



UNIVERSIDADE D
COIMBRA

Cristina Jordão Nazaré

UNIMODALIDADE E MULTIMODALIDADE NAS
REPRESENTAÇÕES DINÂMICAS: FORMATOS
VISUAIS, AUDITIVOS E AUDIOVISUAIS

Tese no âmbito do doutoramento em Psicologia, na especialidade de Psicologia Social e Cognitiva, orientada pelos Professores Doutores Armando Luíz Dinis Mónica de Oliveira, Carlos Manuel Lopes Pires e Michel-Ange Amorim e apresentada à Faculdade de Psicologia e de Ciências da Educação da Universidade de Coimbra.

Julho de 2021



UNIVERSIDADE D
COIMBRA

Cristina Jordão Nazaré

**UNIMODALIDADE E MULTIMODALIDADE
NAS REPRESENTAÇÕES DINÂMICAS:
FORMATOS VISUAIS, AUDITIVOS E
AUDIOVISUAIS**

Tese no âmbito do doutoramento em Psicologia, na especialidade de Psicologia Social e Cognitiva, orientada pelos Professores Doutores Armando Luíz Dinis Mónica de Oliveira, Carlos Manuel Lopes Pires e Michel-Ange Amorim e apresentada à Faculdade de Psicologia e de Ciências da Educação da Universidade de Coimbra.

Julho de 2021

Aos meus pais maravilhosos

Referência:

Nazaré, Cristina Jordão. (2021). *Unimodalidade e multimodalidade nas representações dinâmicas: Formatos visuais, auditivos e audiovisuais*. Tese de doutoramento não-publicada, Faculdade de Psicologia e de Ciências da Educação, Universidade de Coimbra, Coimbra, Portugal.

Endereço eletrónico da autora:

cristinanazare@estescoimbra.pt

Apoio:

AGRADECIMENTOS

Os meus maiores agradecimentos a todos aqueles que, das mais variadas formas, acompanharam, colaboraram e permitiram que esta dissertação se concretizasse e as minhas maiores desculpas a todos aqueles, que de algum modo, foram envolvidos nas dificuldades e privações ao longo deste percurso.

Os meus agradecimentos e reconhecimento dirigem-se em particular ao Professor Doutor Armando Luíz Dinis Mónica de Oliveira, pela dedicada e superior orientação nas distintas fases deste trabalho, pela disponibilidade, ajuda e incentivo constantes, pelos ensinamentos, sugestões e correções e pelo imenso trabalho, generosidade e compreensão. Os meus agradecimentos pela atenção são extensíveis aos orientadores, Professor Doutor Carlos Pires e Professor Doutor Michel-Ange Amorim.

Um forte agradecimento às colegas de Audiologia e aos colegas de gabinete da Escola Superior de Tecnologia da Saúde de Coimbra do Instituto Politécnico de Coimbra (ESTeSC-IPC) pelo trabalho, ajuda, amizade e partilha de momentos. Pelo apoio e/ou participação no estudo, também um grande agradecimento aos colegas docentes, não-docentes, estudantes e alumni da ESTeSC-IPC e estudantes, colegas de doutoramento e colegas do Laboratório de Psicologia Cognitiva Aplicada da Universidade de Coimbra e igualmente à família e amigos.

Um infinito agradecimento aos meus pais, Maria Alice Jordão e Silvério Nazaré pelo amor, carinho, ajuda, força, dedicação e exemplo, assim como um agradecimento de coração a toda a família (mãe, pai, avó [avós], afilhadas, primas, primos, madrinhas, padrinhos, tias e tios) e a todas as amigas e amigos pelo carinho, amizade e apoio, pela compreensão pelas ausências, ajuda e incentivo e pelo entusiasmo, sorrisos e alegrias partilhadas.

Muito obrigada a todos.

RESUMO

O presente estudo investiga o grau em que propriedades temporais e espaciais do som podem influenciar o processamento visual do movimento em tarefas de localização espacial, recorrendo a um paradigma de Momento Representacional (MR) intermodal. Os participantes localizaram a última posição percebida de um alvo visual em movimento que imprevistamente desaparecia, ao mesmo tempo que sons contínuos, mantendo relações temporais e/ou espaciais variáveis com o início (*onset*) e o fim (*offset*) do movimento visual, eram apresentados por auscultadores.

