest thresholds in the O condition, lowest in the A condition and intermediate in the Ta/Tr conditions held largely constant across participants.

Pairwise analyses were employed to test the specific hypothesis of the study. For the tambourine stimuli, the comparison between O and A condition revealed a decrease of -0.58 SNR dB on thresholds ( $p = 0.35$ , non significant), between conditions A and Ta the threshold differed by -2.27 SNR dB ( $p = 0.008$ ) and between O and A conditions this difference was of -2.86 SNR dB ( $p = 0.002$ ). A threshold was significantly lower than the thresholds in the two other conditions, and no significant difference was found between O and Ta conditions (Table 1). Analyzing conditions of the triangle stimuli, there was a threshold decrease of -1.16 SNR dB between O and A conditions ( $p = 0.031$ ), between A and Tr conditions there was a decrease of -1.84 SNR dB ( $p = 0.011$ ) and for O and A conditions this difference was of -3.0 SNR dB ( $p = 0.001$ ). Tr threshold was significantly lower when compared to the thresholds in the two other conditions, and A threshold was also significantly lower than the O threshold (Table 1).

## 2.4 DISCUSSION

A general evaluation of the data reveals that the threshold for detecting the masked auditory signal gradually decreased from no visual cue (white fixation cross), temporal visual cue (changing white/red fixation cross) and object visual cue conditions. The absolute values for thresholds varied between instruments, but some variation is expected, given the two stimuli presented specific physical properties, and different amplitude envelope and spectral acoustic cues are known as to affect the audiovisual perception (Chuen & Schutz, 2016). More specifically, pairwise analysis revealed a significant reduction on detection thresholds of both tambourine and triangle sounds at the object cue condition in relation to the temporal cue condition, evidencing a facilitation effect exerted by a congruent object visually presented on auditory detection performance. For the tambourine stimulus, the object cue improved detection threshold on 2.27 dB, whereas for the triangle this effect was of 1.84 dB. In other words, when presented to a semantically congruent, time-synchronized, visual stimulus, participants were able to detect the auditory signal under increased proportions of noise, in comparison to when no object cue accompanied the sound.

Making use of visible speech cues for testing auditory detection of spoken sentences under noise, the work of Grant & Seitz, 2000 revealed an improvement on detection thresholds of about 1.6 dB when the speechread sentence matched the target sentence (which would be analogous to our object cue condition) in relation to the unimodal auditory condition (which would be analogous to our no visual cue condition), an effect which is lower but coherent to the observed in our experiment. Furthermore, on their study, analyses of correlation between area of mouth opening and acoustic envelopes derived from selected spectral regions of each sentence suggested that the observed threshold reduction may be determined by the degree of auditory-visual temporal coherence, especially between the area of lip opening and the envelope derived from mid-frequency to high-frequency acoustic energy. Therefore, it was proposed that visual cues derived from the dynamic movements during speech production interacted with time-aligned auditory cues to enhance sensitivity in auditory detection, and that the amount of visual influence depended in part on the degree of correlation between acoustic envelopes and visible movement of the speakers. Indeed, a sequential study (Grant, 2001) manipulating the cross-modal correlation through bandpass filtering, suggested that the magnitude of a effect played by vision on speech detection under noise depended on both the spectral and temporal properties of the target speech signal.

Grant's experiments showed that seeing the lips of a speaker also enhances sensitivity to acoustic information, decreasing the auditory detection threshold of speech embedded in noise. However, detection is different from comprehension. Schwartz et al., 2004 therefore designed an experiment to test whether improved sensitivity also results in an intelligibility gain in audio-visual speech perception. Their audio-visual stimuli could not be differentiated by lip reading per se since they contained exactly the same lip gesture matched with different compatible speech sounds. Thus, if visual information improves intelligibility, this improvement is not likely to be due to visual information per se, since the stimuli are visually similar. Rather, the improvement should be mainly due to visually enhanced detection of acoustic cues. Results indeed showed that the noise-masked stimuli were more intelligible in the audio-visual condition than in the audio-only condition due to the contribution of temporal visual information to the extraction of acoustic cues. This mechanism, different from lip-reading, provides an early stage for audio-visual interaction in speech identification. Additionally, the speech-specific nature of this effect was suggested when replacing the lip gesture by a non-speech visual input with exactly the same time course, providing the same temporal cues for extraction, removed the intelligibility benefit.

Sensory processes relying on early-stage interactions are far less vulnerable to response biases than are decisional processes. In turn, response biases are highly susceptible to manipulation of presentation probability (Odgaard et al., 2004). For example, if the magnitude of an effect on auditory detection increases with an increased probability of visual stimulation, the observed effect would likely result from a biased decisional process. On our 2-IFC paradigm, we were able to eliminate the possibility that the observed size of the facilitatory effect depended on the probability that the visual stimulus of the object accompanied the sound, once the sound could equally be presented either on the first or second interval, both containing exactly the same visual stimulation. Additionally, always that a visual stimulus provided a cue about the outset of the auditory signal (either on the temporal or object conditions), participants performed better in comparison to no-visual cue conditions, but even though both stimuli provided the same visual cue about the starting time of the sound, participants performed better under visual object stimulation. Moreover, our results provide evidence of a facilitatory effect observed when the object was presented together to the sound which cannot be explained by a decisional bias. Rather, the effect exerted by the presentation of the semantically congruent object cue itself is more consistent with a true, early-stage, sensory interaction.

At least in some situations and particular experimental configurations, speech-specificity explanations may be extrapolated to a mechanism underlying the multisensory processing of higher-complexity semantic stimuli in general. Mouth movements during speech may interact with time-aligned auditory cues so as the movements of instruments being played may interact to correspondent time-aligned auditory cues, depending on the correlation between audiovisual physical properties of the stimuli. Furthermore, as non-speech visual input providing the same temporal cues than visual lips removed the cross-modal benefit on Schwartz et al., 2004, in our case a simple visual input providing the same temporal cues about the outset of the sound than a complex object also led to a smaller cross-modal effect (indeed non significant from the no-visual cue condition, for one instrument). Temporal cues are important, but are not alone sufficient to fully explain a cross-modal enhancement on behavioral performance of semantically complex, realistic stimuli. Rather, timing may play a role more closely related to the coordination of multisensory integration.

A hypothetical explanation underlying the facilitatory effect observed on our experiment, and likely cross-modal effects in general, could be depicted in terms of the following mechanism:

initially, the auditory noisy input would converge through a processing hierarchy, parting from the cochlea to specific subcortical nuclei, to auditory cortices and then to multisensory association areas. Reaching these high-order association regions, the sensory signal would match a stored multisensory semantic representation of the object, possibly organized in the form of neural dispositions (i.e. convergence-divergence zones). In addition to the auditory stimulus, the co-presentation of a time-synchronized visual stimulus could also activate, at the correspondent timing, the stored multisensory representation of the object in association areas. The reinforced multisensory representation would then send enhanced divergent signals back to primary unisensorial cortices, straightening the primary auditory correlates of the instrument. What would be these correlates? The sensory maps, which only can exist in the primary cortices and are the unique form by which any stimulus can be consciously perceived (Damasio, 2010; Meyer & Damasio, 2009). The activation of the auditory maps relative to a specific instrument would enhance the awareness about its presence in the environment, thus facilitating auditory detection even under increased noise levels. Otherwise, a relatively larger intensity of unisensory stimulation would be necessary (i.e. higher signal-to-noise ratio) to exert the same effects on detection as in audiovisual conditions.

A study performed by Meyer et al., 2010, making use of multivariate pattern analysis (MVPA) of functional magnetic resonance imaging data in human subjects, supported the idea that early sensory cortex activity reflects perceptual experience, rather than sensory stimulation alone. On this study, it was possible to predict which visual stimuli participants were seeing solely on the basis of activity in auditory cortices. As subjects viewed sound-implying, but silent, visual stimuli, activity in auditory cortex differentiated among sounds related to various animals, musical instruments and objects. Thus, the subjective experience of sound in the absence of auditory stimulation, i.e. hearing sounds in the “mind’s ear”, was associated with content-specific activity in very early auditory cortices in humans. The degree of specificity of the auditory representations is suggested by the successful discrimination not only among categories of video clips, but among individual clips as well, further indicating that the representations of sounds in early auditory cortices respect categorical boundaries.

Sensory processes relying on early-stage interactions are far less vulnerable to response biases than are decisional processes. In turn, response biases are highly susceptible to manipulation of presentation probability (Odgaard et al., 2004). For example, if the magnitude of an effect on

auditory detection increases with an increased probability of visual stimulation, the observed effect would likely result from a biased decisional process. On our 2-IFC paradigm, we were able to eliminate the possibility that the observed size of the facilitatory effect depended on the probability that the visual stimulus of the object accompanied the sound, once the sound could equally be presented either on the first or second interval, both containing exactly the same visual stimulation. Additionally, always that a visual stimulus provided a cue about the outset of the auditory signal (either on the temporal or object conditions), participants performed better in comparison to no-visual cue conditions, but even though both stimuli provided the same visual cue about the starting time of the sound, participants performed better under visual object stimulation.

## 2.5 CONCLUSIONS

Here we were able to provide evidence of a bias-free cross-modal effect observed when an object was visually presented synchronously to its corresponding sound, which can be plausibly explained by an early-stage sensory interaction between semantically congruent audiovisual stimuli. However, to further explore the veracity of the hypothetical mechanism underlying this interaction as proposed, more evidence is still required. Multisensory research often makes use of congruence/incongruence distinctions between stimuli to corroborate semantic assumptions. On that real, it would be highly propitious to test auditory detection performance as varying object audiovisual congruence on different experimental conditions. Furthermore, to more closely investigate the engagement of multisensory representations and underlying neural architecture on the recognition of specific objects, performance on identification tasks could be measured. The additional analysis of metacognitive feelings, estimating to what extend participants could unambiguously access their perceptual representations by asking how confident they felt about their own perceptual reports, is a desirable and promising strategy to evaluate the relation with the awareness as manifested in multisensory situations.

## 2.6 APPENDIX – WHAT’S NEXT?

It has been shown that correspondent visual information can aid auditory detection of an object, by lowering detection thresholds on a 2-IFC detection paradigm under auditory masking. In order to investigate more closely whether this cross-modal effect relate to enhancing the neural correlates of the object represented in the brain, a crucial remaining question asks the role of semantic congruency for the effect to happen. From previous results, the possibility that participants would rely on temporal cues provided by any concurrent visual stimulation has been shown unlikely to account for the perceptual improvement observed on multimodal conditions. However, it could be additionally argued that the presentation of any visual stimuli belonging to the same class of objects (i.e. musical instruments) may have enhanced auditory detection, regardless of its specific semantic content. Thus, in order to test this assumption, here we compare the magnitude of the cross-modal effect among congruent and incongruent conditions. On congruent conditions, either an audiovisual tambourine or triangle were presented, whereas on incongruent conditions, the auditory tambourine was presented with the visual triangle, and vice-versa.