As tarefas de MR visual com movimentos horizontais dão tipicamente lugar a erros de localização para diante, no sentido do movimento (deslocamento M), e para baixo, ortogonais à direção do movimento (deslocamento O, alinhado com a direção da gravidade). Examinar a forma como estes deslocamentos são modulados pela informação auditiva convoca várias questões em aberto no domínio do processamento audiovisual (AV). Primeiro, a de saber se a audição pode afetar a localização espacial de estímulos dinâmicos em visão central (a evidência encontra-se limitada a estímulos estáticos degradados). Segundo, a de saber se os sons podem modular a localização visualmente percebida de um alvo móvel através da informação temporal, espacial, ou de ambos os tipos de informação (apenas efeitos temporalmente mediados foram demonstrados). Por fim, a de saber se as pistas auditivas podem induzir efeitos de melhoria audiovisual na localização de estímulos dinâmicos visuais (apenas enviesamentos de atração foram reportados em estudos anteriores de MR audiovisual).

O estudo compreendeu três tarefas contrabalanceadas obedecendo a um mesmo desenho fatorial *onset* (disparidades AV no início) × *offset* (disparidades AV no final) mas diferindo no tipo de pista auditiva fornecida: temporal (tarefa T), espacial (tarefa E) e espaço-temporal (tarefa ET). Uma manipulação adicional envolveu o tamanho e número das

disparidades AV, maiores (de modo a aumentar a sua saliência perceptiva) e em menor número (de modo a aumentar a sua fiabilidade) do que nos estudos anteriores. A hipótese subjacente a esta manipulação foi a de que poderia promover a relevância percebida para a tarefa das contingências AV e mobilizar mecanismos intermodais distintos dos que suportam os enviesamentos de atração para o som. O seguimento oculomotor dos alvos foi permitido, de modo a assegurar que a resposta de localização ocorria em condições de visão central.

Os resultados relativos ao deslocamento M documentaram consistentemente a capacidade dos sons para modular o MR visual por meio de pistas auditivas temporais (tarefa T) e espaciais (tarefa E). Este efeito foi atribuível às disparidades AV no *offset*, apresentando as disparidades no *onset* efeitos nulos em todas as tarefas. Ao invés de enviesamentos de atração, foi observado um benefício audiovisual generalizado (localização mais exata do alvo nas condições AV). A hipótese de que disparidades AV mais salientes e em menor número poderiam promover a sua relevância para a tarefa e induzir uma alteração qualitativa do padrão dos efeitos modulatórios foi assim suportada. A localização nas condições AV da tarefa E não correspondeu a uma combinação da informação posicional da visão e da audição (avaliadas em ensaios visuais intercalados e numa tarefa suplementar com alvos auditivos unimodais), excluindo o modelo Bayesiano da integração multissensorial como uma explicação da melhoria AV. No conjunto, os resultados mostraram-se favoráveis a uma interpretação baseada numa articulação flexível, mediada pela tarefa, entre processamento multissensorial e atenção endógena intermodal.

Os resultados relativos ao deslocamento O ilustraram o típico erro para baixo nas condições audiovisuais, visuais intercaladas e auditivas unimodais. A juntar ao paralelismo global do MR nas duas modalidades, a semelhança no deslocamento O reforça a perspectiva de que mecanismos amodais comuns (possivelmente preditivos) subjazem em parte à localização de alvos dinâmicos na visão e na audição. Não se observaram benefícios AV no deslocamento O. Diferentemente do deslocamento M, o *offset* percebido nas condições AV das tarefas E e ET pode ser modelizado como uma média ponderada da informação posicional da visão e da audição, envolvendo uma ponderação não ótima (à luz de um critério Bayesiano). Não se verificaram correlações entre a magnitude dos deslocamentos M e O. Embora a independência estatística não exclua a possibilidade de mecanismos comuns, é em todo o caso contrária a mecanismos partilhados que prevejam correlações entre os dois deslocamentos.