A further hypothetical assumption is that an auditory stimulus rendered unconscious by noise masking can have its neural representation reinforced by a congruent stimulation through another modality, thus exerting a facilitatory effect for the perceptual awareness of the object. In other words, the cross-modal stimulation could bring the object representation back to awareness more efficiently, easing its arousal on consciousness. However, the conceptual relationship between multisensory integration and perception awareness is still unclear, and posits methodological challenges to the most common experimental approaches on multisensory research (Deroy et al., 2014; Noppeney et al., 2016). Event though a fundamental aspect of our perceptual experience is that multisensory signals are integrated into a coherent percept, and despite leading theories largely support a strong link between information integration and perceptual awareness (Balduzzi & Tononi, 2008; Dehaene & Naccache, 2001; Lamme & Roelfsema, 2000), a remaining conceptual and methodological challenge is to clarify the relationship between multisensory integration and perception awareness.

Metacognition refers to cognitive processes about other cognitive processes. A key aspect of metacognition is the ability to recognize one’s own successful cognitive processing, i.e. to

monitor one's decisions and representations, their accuracy and uncertainty. Experimental approaches on metacognition ask to which extent observers can assess the uncertainty or accuracy of their perceptual representations and judgments, thus recognizing their own performance. Metacognitive accuracy can be measured based on the fidelity of subjects' trial-by-trial confidence judgments with respect to their objective task performance (Deroy, Spence, & Noppeney, 2016; Fleming & Lau, 2014). However, most studies to date had focused on simple visual or auditory tasks that were based on a single signal stream and little is known about metacognition in a multisensory environment. On multisensory research, the analysis of metacognitive feelings has been suggested as a useful implement to investigate the relation with the awareness that manifests itself in multisensory situations (Deroy et al., 2014).

Multisensory perception relies on inferring the world's causal structure, integrating signals from a common cause but segregating those from independent causes. Stimuli in the real world thereby produce first-order multisensory representations of objects and events. In turn, metacognition relates to the formation of second-order representations, which allow observers to monitor their first-order representations. Thereby, observers can assess their uncertainty about the perceptual representations that are formed on the basis of noisy sensory evidence. In other words, metacognition can be described as the ability to introspect about the perceptual inference processes that led to the first-order perceptual representations about the environment (Deroy et al., 2016).

Humans can indeed access the uncertainty of their perceptual representations and adjust their decisional confidence accordingly, as showed by unimodal research (Aitchison, Bang, Bahrami, & Latham, 2015; Clifford, Arabzadeh, & Harris, 2008; Dienes & Seth, 2010; Massoni, Gajdos, & Vergnaud, 2014; Persaud, McLeod, & Cowey, 2007). An observer's decisional confidence level reflects the probability that his/her perceptual choices are correct, given the sensory signals available at the time. Furthermore, one's decisional confidence depends on the uncertainty of the related first-order perceptual representations. In other words, the perceptual sensitivity (e.g., the ability to detect the sound of triangle) constrains the maximally possible metacognitive sensitivity (i.e., the ability to discriminate between their correct and incorrect choices) of an observer (Fleming & Lau, 2014; Maniscalco & Lau, 2012). Thus, in order to measure the metacognitive accuracy of participants when detecting and identifying real-world multisensory stimuli, we developed a new psychophysical experimental paradigm. The experiment consists on forced-choice perceptual tasks accompanied by confidence ratings on a trial-by-trial basis.

Thereby, we aim to evaluate whether observers are appropriately confident about the accuracy of their perceptual choices and how their confidence correlate with their accuracy on different stimuli settings.

Furthermore, in order to investigate more closely whether a cross-modal effect would relate to enhancing the neural correlates of the object represented in the brain, a crucial remaining question asks the role of semantic congruency for the effect to happen. From previous results, the possibility that participants would rely on temporal cues provided by any concurrent visual stimulation has been shown unlikely to account for the perceptual improvement observed on multimodal conditions. However, it could be additionally argued that the presentation of any visual stimuli belonging to the same class of objects (i.e. musical instruments) may have enhanced auditory detection, regardless of its specific semantic content. Thus, to evaluate this assumption, we compare the magnitude of the cross-modal effect among congruent and incongruent conditions. On congruent conditions, either an audiovisual tambourine or triangle were presented, whereas on incongruent conditions, the auditory tambourine was presented with the visual triangle, and vice-versa. Both congruent and incongruent settings are tested on detection and identification tasks in association to the evaluation of metacognitive accuracy. The detailed methodology is presented next.

### 2.6.1 A psychophysical evaluation of the influence of vision on auditory detection and identification of objects

The experiment proposed here consists of signal detection and signal identification forced-choice tasks. The apparatus and stimuli are the same to the used on the threshold experiment. For taking part on the present experiment, participants must have their auditory thresholds previously determined.

The detection task consists of a 2-interval forced-choice (2-IFC) design and tests whether the co-presentation of a visual stimulus affects the detection of an auditory signal embedded in noise. The experimental paradigm is schematized on Figure A1. The task is performed separately for each object. The 2-IFC paradigm contains auditory-only trials and audio-visual trials. The previously determined 70.7% auditory threshold (visual condition 2) is inputted for both instruments. Each trial consists of two noise clips played one after the other, with a short gap in



between them. One of the clips contains an auditory signal embedded, while the other contains pure white noise. Half of the trials are presented accompanied by movies. Four conditions are tested:

- 1) Auditory-only containing the signal in the first interval;
- 2) Auditory-only containing the signal in the second interval;
- 3) Audio-visual containing the signal in the first interval;
- 4) Audio-visual containing the signal in the second interval.

All conditions are presented the same number of times, and the trial sequence is shuffled across all conditions. After each trial, subjects are asked to report which clip contained the signal (1 or 2), and to estimate their confidence in their response on a 1-5 scale.

The identification task consists of a 2-alternative forced choice (2-AFC) design and tests whether a visual input may affect auditory identification. The paradigm is schematized on Figure A2. Again, the previously determined 70.7% detection thresholds were used. The auditory signals are presented in intermixed trials with co-presentation of visual input. The visually presented objects are either congruent with, incongruent with, or neutral to the auditory presented objects. Nine conditions are tested:

- 1) Noise-only;
- 2) Auditory-only triangle;
- 3) Auditory-only tambourine;
- 4) Visual-only triangle;
- 5) Audio-visual triangle (congruent);
- 6) Audio-visual triangle-tambourine (incongruent);
- 7) Visual-only tambourine;
- 8) Audio-visual tambourine (congruent);
- 9) Audio-visual tambourine-triangle (incongruent).

All conditions are presented the same number of times. Additionally, 20% of the trials are catch trials with no auditory signal. The sequence of trials is randomized. After each trial, subjects are asked to report which sound was present (triangle or tambourine), and then to report their confidence in their response on a 1-5 scale.

## REFERENCES

- Aitchison, L., Bang, D., Bahrami, B., & Latham, P. E. (2015). Doubly Bayesian Analysis of Confidence in Perceptivo Decision-Making. *PLoS Computational Biology*, *11*(10), e1004519. Journal Article, Research Support, Non-U.S. Gov't. <http://doi.org/10.1371/journal.pcbi.1004519>
- Balduzzi, D., & Tononi, G. (2008). Integrated information in discrete dynamical systems: motivation and theoretical framework. *PLoS Computational Biology*, *4*(6), e1000091. Journal Article, Research Support, N.I.H., Extramural, Research Support, Non-U.S. Gov't. <http://doi.org/10.1371/journal.pcbi.1000091>
- Bernstein, J., & Grant, K. (2009). Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *125*(5), 3358–3372. Journal Article, Research Support, Non-U.S. Gov't. <http://doi.org/10.1121/1.3110132>
- Bernstein, L. E., Auer, E. T., & Takayanagi, S. (2004). Auditory speech detection in noise enhanced by lipreading. *Speech Communication*, *44*(1–4 SPEC. ISS.), 5–18. <http://doi.org/10.1016/j.specom.2004.10.011>
- Binnie, C. (1973). Bi-sensory articulation functions for normal hearing and sensorineural hearing loss patients. *Journal of the Academy of Rehabilitative Audiology*, *6*, 43–53.
- Brunel, L., Carvalho, P. F., & Goldstone, R. L. (2015). It does belong together: Cross-modal correspondences influence cross-modal integration during perceptivo learning. *Frontiers in Psychology*, *6*(MAR), 1–10. <http://doi.org/10.3389/fpsyg.2015.00358>
- Brunel, L., Goldstone, R. L., Vallet, G., Riou, B., & Versace, R. (2013). When seeing a dog activates the bark multisensory generalization and distinctiveness effects. *Experimental Psychology*, *60*(2), 100–112. <http://doi.org/10.1027/1618-3169/a000176>
- Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., ... David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science (New York, N.Y.)*, *276*(5312), 593–6. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9110978>
- Chang, A. Y. C., Kanai, R., & Seth, A. K. (2015). Cross-modal prediction changes the timing of conscious access during the motion-induced blindness. *Consciousness and Cognition*, *31*(May 2016), 139–147. <http://doi.org/10.1016/j.concog.2014.11.005>
- Chen, Y.-C., & Spence, C. (2010). When hearing the bark helps to identify the dog: semantically-congruent sounds modulate the identification of masked pictures. *Cognition*, *114*(3), 389–404. <http://doi.org/10.1016/j.cognition.2009.10.012>
- Chen, Y.-C., & Spence, C. (2011a). Crossmodal semantic priming by naturalistic sounds and spoken words enhances visual sensitivity. *Journal of Experimental Psychology. Human*

- Perception and Performance*, 37(5), 1554–68. <http://doi.org/10.1037/a0024329>
- Chen, Y.-C., & Spence, C. (2011b). The crossmodal facilitation of visual object representations by sound: evidence from the backward masking paradigm. *Journal of Experimental Psychology. Human Perception and Performance*, 37(6), 1784–802. <http://doi.org/10.1037/a0025638>
- Chen, Y.-C., & Spence, C. (2013). The time-course of the cross-modal semantic modulation of visual picture processing by naturalistic sounds and spoken words. *Multisensory Research*, 26(4), 371–86. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/24319929>
- Chen, Y.-C., Yeh, S.-L., & Spence, C. (2011). Crossmodal constraints on human perceptivo awareness: auditory semantic modulation of binocular rivalry. *Frontiers in Psychology*, 2, 212. <http://doi.org/10.3389/fpsyg.2011.00212>
- Chuen, L., & Schutz, M. (2016). The unity assumption facilitates cross-modal binding of musical, non-speech stimuli: The role of spectral and amplitude envelope cues. *Attention, Perception & Psychophysics*, 78, 1512–1528. <http://doi.org/10.3758/s13414-016-1088-5>
- Clifford, C. W. G., Arabzadeh, E., & Harris, J. A. (2008). Getting technical about awareness. *Trends in Cognitive Sciences*, 12(2), 54–58. Journal Article, Research Support, Non-U.S. Gov't. <http://doi.org/10.1016/j.tics.2007.11.009>
- Conrad, V., Bartels, A., Kleiner, M., & Noppeney, U. (2010). Audiovisual interactions in binocular rivalry. *Journal of Vision*, 10(10), 27. <http://doi.org/10.1167/10.10.27>
- Cox, D., & Hong, S. W. (2015). Semantic-based crossmodal processing during visual suppression. *Frontiers in Psychology*, 6, 722. <http://doi.org/10.3389/fpsyg.2015.00722>
- Damasio, A. (2010). *Self Comes to Mind*. Vintage Books.
- Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, 79(1–2), 1–37. [http://doi.org/10.1016/S0010-0277\(00\)00123-2](http://doi.org/10.1016/S0010-0277(00)00123-2)
- Delgutte, B. (1990). Physiological mechanisms of psychophysical masking: observations from auditory-nerve fibers. *The Journal of the Acoustical Society of America*, 87(2), 791–809. Journal Article, Research Support, U.S. Gov't, P.H.S.
- Deroy, O., Chen, Y.-C., & Spence, C. (2014). Multisensory constraints on awareness. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369(1641), 20130207. <http://doi.org/10.1098/rstb.2013.0207>
- Deroy, O., Spence, C., & Noppeney, U. (2016). Metacognition in multisensory perception. *Trends in Cognitive Sciences*, xx, 1–12. <http://doi.org/10.1016/j.tics.2016.08.006>
- Dienes, Z., & Seth, A. (2010). Gambling on the unconscious: a comparison of wagering and confidence ratings as measures of awareness in an artificial grammar task. *Consciousness*

- and Cognition*, 19(2), 674–681. Comparative Study, Journal Article.  
<http://doi.org/10.1016/j.concog.2009.09.009>
- Doehrmann, O., & Naumer, M. J. (2008). Semantics and the multisensory brain: How meaning modulates processes of audio-visual integration. *Brain Research*, 1242, 136–150.  
<http://doi.org/10.1016/j.brainres.2008.03.071>
- Eramudugolla, R., Henderson, R., & Mattingley, J. B. (2011). Effects of audio-visual integration on the detection of masked speech and non-speech sounds. *Brain and Cognition*, 75(1), 60–66. <http://doi.org/10.1016/j.bandc.2010.09.005>
- Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, 8(July), 443. <http://doi.org/10.3389/fnhum.2014.00443>
- Forster, B., Cavina-Pratesi, C., Aglioti, S. M., & Berlucchi, G. (2002). Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time. *Experimental Brain Research*, 143(4), 480–487. Journal Article, Research Support, Non-U.S. Gov't. <http://doi.org/10.1007/s00221-002-1017-9>
- Frassinetti, F., Bolognini, N., & Ladavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research*, 147(3), 332–343. Comparative Study, Journal Article, Research Support, Non-U.S. Gov't.  
<http://doi.org/10.1007/s00221-002-1262-y>
- Giard, M. H., & Peronnet, F. (2006). Auditory-Visual Integration during Multimodal Object Recognition in Humans: A Behavioral and Electrophysiological Study. *Dx.Doi.Org*, 11(5), 473–490. <http://doi.org/10.1162/089892999563544>
- Grant, K. W. (2001). The effect of speechreading on masked detection thresholds for filtered speech. *The Journal of the Acoustical Society of America*, 109(5 Pt 1), 2272–2275.  
<http://doi.org/10.1121/1.1362687>
- Grant, K. W., & Seitz, P. F. P. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *The Journal of the Acoustical Society of America*, 108(3 Pt 1), 1197–1208. <http://doi.org/10.1121/1.422512>
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, US: Wiley.
- Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences*, 9(9), 416–423. Journal Article, Review. <http://doi.org/10.1016/j.tics.2005.07.004>
- Haynes, J.-D., & Rees, G. (2005). Predicting the stream of consciousness from activity in human visual cortex. *Current Biology : CB*, 15(14), 1301–7.  
<http://doi.org/10.1016/j.cub.2005.06.026>
- Helfer, K. S., & Freyman, R. L. (2005). The role of visual speech cues in reducing energetic and informational masking. *The Journal of the Acoustical Society of America*, 117(2), 842–849.

Journal Article, Research Support, N.I.H., Extramural, Research Support, U.S. Gov't, P.H.S.

- Hsiao, J.-Y., Chen, Y.-C., Spence, C., & Yeh, S.-L. (2012). Assessing the effects of audiovisual semantic congruency on the perception of a bistable figure. *Consciousness and Cognition*, 21(2), 775–87. <http://doi.org/10.1016/j.concog.2012.02.001>
- Hughes, H. C., Reuter-Lorenz, P. A., Nozawa, G., & Fendrich, R. (1994). Visual-auditory interactions in sensorimotor processing: saccades versus manual responses. *Journal of Experimental Psychology. Human Perception and Performance*, 20(1), 131–153. Comparative Study, Journal Article, Research Support, U.S. Gov't, Non-P.H.S.
- Iordanescu, L., Guzman-Martinez, E., Grabowecky, M., & Suzuki, S. (2008). Characteristic sounds facilitate visual search. *Psychonomic Bulletin & Review*, 15(3), 548–54. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2647585&tool=pmcentrez&rendertype=abstract>
- Kang, M., & Blake, R. (2005). Perceptivo synergy between seeing and hearing revealed during binocular rivalry. *Psychologija*, 32, 7–15. <http://doi.org/10.1167/9.1.17>
- Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, 23(11), 571–579. [http://doi.org/10.1016/S0166-2236\(00\)01657-X](http://doi.org/10.1016/S0166-2236(00)01657-X)
- Levitt, H. (1971). Transformed Up-Down Methods in Psychoacoustics. *The Journal of the Acoustical Society of America*, 467–477. <http://doi.org/10.1121/1.1912375>
- Lovelace, C. T., Stein, B. E., & Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. *Brain Research. Cognitive Brain Research*, 17(2), 447–453. Comparative Study, Journal Article, Research Support, Non-U.S. Gov't, Research Support, U.S. Gov't, P.H.S.
- Maddox, R. K., Atilgan, H., Bizley, J. K., & Lee, A. K. (2015). Auditory selective attention is enhanced by a task-irrelevant temporally coherent visual stimulus in human listeners. *eLife*, 2015(4), 1–11. <http://doi.org/10.7554/eLife.04995.001>
- Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, 21(1), 422–430. <http://doi.org/10.1016/j.concog.2011.09.021>
- Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, 58, 25–45. Journal Article, Research Support, N.I.H., Intramural. <http://doi.org/10.1146/annurev.psych.57.102904.190143>
- Massoni, S., Gajdos, T., & Vergnaud, J. C. (2014). Confidence measurement in the light of signal detection theory. *Frontiers in Psychology*, 5(DEC), 1–13. <http://doi.org/10.3389/fpsyg.2014.01455>

- Meyer, K., & Damasio, A. (2009). Convergence and divergence in a neural architecture for recognition and memory. *Trends in Neurosciences*, 32(7), 376–82. <http://doi.org/10.1016/j.tins.2009.04.002>
- Meyer, K., Kaplan, J. T., Essex, R., Webber, C., Damasio, H., & Damasio, A. (2010). Predicting visual stimuli on the basis of activity in auditory cortices. *Nature Neuroscience*, 13(6), 667–8. <http://doi.org/10.1038/nn.2533>
- Miller, J. (1982). Divided attention: evidence for coactivation with redundant signals. *Cognitive Psychology*, 14(2), 247–279. Journal Article, Research Support, U.S. Gov't, Non-P.H.S.
- Molholm, S., Ritter, W., Javitt, D. C., & Foxe, J. J. (2004). Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study. *Cerebral Cortex (New York, N.Y. : 1991)*, 14(4), 452–465. Journal Article, Research Support, Non-U.S. Gov't, Research Support, U.S. Gov't, P.H.S.
- Mudrik, L., Faivre, N., & Koch, C. (2014). Information integration without awareness. *Trends in Cognitive Sciences*, 18(9), 488–496. <http://doi.org/10.1016/j.tics.2014.04.009>
- Noppeney, U., Faivre, N., Lunghi, C., Deroy, O., Spence, C., & Allier, M. (2016). The Complex Interplay Between Multisensory Integration and Perceptivo Awareness. *Multisensory Research*, 1–22. <http://doi.org/10.1163/22134808-00002529>
- Odgaard, E. C., Arieh, Y., & Marks, L. E. (2004). Brighter noise: sensory enhancement of perceived loudness by concurrent visual stimulation. *Cognitive, Affective & Behavioral Neuroscience*, 4(2), 127–132. <http://doi.org/10.3758/CABN.4.2.127>
- Oxenham, A. J., & Moore, B. C. (1995). Additivity of masking in normally hearing and hearing-impaired subjects. *The Journal of the Acoustical Society of America*, 98(4), 1921–1934. Comparative Study, Journal Article, Research Support, Non-U.S. Gov't.
- Oxenham, A. J., & Plack, C. J. (1998). Suppression and the upward spread of masking. *The Journal of the Acoustical Society of America*, 104(6), 3500–10. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9857509>
- Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews. Neuroscience*, 8(12), 976–987. Journal Article, Review. <http://doi.org/10.1038/nrn2277>
- Persaud, N., McLeod, P., & Cowey, A. (2007). Post-decision wagering objectively measures awareness. *Nature Neuroscience*, 10(2), 257–261. Journal Article, Research Support, Non-U.S. Gov't. <http://doi.org/10.1038/nn1840>
- Plass, J., Guzman-Martinez, E., Ortega, L., Grabowecky, M., & Suzuki, S. (2014). Lip reading without awareness. *Psychological Science*, 25(9), 1835–7. <http://doi.org/10.1177/0956797614542132>
- Schroger, E., & Widmann, A. (1998). Speeded responses to audiovisual signal changes result

from bimodal integration. *Psychophysiology*, 35(6), 755–759. Journal Article, Research Support, Non-U.S. Gov't.

Schwartz, J. L., Berthommier, F., & Savariaux, C. (2004). Seeing to hear better: Evidence for early audio-visual interactions in speech identification. *Cognition*, 93(2), 69–78. <http://doi.org/10.1016/j.cognition.2004.01.006>

Seitz, A. R., Kim, R., & Shams, L. (2006). Sound facilitates visual learning. *Current Biology : CB*, 16(14), 1422–1427. Comparative Study, Journal Article, Research Support, N.I.H., Extramural, Research Support, U.S. Gov't, Non-P.H.S. <http://doi.org/10.1016/j.cub.2006.05.048>

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature*. <http://doi.org/10.1038/385308a0>

Stein, B., London, N., Wilkinson, L., & Price, D. (1996). Enhancement of Perceived Visual Intensity by Auditory Stimuli: A Psychophysical Analysis. *Journal of Cognitive Neuroscience*, 8(6), 497–506. <http://doi.org/10.1162/jocn.1996.8.6.497>

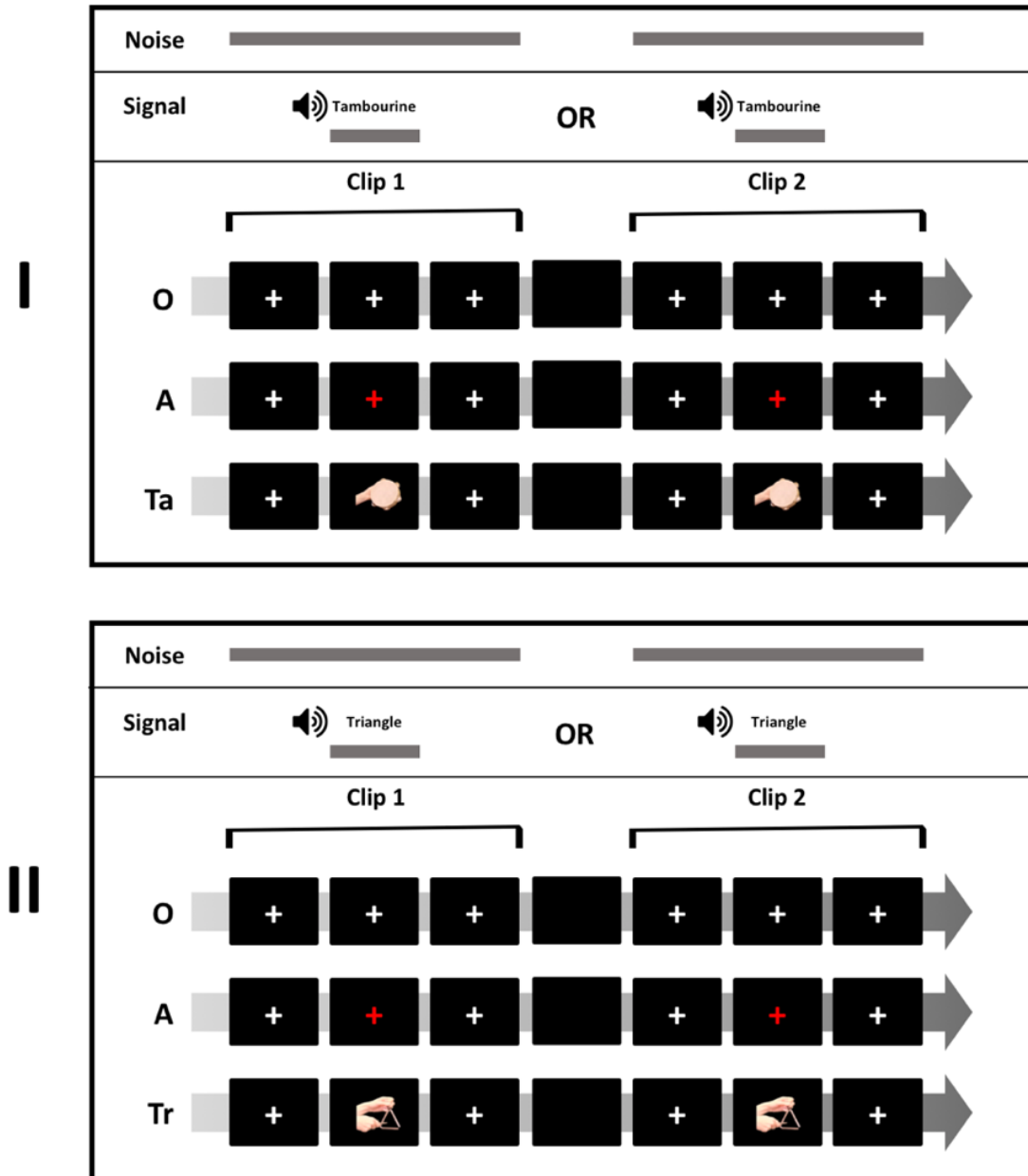
Stein, B., & Meredith, M. (1993). *The merging of the senses*. Cambridge, Mass: MIT Press.

Tjan, B. S., Chao, E., & Bernstein, L. E. (2014). A visual or tactile signal makes auditory speech detection more efficient by reducing uncertainty. *European Journal of Neuroscience*, 39(8), 1323–1331. <http://doi.org/10.1111/ejn.12471>

Van Wassenhove, V., Grant, K. W., Poeppel, D., & Halle, M. (2005). Visual Speech Speeds up the Neural Processing of Auditory Speech. *Source: Proceedings of the National Academy of Sciences of the United States of America*, 102(4), 1181–1186. <http://doi.org/10.1073/pnas.0408949102>

Vroomen, J., & de Gelder, B. (2000). Sound Enhances Visual Perception: Cross-Modal Effects of Auditory Organization on Vision. *Journal of Experimental Psychology: Human Perception and Performance*, 26(5), 1583–1590. <http://doi.org/10.1037/0096-1523.26.5.1583>

## FIGURES



*Figure 1.* Two-interval forced-choice detection task within the three audiovisual conditions for the tambourine (I) and the triangle (II) stimuli. Each trial consisted on the presentation of two audiovisual clips with a short interval between them. Each clip contained acoustic noise and a target auditory signal, presented together to a white fixation cross, a white/red temporal cue fixation cross or a synchronous congruent object. A white fixation cross was presented at the beginning and at the ending of each clip. The sequence of trials was randomly selected between the three conditions.



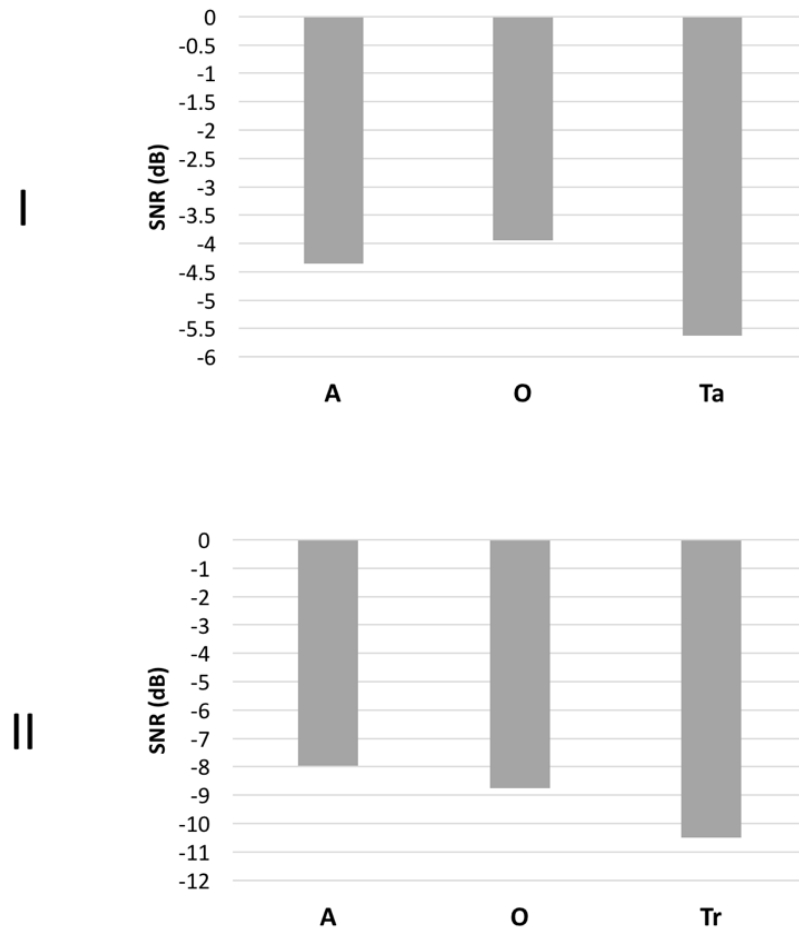


Figure 2. Group means for thresholds across all participants for the tambourine (I) and the triangle (II) conditions. Means are dB signal-to-noise ratio at the estimated 70.7% thresholds.

Table 1 Multiple pairwise comparisons test between conditions.

Pairs of conditions compared		Mean difference	95% confidence interval	Adjusted P value
Tambourine	<b>O vs A</b>	0.59	0.37 to 11.55	0.2408
	<b>A vs Ta</b>	2.27**	0.79 to 3.75	0.0067
	<b>O vs Ta</b>	2.86**	1.41 to 4.31	0.0016
Triangle	<b>O vs A</b>	1.16*	0.17 to 2.14	0.0244
	<b>A vs Tr</b>	1.84**	0.56 to 3.11	0.0092
	<b>O vs Tr</b>	3.00***	1.69 to 4.30	0.0006

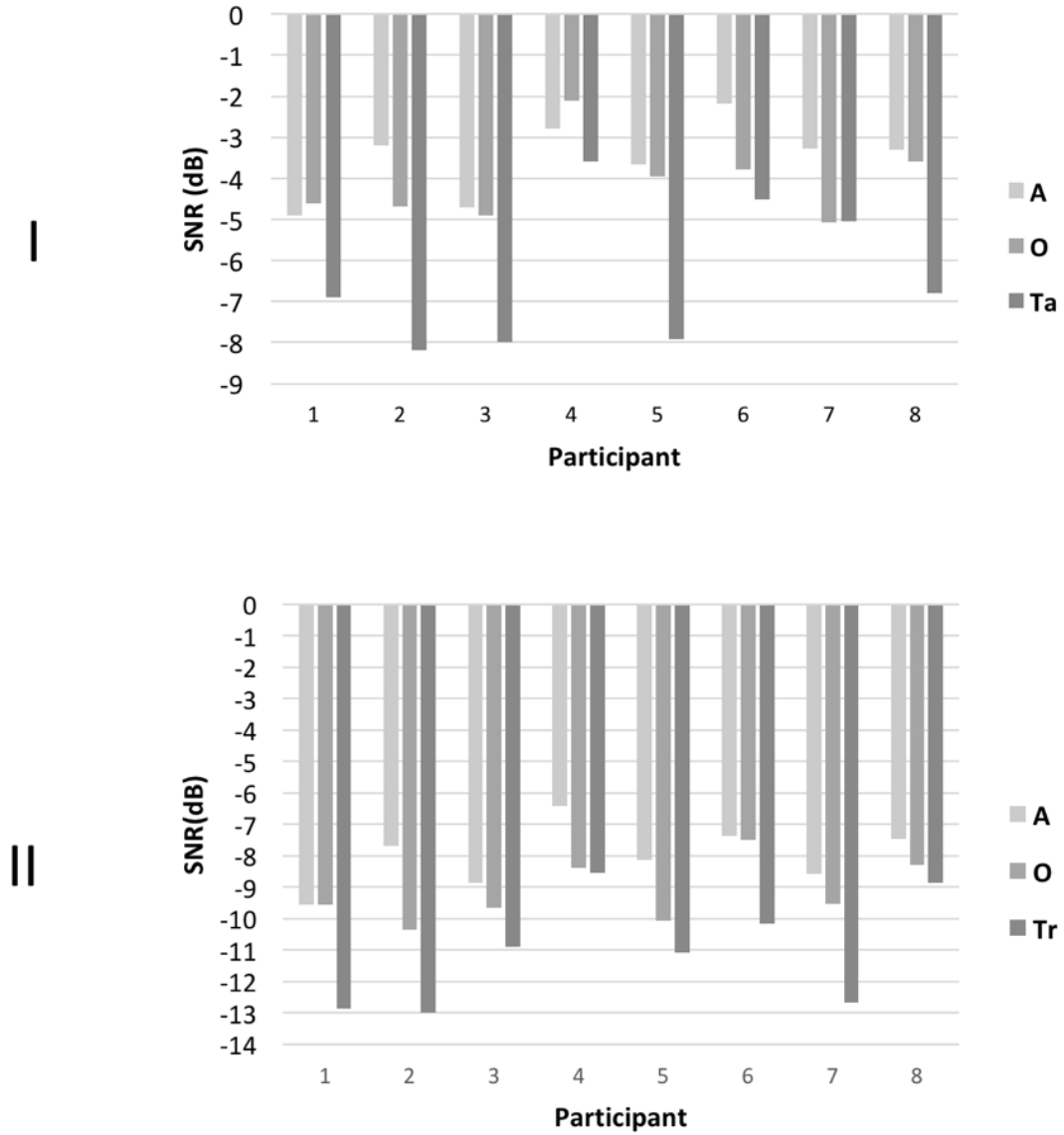
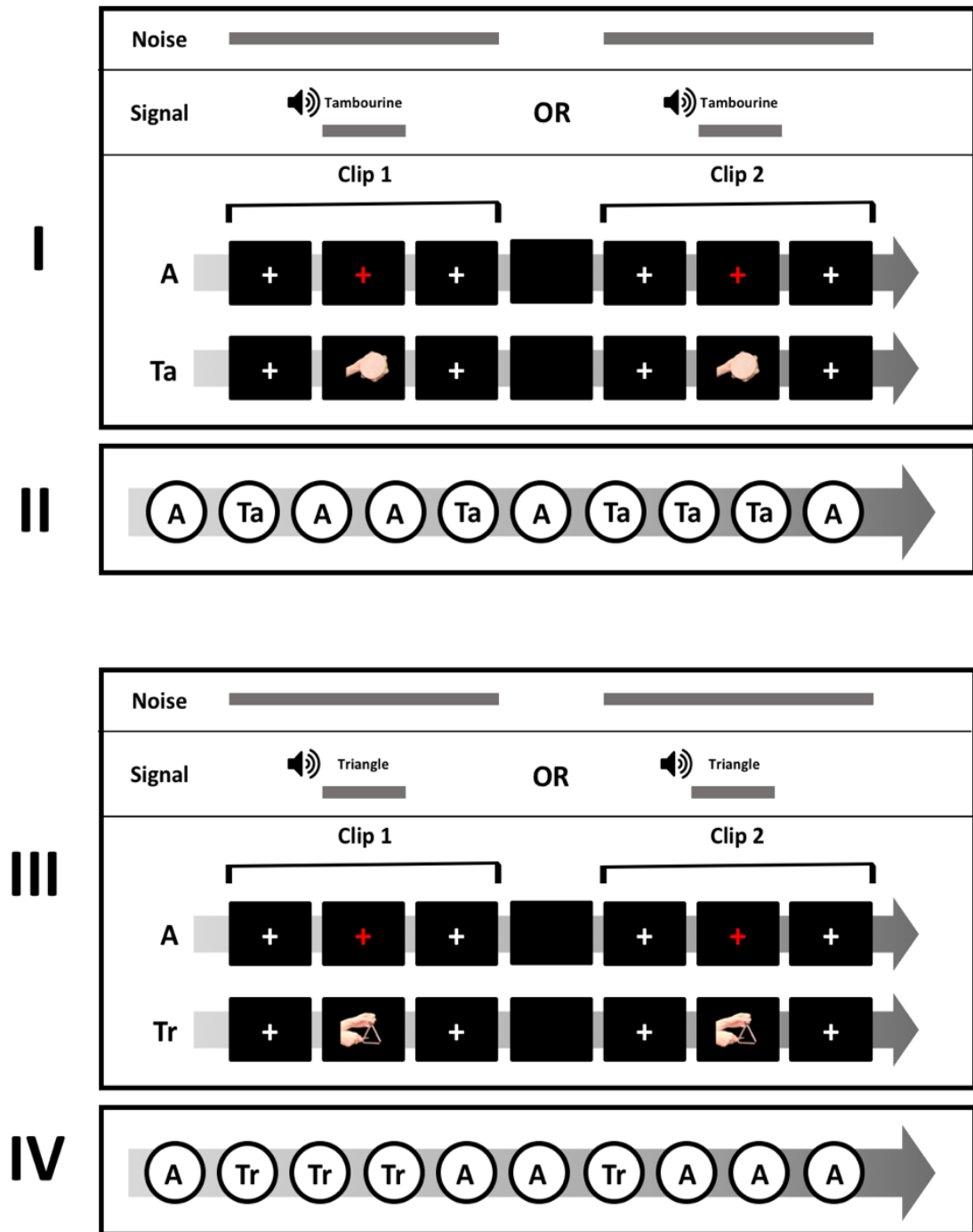
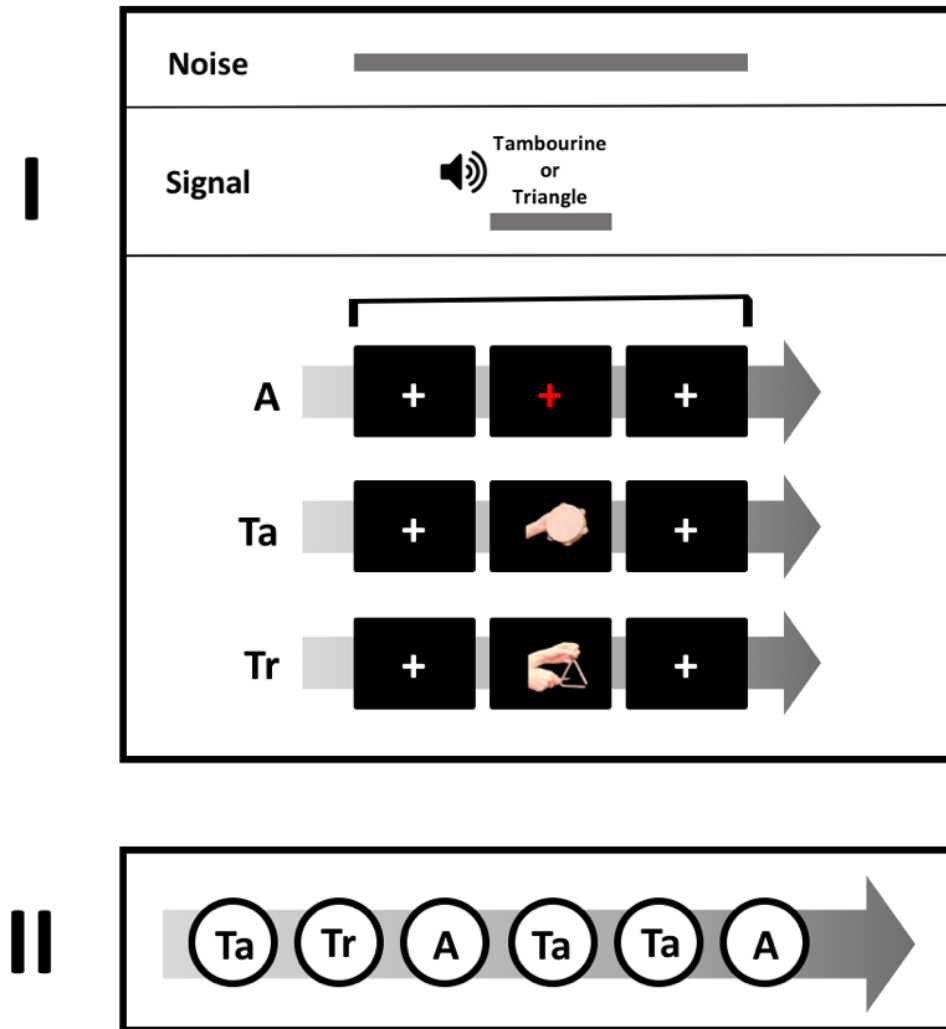


Figure 3. Individual means for thresholds for each participant in each tambourine (I) and the triangle (II) conditions. Means are dB signal-to-noise ratio at the estimated 70.7% threshold.



*Figure A1.* Two-interval forced-choice detection task within the three audiovisual conditions for the tambourine (I) and the triangle (III) stimuli. Each trial consisted on the presentation of two audiovisual clips with a short interval between them. Each clip contained acoustic noise and a target auditory signal, presented together to a white fixation cross, a white/red temporal cue fixation cross or a synchronous congruent object. A white fixation cross was presented at the beginning and at the ending of each clip. The sequence of trials was randomly determined across the three tambourine (II) and triangle (IV) conditions.



*Figure A2.* Two-alternative forced-choice discrimination task within the three audiovisual conditions for both tambourine and triangle stimuli (I). Each trial consisted on the presentation a clip containing acoustic noise and a target auditory signal, presented together to a white fixation cross, a white/red temporal cue fixation cross or a synchronous visual object. A white fixation cross was presented at the beginning and at the ending of each clip. The auditory signal could be either congruent or incongruent to the visual stimulus. The sequence of trials was randomly determined across all conditions for both instruments (II).

## 5. CONCLUSÕES E PERSPECTIVAS

Os paradigmas experimentais desenvolvidos nesse trabalho direcionaram-se a investigação efeitos intermodais entre a visão e audição em humanos, mais especificamente, de que forma estímulos visuais podem modular a percepção de sons semanticamente congruentes. Os resultados obtidos, ainda que preliminares, indicaram um efeito facilitador bastante considerável exercido pela visão na redução do limiar de percepção auditiva. Tais achados são motivadores para a continuidade desse projeto de pesquisa e a realização dos experimentos de detecção e de identificação perceptivo. Notavelmente, esses outros dois paradigmas experimentais dedicam-se também a mensurar a capacidade metacognitiva dos participantes, e relacioná-la com a performance comportamental de cada um ao detectar/identificar objetos e ao avaliar a confiabilidade de suas próprias respostas. Dessa forma, pretende-se investigar de maneira mais próxima as relações entre percepção e consciência multissensoriais.

Contudo, esse trabalho de conclusão de curso, mais do que resultar em alguns novo dados significativos, foi a minha primeira experiência num projeto ao qual eu **realmente** me interesse e tenho a perspectiva de continuar trabalhando no futuro. Enquanto revisava a literatura científica para a escrita desse texto, tive a oportunidade de aprender mais profundamente sobre meus temas de interesse, e inclusive mudei minhas ideias sobre alguns tópicos particulares. O desenvolvimento de novos paradigmas experimentais me obrigou a aprender noções de linguagem de programação (não com pouco esforço), o que na verdade provavelmente será um atributo muito útil em projetos acadêmicos futuros. Durante o desenvolvimento e aplicação dos experimentos, uma série de obstáculos foram enfrentados, os quais serviram como tarefas práticas de resiliência. A dedicação desse ano ao meu trabalho de conclusão de curso me trouxe uma rica experiência da prática científica, vendo a ciência muito mais como um processo contínuo de construção de conhecimento do que como obtenção de resultados diretos.

## REFERÊNCIAS

- AITCHISON, L.; BANG, D.; BAHRAMI, B.; LATHAM, P. E. Doubly Bayesian Analysis of Confidence in Perceptivo Decision-Making. **PLoS computational biology**, v. 11, n. 10, p. e1004519, 2015. Journal Article, Research Support, Non-U.S. Gov't, United States.
- ALSIUS, A.; MUNHALL, K. G. Detection of audiovisual speech correspondences without visual awareness. **Psychological science**, v. 24, n. 4, p. 423–31, 2013. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/23462756>>. Acesso em: 1/4/2016.
- ARDEN, G. B.; WOLF, J. E.; MESSITER, C. Electrical activity in visual cortex associated with combined auditory and visual stimulation in temporal sequences known to be associated with a visual illusion. **Vision research**, v. 43, n. 23, p. 2469–2478, 2003. Journal Article, England.
- BAARS, B. J. The conscious access hypothesis: origins and recent evidence. **Trends in cognitive sciences**, v. 6, n. 1, p. 47–52, 2002. JOURNAL ARTICLE, England.
- BAARS, B. J.; FRANKLIN, S.; RAMSOY, T. Z. Global workspace dynamics: cortical “binding and propagation” enables conscious contents. **Frontiers in psychology**, v. 4, p. 200, 2013. Journal Article, Switzerland.
- BALDUZZI, D.; TONONI, G. Integrated information in discrete dynamical systems: motivation and theoretical framework. **PLoS computational biology**, v. 4, n. 6, p. e1000091, 2008. Journal Article, Research Support, N.I.H., Extramural, Research Support, Non-U.S. Gov't, United States.
- BLAKE, R.; TONG, F. Binocular rivalry. **Scholarpedia**, v. 3, n. 12, p. 1578, 2008.
- BRUNEL, L.; CARVALHO, P. F.; GOLDSTONE, R. L. It does belong together: Cross-modal correspondences influence cross-modal integration during perceptivo learning. **Frontiers in Psychology**, v. 6, n. MAR, p. 1–10, 2015.
- CHEN, Y.-C.; SPENCE, C. Crossmodal semantic priming by naturalistic sounds and spoken words enhances visual sensitivity. **Journal of experimental psychology. Human perception and performance**, v. 37, n. 5, p. 1554–68, 2011a. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/21688942>>. Acesso em: 1/4/2016.
- CHEN, Y.-C.; SPENCE, C. The crossmodal facilitation of visual object representations by sound: evidence from the backward masking paradigm. **Journal of experimental psychology. Human perception and performance**, v. 37, n. 6, p. 1784–802, 2011b. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/21942418>>. Acesso em: 1/4/2016.
- CHEN, Y. C.; YEH, S. L.; SPENCE, C. Crossmodal constraints on human perceptivo awareness: Auditory semantic modulation of binocular rivalry. **Frontiers in Psychology**, v. 2, n. SEP, p. 1–13, 2011.
- CHOU, W.-L.; YEH, S.-L. Object-based attention occurs regardless of object awareness. **Psychonomic bulletin & review**, v. 19, n. 2, p. 225–31, 2012. Disponível em:

<<http://www.ncbi.nlm.nih.gov/pubmed/22237418>>. Acesso em: 1/4/2016.

CLARKE, F. R.; BIRDSALL, T. G.; JR., W. P. T. Two Types of ROC Curves and Definitions of Parameters. **J. Acoust. Soc. Am.**, v. 31, 1959.

CLIFFORD, C. W. G.; ARABZADEH, E.; HARRIS, J. A. Getting technical about awareness. **Trends in cognitive sciences**, v. 12, n. 2, p. 54–58, 2008. Journal Article, Research Support, Non-U.S. Gov't, England.

CONRAD, V.; BARTELS, A.; KLEINER, M.; NOPPENY, U. Audiovisual interactions in binocular rivalry. **Journal of Vision**, v. 10, n. 10, p. 27, 2010.

DAMASIO, A. **Self Comes to Mind**. Vintage Books, 2010.

DAMASIO, A.; CARVALHO, G. B. The nature of feelings: evolutionary and neurobiological origins. **Nature reviews. Neuroscience**, v. 14, n. 2, p. 143–152, 2013. Journal Article, Research Support, N.I.H., Extramural, Research Support, Non-U.S. Gov't, Review, England.

DAMASIO, A. R. Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition. **Cognition**, v. 33, n. 1–2, p. 25–62, 1989. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/2691184>>. Acesso em: 26/11/2015.

DAMASIO, A. R. **Descartes Error**. Penguin Books, 1994.

DAMASIO, A. R. **The Feeling of What Happens: Body and Emotion in the Making of Consciousness**. New York, USA: Harcourt Brace, 1999.

DEHAENE, S.; CHANGEUX, J.-P. Experimental and theoretical approaches to conscious processing. **Neuron**, v. 70, n. 2, p. 200–227, 2011. Journal Article, Research Support, Non-U.S. Gov't, Review, United States.

DEHAENE, S.; CHARLES, L.; KING, J.-R.; MARTI, S. Toward a computational theory of conscious processing. **Current opinion in neurobiology**, v. 25, p. 76–84, 2014. Journal Article, Research Support, Non-U.S. Gov't, Review, England.

DEROY, O.; CHEN, Y.-C.; SPENCE, C. Multisensory constraints on awareness. **Philosophical transactions of the Royal Society of London. Series B, Biological sciences**, v. 369, n. 1641, p. 20130207, 2014. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3965162&tool=pmcentrez&rendertype=abstract>>. Acesso em: 29/3/2016.

DEROY, O.; SPENCE, C.; NOPPENY, U. Metacognition in multisensory perception. **Trends in Cognitive Sciences**, v. xx, p. 1–12, 2016. Elsevier Ltd. Disponível em: <<http://dx.doi.org/10.1016/j.tics.2016.08.006>>. .

DIENES, Z.; SETH, A. Gambling on the unconscious: a comparison of wagering and confidence ratings as measures of awareness in an artificial grammar task. **Consciousness and cognition**, v. 19, n. 2, p. 674–681, 2010. Comparative Study, Journal Article, United States.

DOEHRMANN, O.; NAUMER, M. J. Semantics and the multisensory brain: How meaning modulates processes of audio-visual integration. **Brain Research**, v. 1242, p. 136–150, 2008. Elsevier B.V. Disponível em: <<http://dx.doi.org/10.1016/j.brainres.2008.03.071>>. .

DRIVER, J.; SPENCE, C. Multisensory perception: Beyond modularity and convergence. **Current Biology**, v. 10, n. 20, p. 10–12, 2000.

EDELMAN, G. M.; TONONI, G. **A universe of consciousness: How matter becomes imagination**. New York, USA: Basic Books, 2000.

ENGEL, A. K.; FRIES, P.; KONIG, P.; BRECHT, M.; SINGER, W. Temporal binding, binocular rivalry, and consciousness. **Consciousness and cognition**, v. 8, n. 2, p. 128–151, 1999. Journal Article, Research Support, Non-U.S. Gov't, United States.

ERNST, M. O.; BANKS, M. S. Humans integrate visual and haptic information in a statistically optimal fashion. **Nature**, v. 415, n. 6870, p. 429–433, 2002.

FAHRENFORT, J. J.; LAMME, V. A. F. A true science of consciousness explains phenomenology: comment on Cohen and Dennett. **Trends in cognitive sciences**, Mar. 2012. Comment, Letter, England.

FAIVRE, N.; MUDRIK, L.; SCHWARTZ, N.; KOCH, C. Multisensory integration in complete unawareness: evidence from audiovisual congruency priming. **Psychological science**, v. 25, n. 11, p. 2006–16, 2014. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/25269620>>. Acesso em: 1/4/2016.

FLEMING, S. M.; LAU, H. C. How to measure metacognition. **Frontiers in Human Neuroscience**, v. 8, n. July, p. 443, 2014. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4097944&tool=pmcentrez&rendertype=abstract>>. .

GAU, R.; NOPPENY, U. How prior expectations shape multisensory perception. **NeuroImage**, v. 124, n. Pt A, p. 876–886, 2016. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/26419391>>. Acesso em: 5/4/2016.

GHAZANFAR, A. A.; SCHROEDER, C. E. Is neocortex essentially multisensory? **Trends in Cognitive Sciences**, v. 10, n. 6, p. 278–285, 2006.

GILLMEISTER, H.; EIMER, M. Tactile enhancement of auditory detection and perceived loudness. **Brain Research**, v. 1160, n. 1, p. 58–68, 2007.

GRANT, K. W. The effect of speechreading on masked detection thresholds for filtered speech. **The Journal of the Acoustical Society of America**, v. 109, n. 5 Pt 1, p. 2272–2275, 2001.

GROEN, O. VAN DER; BURG, E. VAN DER; LUNGHI, C.; ALAIS, D. Touch influences visual perception with a tight orientation-tuning. **PloS one**, v. 8, n. 11, p. e79558, 2013.

Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3828350&tool=pmcentrez&rendertype=abstract>>. .



ype=abstract>. Acesso em: 1/4/2016.

HELMHOLTZ, H. VON. **Handbuch der physiologischen Optik**. Leipzig: Leipzig: Leopold Voss, 1867.

HOLMES, N. P.; SPENCE, C. Multisensory integration: Space, time and superadditivity. **Current Biology**, v. 15, n. 18, 2005.

HSIAO, J.-Y.; CHEN, Y.-C.; SPENCE, C.; YEH, S.-L. Assessing the effects of audiovisual semantic congruency on the perception of a bistable figure. **Consciousness and cognition**, v. 21, n. 2, p. 775–87, 2012. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/22386069>>. Acesso em: 1/4/2016.

IORDANESCU, L.; GUZMAN-MARTINEZ, E.; GRABOWECKY, M.; SUZUKI, S. Characteristic sounds facilitate visual search. **Psychonomic bulletin & review**, v. 15, n. 3, p. 548–54, 2008. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2647585&tool=pmcentrez&rendertype=abstract>>. Acesso em: 1/4/2016.

KADUNCE, D. C.; VAUGHAN, J. W.; WALLACE, M. T.; BENEDEK, G.; STEIN, B. E. Mechanisms of within- and cross-modality suppression in the superior colliculus. **Journal of neurophysiology**, v. 78, n. 6, p. 2834–2847, 1997. Journal Article, Research Support, Non-U.S. Gov't, Research Support, U.S. Gov't, Non-P.H.S., Research Support, U.S. Gov't, P.H.S., United States.

KADUNCE, D. C.; VAUGHAN, J. W.; WALLACE, M. T.; STEIN, B. E. The influence of visual and auditory receptive field organization on multisensory integration in the superior colliculus. **Experimental brain research**, v. 139, n. 3, p. 303–310, 2001. Journal Article, Research Support, U.S. Gov't, P.H.S., Germany.

KANG, M.; BLAKE, R. Perceptivo synergy between seeing and hearing revealed during binocular rivalry. **Psychologija**, v. 32, p. 7–15, 2005.

KANWISHER, N. Neural events and perceptivo awareness. **Cognition**, v. 79, n. 1–2, p. 89–113, 2001. Journal Article, Research Support, Non-U.S. Gov't, Research Support, U.S. Gov't, P.H.S., Netherlands.

KAYSER, C.; LOGOTHETIS, N. K. Do early sensory cortices integrate cross-modal information? **Brain Structure and Function**, v. 212, n. 2, p. 121–132, 2007.

KAYSER, C.; PETKOV, C. I.; LOGOTHETIS, N. K. Visual modulation of neurons in auditory cortex. **Cerebral Cortex**, v. 18, n. 7, p. 1560–1574, 2008.

KAYSER, C.; SHAMS, L. Multisensory Causal Inference in the Brain. **PLoS Biology**, v. 13, n. 2, p. 1–7, 2015.

KEANE, B.; SPENCE, M.; YARROW, K.; ARNOLD, D. Perceptivo confidence demonstrates trial-by-trial insight into the precision of audio-visual timing encoding. **Consciousness and**

**Cognition**, v. 38, p. 107–117, 2015. Elsevier Inc. Disponível em: <<http://dx.doi.org/10.1016/j.concog.2015.10.010>> . .

KERSTEN, D.; MAMASSIAN, P.; YUILLE, A. Object perception as Bayesian inference. **Annual review of psychology**, v. 55, p. 271–304, 2004. Journal Article, Research Support, Non-U.S. Gov't, Research Support, U.S. Gov't, P.H.S., Review, United States.

KOCH, C. **Consciousness – Confessions of a Romantic Reductionist**. MIT Press, 2012.

KORDING, K. P. Bayesian statistics: Relevant for the brain? **Current Opinion in Neurobiology**, v. 25, p. 130–133, 2014.

KÖRDING, K. P.; BEIERHOLM, U.; MA, W. J.; et al. Causal inference in multisensory perception. **PLoS ONE**, v. 2, n. 9, 2007.

LAMME, V. A. F. Towards a true neural stance on consciousness. **Trends in cognitive sciences**, v. 10, n. 11, p. 494–501, 2006. Journal Article, Review, England.

LUNGI, C.; BINDA, P.; MORRONE, M. C. Touch disambiguates rivalrous perception at early stages of visual analysis. **Current biology : CB**, v. 20, n. 4, p. R143-4, 2010. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/20178754>>. Acesso em: 31/12/2015.

MAN, K.; DAMASIO, A.; MEYER, K.; KAPLAN, J. T. Convergent and invariant object representations for sight, sound, and touch. **Human Brain Mapping**, v. 36, n. 9, p. 3629–3640, 2015.

MAN, K.; KAPLAN, J. T.; DAMASIO, A.; MEYER, K. Sight and sound converge to form modality-invariant representations in temporoparietal cortex. **The Journal of neuroscience : the official journal of the Society for Neuroscience**, v. 32, n. 47, p. 16629–36, 2012. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3667662&tool=pmcentrez&rendertype=abstract>>. Acesso em: 26/11/2015.

MANISCALCO, B.; LAU, H. A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. **Consciousness and Cognition**, v. 21, n. 1, p. 422–430, 2012. Elsevier Inc. Disponível em: <<http://dx.doi.org/10.1016/j.concog.2011.09.021>> . .

MARCHANT, J. L.; RUFF, C. C.; DRIVER, J. Audiovisual synchrony enhances BOLD responses in a brain network including multisensory STS while also enhancing target-detection performance for both modalities. **Human Brain Mapping**, v. 33, n. 5, p. 1212–1224, 2012.

MASSONI, S.; GAJDOS, T.; VERGNAUD, J. C. Confidence measurement in the light of signal detection theory. **Frontiers in Psychology**, v. 5, n. DEC, p. 1–13, 2014.

MCGURK, H.; MACDONALD, J. Hearing lips and seeing voices. **Nature**, v. 264, n. 5588, p. 746–748, 1976. Journal Article, England.

MEREDITH, M. A.; NEMITZ, J. W.; STEIN, B. E. Determinants of multisensory integration in

superior colliculus neurons. I. Temporal factors. **The Journal of neuroscience : the official journal of the Society for Neuroscience**, v. 7, n. 10, p. 3215–3229, 1987. Journal Article, Research Support, U.S. Gov't, P.H.S., United States.

MEREDITH, M. A.; STEIN, B. E. Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. **Journal of neurophysiology**, v. 56, n. 3, p. 640–662, 1986. Journal Article, Research Support, U.S. Gov't, P.H.S., United States.

MEYER, K.; DAMASIO, A. Convergence and divergence in a neural architecture for recognition and memory. **Trends in neurosciences**, v. 32, n. 7, p. 376–82, 2009. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/19520438>>. Acesso em: 26/11/2015.

MEYER, K.; KAPLAN, J. T.; ESSEX, R.; et al. Predicting visual stimuli on the basis of activity in auditory cortices. **Nature neuroscience**, v. 13, n. 6, p. 667–8, 2010. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/20436482>>. Acesso em: 26/11/2015.

MEYER, K.; KAPLAN, J. T.; ESSEX, R.; DAMASIO, H.; DAMASIO, A. Seeing touch is correlated with content-specific activity in primary somatosensory cortex. **Cerebral cortex (New York, N.Y. : 1991)**, v. 21, n. 9, p. 2113–2121, 2011. Journal Article, Research Support, N.I.H., Extramural, Research Support, Non-U.S. Gov't, United States.

MUDRIK, L.; FAIVRE, N.; KOCH, C. Information integration without awareness. **Trends in Cognitive Sciences**, v. 18, n. 9, p. 488–496, 2014. Elsevier Ltd. Disponível em: <<http://dx.doi.org/10.1016/j.tics.2014.04.009>>. .

MURRAY, M. M.; WALLACE, M. T. **Frontiers in the Neural Bases of Multisensory Processes**. Boca Ratan, FL: CRC, 2012.

NOESSELT, T.; RIEGER, J. W.; SCHOENFELD, M. A.; et al. Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. **The Journal of neuroscience : the official journal of the Society for Neuroscience**, v. 27, n. 42, p. 11431–11441, 2007. Comparative Study, Journal Article, Research Support, Non-U.S. Gov't, United States.

NOPPENY, U.; FAIVRE, N.; LUNGHI, C.; et al. The Complex Interplay Between Multisensory Integration and Perceptivo Awareness. **Multisensory Research**, p. 1–22, 2016. Disponível em: <<http://booksandjournals.brillonline.com/content/journals/10.1163/22134808-00002529>>. .

O'HARE, J. J. Perceptivo integration. **Journal of the Washington Academy of Sciences**, v. 81, n. 1, p. 44–59, 1991.

ODGAARD, E. C.; ARIEH, Y.; MARKS, L. E. Brighter noise: sensory enhancement of perceived loudness by concurrent visual stimulation. **Cognitive, affective & behavioral neuroscience**, v. 4, n. 2, p. 127–132, 2004.

PALMER, T. D.; RAMSEY, A. K. The function of consciousness in multisensory integration. **Cognition**, v. 125, n. 3, p. 353–364, 2012. Elsevier B.V. Disponível em:

<<http://dx.doi.org/10.1016/j.cognition.2012.08.003>>. .

PASCUAL-LEONE, A.; HAMILTON, R. The metamodal organization of the brain. **Progress in brain research**, v. 134, p. 427–445, 2001. Historical Article, Journal Article, Research Support, U.S. Gov't, P.H.S., Review, Netherlands.

PERSAUD, N.; MCLEOD, P.; COWEY, A. Post-decision wagering objectively measures awareness. **Nature neuroscience**, v. 10, n. 2, p. 257–261, 2007. Journal Article, Research Support, Non-U.S. Gov't, United States.

PLASS, J.; GUZMAN-MARTINEZ, E.; ORTEGA, L.; GRABOWECKY, M.; SUZUKI, S. Lip reading without awareness. **Psychological science**, v. 25, n. 9, p. 1835–7, 2014. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4303247&tool=pmcentrez&rendertype=abstract>>. Acesso em: 1/4/2016.

SCHIRILLO, J. A. Cross-modal detection using various temporal and spatial configurations. **Attention, perception & psychophysics**, v. 73, n. 1, p. 237–246, 2011.

SCHNEIDER, T. R.; ENGEL, A. K.; DEBENER, S. Multisensory identification of natural objects in a two-way crossmodal priming paradigm. **Experimental psychology**, v. 55, n. 2, p. 121–32, 2008. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/18444522>>. Acesso em: 1/4/2016.

SCHWARTZ, J. L.; BERTHOMMIER, F.; SAVARIAUX, C. Seeing to hear better: Evidence for early audio-visual interactions in speech identification. **Cognition**, v. 93, n. 2, p. 69–78, 2004.

SEKULER, R.; SEKULER, A. B.; LAU, R. Sound alters visual motion perception. **Nature**, 1997.

SHAMS, L.; IWAKI, S.; CHAWLA, A.; BHATTACHARYA, J. Early modulation of visual cortex by sound: An MEG study. **Neuroscience Letters**, v. 378, n. 2, p. 76–81, 2005.

SHAMS, L.; KAMITANI, Y.; SHIMOJO, S. Illusions. What you see is what you hear. **Nature**, v. 408, n. 6814, p. 788, 2000. Journal Article, England.

SPENCE, C.; BAYNE, T. **Is consciousness multisensory? in: Perception and Its Modalities**. Oxford, UK: Oxford University Press, 2015.

STEIN, B. E.; ARIGBEDE, M. O. Unimodal and multimodal response properties of neurons in the cat's superior colliculus. **Experimental neurology**, v. 36, n. 1, p. 179–196, 1972. Journal Article, United States.

STEIN, B. E.; STANFORD, T. R.; ROWLAND, B. A. The neural basis of multisensory integration in the midbrain: Its organization and maturation. **Hearing Research**, v. 258, n. 1–2, p. 4–15, 2009. Elsevier B.V. Disponível em: <<http://dx.doi.org/10.1016/j.heares.2009.03.012>>. .

STEIN, B. E.; STANFORD, T. R.; ROWLAND, B. A. Development of multisensory integration from the perspective of the individual neuron. , v. 15, n. 8, p. 520–535, 2014.

STEIN, B.; MEREDITH, M. **The merging of the senses**. Cambridge, Mass: MIT Press, 1993.

TAN, J.-S.; YEH, S.-L. Audiovisual integration facilitates unconscious visual scene processing. **Journal of experimental psychology. Human perception and performance**, v. 41, n. 5, p. 1325–35, 2015. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/26076179>>. Acesso em: 1/4/2016.

TJAN, B. S.; CHAO, E.; BERNSTEIN, L. E. A visual or tactile signal makes auditory speech detection more efficient by reducing uncertainty. **European Journal of Neuroscience**, v. 39, n. 8, p. 1323–1331, 2014.

TONONI, G. Consciousness as integrated information: a provisional manifesto. **The Biological bulletin**, v. 215, n. 3, p. 216–242, 2008. Journal Article, United States.

TONONI, G. Information integration: its relevance to brain function and consciousness. **Archives italiennes de biologie**, v. 148, n. 3, p. 299–322, 2010. Journal Article, Research Support, Non-U.S. Gov't, Italy.

TREISMAN, A. Consciousness and perceptivo binding. **The unity of consciousness: Binding, integration, and dissociation**. p.95–113, 2003. Oxford, UK: Oxford University Press.

TSUCHIYA, N. Flash suppression. **Scholarpedia**, v. 3, n. 2, p. 5640, 2008.

TURK, M. Multimodal interaction: A review. **Pattern Recognition Letters**, v. 36, n. 1, p. 189–195, 2014. Elsevier B.V. Disponível em: <<http://dx.doi.org/10.1016/j.patrec.2013.07.003>>. .

VROOMEN, J.; GELDER, B. DE. Sound Enhances Visual Perception: Cross-Modal Effects of Auditory Organization on Vision. **Journal of Experimental Psychology: Human Perception and Performance**, v. 26, n. 5, p. 1583–1590, 2000.

WASSENHOVE, V. VAN; GRANT, K. W.; POEPEL, D.; HALLE, M. Visual Speech Speeds up the Neural Processing of Auditory Speech. **Source: Proceedings of the National Academy of Sciences of the United States of America**, v. 102, n. 4, p. 1181–1186, 2005. Disponível em: <<http://www.jstor.org/stable/3374398>\n[http://www.jstor.org/stable/3374398?seq=1&cid=pdf-reference#references\\_tab\\_contents](http://www.jstor.org/stable/3374398?seq=1&cid=pdf-reference#references_tab_contents)\n<http://about.jstor.org/terms>>. .

WATKINS, S.; SHAMS, L.; JOSEPHS, O.; REES, G. Activity in human V1 follows multisensory perception. **NeuroImage**, v. 37, n. 2, p. 572–578, 2007. Journal Article, Research Support, Non-U.S. Gov't, United States.

WATKINS, S.; SHAMS, L.; TANAKA, S.; HAYNES, J.-D.; REES, G. Sound alters activity in human V1 in association with illusory visual perception. **NeuroImage**, v. 31, n. 3, p. 1247–1256, 2006. Journal Article, Research Support, Non-U.S. Gov't, United States.

ZHOU, W.; JIANG, Y.; HE, S.; CHEN, D. Olfaction modulates visual perception in binocular rivalry. **Current biology : CB**, v. 20, n. 15, p. 1356–8, 2010. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4226334&tool=pmcentrez&rendertype=abstract>>. Acesso em: 1/4/2016.

## ANEXO A – NORMAS DE PUBLICAÇÃO

O artigo científico apresentado nesse trabalho foi escrito como manuscrito a ser submetido para publicação na revista *Cognition*, da editora Elsevier. As normas para autores relevantes para a escrita do manuscrito são descritas a seguir, conforme indicado consta no “Guide for Authors – Authors information pack” no site da revista (<https://www.elsevier.com/journals/cognition/0010-0277/guide-for-authors>). Uma vez que o artigo foi escrito na língua inglesa, as normas são aqui descritas também em inglês, para garantir a adequação terminológica utilizada no texto.

Article structure:

Subdivision - numbered sections: Divide your article into clearly defined and numbered sections. Subsections should be numbered 1.1 (then 1.1.1, 1.1.2, ...), 1.2, etc. (the abstract is not included in section numbering). Use this numbering also for internal cross-referencing: do not just refer to 'the text'. Any subsection may be given a brief heading. Each heading should appear on its own separate line.

Introduction: State the objectives of the work and provide an adequate background, avoiding a detailed literature survey or a summary of the results.

Material and methods: Provide sufficient detail to allow the work to be reproduced. Methods already published should be indicated by a reference: only relevant modifications should be described.

Theory/calculation: A Theory section should extend, not repeat, the background to the article already dealt with in the Introduction and lay the foundation for further work. In contrast, a Calculation section represents a practical development from a theoretical basis.

Results: Results should be clear and concise.

Discussion: This should explore the significance of the results of the work, not repeat them. A combined Results and Discussion section is often appropriate. Avoid extensive citations and discussion of published literature.

Conclusions: The main conclusions of the study may be presented in a short Conclusions section, which may stand alone or form a subsection of a Discussion or Results and Discussion.

Appendices: If there is more than one appendix, they should be identified as A, B, etc. Formulae and equations in appendices should be given separate numbering: Eq. (A.1), Eq. (A.2),

etc.; in a subsequent appendix, Eq. (B.1) and so on. Similarly for tables and figures: Table A.1; Fig. A.1, etc.

Essential title page information:

**Title:** Concise and informative. Titles are often used in information-retrieval systems. Avoid abbreviations and formulae where possible.

**Author names and affiliations:** Please clearly indicate the given name(s) and family name(s) of each author and check that all names are accurately spelled. Present the authors' affiliation addresses (where the actual work was done) below the names. Indicate all affiliations with a lower-case superscript letter immediately after the author's name and in front of the appropriate address. Provide the full postal address of each affiliation, including the country name and, if available, the e-mail address of each author.

**Corresponding author:** Clearly indicate who will handle correspondence at all stages of refereeing and publication, also post-publication. Ensure that the e-mail address is given and that contact details are kept up to date by the corresponding author.

**Present/permanent address:** If an author has moved since the work described in the article was done, or was visiting at the time, a 'Present address' (or 'Permanent address') may be indicated as a footnote to that author's name. The address at which the author actually did the work must be retained as the main, affiliation address. Superscript Arabic numerals are used for such footnotes.

**Abstract:** A concise and factual abstract is required. The abstract should state briefly the purpose of the research, the principal results and major conclusions. An abstract is often presented separately from the article, so it must be able to stand alone. For this reason, References should be avoided, but if essential, then cite the author(s) and year(s). Also, non-standard or uncommon abbreviations should be avoided, but if essential they must be defined at their first mention in the abstract itself.

**Keywords:** Immediately after the abstract, provide a maximum of 6 keywords, using British spelling and avoiding general and plural terms and multiple concepts.

**Abbreviations:** Define abbreviations that are not standard in this field in a footnote to be placed on the first page of the article. Such abbreviations that are unavoidable in the abstract must be defined at their first mention there, as well as in the footnote. Ensure consistency of

abbreviations throughout the article.

**Acknowledgements:** Collate acknowledgements in a separate section at the end of the article before the references and do not, therefore, include them on the title page, as a footnote to the title or otherwise. List here those individuals who provided help during the research (e.g., providing language help, writing assistance or proof reading the article, etc.).

**Formatting of funding sources:** List funding sources in this standard way to facilitate compliance to funder's requirements: **Funding:** This work was supported by the National Institutes of Health [grant numbers xxxx, yyyy]; the Bill & Melinda Gates Foundation, Seattle, WA [grant number zzzz]; and the United States Institutes of Peace [grant number aaaa]. It is not necessary to include detailed descriptions on the program or type of grants and awards. When funding is from a block grant or other resources available to a university, college, or other research institution, submit the name of the institute or organization that provided the funding. If no funding has been provided for the research, please include the following sentence: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Footnotes:** Footnotes should be used sparingly. Number them consecutively throughout the article. Many word processors build footnotes into the text, and this feature may be used. Should this not be the case, indicate the position of footnotes in the text and present the footnotes themselves separately at the end of the article.

**Artwork:** Make sure you use uniform lettering and sizing of your original artwork. Preferred fonts: Arial (or Helvetica), Times New Roman (or Times), Symbol, Courier. Number the illustrations according to their sequence in the text. Use a logical naming convention for your artwork files. Indicate per figure if it is a single, 1.5 or 2-column fitting image. For Word submissions only, you may still provide figures and their captions, and tables within a single file at the revision stage. Please note that individual figure files larger than 10 MB must be provided in separate source files. A detailed guide on electronic artwork is available. You are urged to visit this site; some excerpts from the detailed information are given here.

**Formats** Regardless of the application used, when your electronic artwork is finalized, please 'save as' or convert the images to one of the following formats (note the resolution requirements for line drawings, halftones, and line/halftone combinations given below): EPS (or PDF): Vector drawings. Embed the font or save the text as 'graphics'. TIFF (or JPG): Color or grayscale



photographs (halftones): always use a minimum of 300 dpi. TIFF (or JPG): Bitmapped line drawings: use a minimum of 1000 dpi. TIFF (or JPG): Combinations bitmapped line/half-tone (color or grayscale): a minimum of 500 dpi is required. Please do not: Supply files that are optimized for screen use (e.g., GIF, BMP, PICT, WPG); the resolution is too low. Supply files that are too low in resolution. Submit graphics that are disproportionately large for the content.

Figure captions: Ensure that each illustration has a caption. A caption should comprise a brief title (not on the figure itself) and a description of the illustration. Keep text in the illustrations themselves to a minimum but explain all symbols and abbreviations used.

Tables: Please submit tables as editable text and not as images. Tables can be placed either next to the relevant text in the article, or on separate page(s) at the end. Number tables consecutively in accordance with their appearance in the text and place any table notes below the table body. Be sparing in the use of tables and ensure that the data presented in them do not duplicate results described elsewhere in the article. Please avoid using vertical rules.

#### References:

Citation in text: Please ensure that every reference cited in the text is also present in the reference list (and vice versa). Any references cited in the abstract must be given in full. Unpublished results and personal communications are not recommended in the reference list, but may be mentioned in the text. If these references are included in the reference list they should follow the standard reference style of the journal and should include a substitution of the publication date with either 'Unpublished results' or 'Personal communication'. Citation of a reference as 'in press' implies that the item has been accepted for publication.

Reference management software: Most Elsevier journals have their reference template available in many of the most popular reference management software products. These include all products that support Citation Style Language styles, such as Mendeley and Zotero, as well as EndNote. Using the word processor plug-ins from these products, authors only need to select the appropriate journal template when preparing their article, after which citations and bibliographies will be automatically formatted in the journal's style. If no template is yet available for this journal, please follow the format of the sample references and citations as shown in this Guide. Users of Mendeley Desktop can easily install the reference style for this journal by clicking the following link: <http://open.mendeley.com/use-citation-style/cognition>.