Vários dos resultados encontrados contêm implicações práticas para a reabilitação compensatória de défices do processamento visual do movimento, brevemente discutidas na secção final da dissertação.

Palavras-chave:

Momento representacional; Interação-integração audiovisual; Perceção do movimento; Atenção intermodal

ABSTRACT

The present study examines the extent to which temporal and spatial properties of sound modulate visual motion processing in spatial localization tasks. To that end, it resorts to a crossmodal Representational Momentum (RM) paradigm. Participants were asked to locate the place at which a moving visual target unexpectedly vanished, while continuous sounds were presented under variable temporal and/or spatial relationships with the onset and offset of visual motion.

The typical result in visual RM tasks with horizontal motions is a forward mislocalization in the direction of motion (M displacement), along with some mislocalization downward, orthogonal to motion direction (O displacement, aligned with gravity for horizontal motions). Studying the way these displacements, particularly the first one, are modulated by auditory information, can bring together several open issues in the realm of audiovisual processing. First, whether audition can influence vision in spatial tasks with dynamic stimuli under central vision (clear answers exist only for degraded static visual stimuli). Second, whether sounds can modulate the perceived location of a moving visual stimulus by means of temporal, spatial, or both kinds of auditory cues (only temporally-mediated effects have been demonstrated). Third, whether auditory cues can also lead to audiovisual enhancement in the visual localization of dynamic stimuli and under which conditions (only sound attraction biases have been reported in previous RM studies).

These issues were addressed by way of three counterbalanced tasks obeying a similar onset (AV disparities at the onset) \times offset (AV disparities at the offset) factorial design, but differing in the type of auditory cues provided: temporal only (task T), spatial only (task S), spatiotemporal (task ST). An additional core manipulation concerned the size and number of AV disparities, which were larger (expectedly increasing their salience) and fewer in number (expectedly increasing their reliability) than in prior studies. This was intended to promote

the perceived relevance of AV contingencies to the task and thereby potentially tap into crossmodal mechanisms other than those underpinning sound attraction biases. Oculomotor tracking of the target was allowed, ensuring that localization took place in central vision.

Results for M displacement consistently supported the ability of sounds to modulate visual RM through both temporal (task T) and spatial (task S) auditory cues. This effect was entirely due to AV disparities at the offset, with null results for those at the onset across all tasks. In contrast to previous studies, widespread audiovisual enhancement (more accurate localization of the target) was observed, rather than sound attraction biases. This concurred with the expectation that using more salient disparities in fewer numbers could promote their task-relevance and qualitatively shift the pattern of modulatory effects. Localization in AV conditions of task S could not be derived as a combination of visual and auditory positional information (assessed with visual-only trials and unimodal auditory targets in a supplementary task), excluding Bayesian-inspired models of multisensory integration as an account. Taken together, the findings were mostly suggestive of a task-dependent, flexible interplay between endogenous crossmodal attention and multisensory processing.

Results for O displacement disclosed the typical downward mislocalization both in audiovisual, visual-only and auditory unimodal condition. Adding to the overall parallelism of visual and auditory RM, this strengthens the view that common amodal mechanisms (possibly predictive) underlie to some extent the localization of dynamic targets in both vision and audition. No AV enhancement was found for O displacement. Differently from M displacement, the perceived offset in AV conditions of tasks S and ST could be modeled by a weighted average of visual and auditory positional information, although with non-optimal weights (as predicted by optimal Bayesian approaches). No correlations between the magnitude of M and O displacements were found. While statistical independence does not rule out the possibility of common mechanisms, it speaks against shared mechanisms that would predict correlations between the two types of displacement.

Several of the outcomes reported in the dissertation may hold practical implications for the compensatory rehabilitation of visual motion processing, which are briefly outlined in the final discussion.

Keywords:

Representational momentum; Audiovisual interaction-integration; Motion perception; Crossmodal attention

ÍNDICE

Introdução	19
PARTE I - ESTADO DA ARTE	27
1. Momento representacional: Um erro de localização do movimento.....	29
1.1. Definição, propostas de explicação e paradigma operacional.....	31
1.1.1. <i>Definição e propostas de explicação</i>	<i>31</i>
1.1.2. <i>Paradigma operacional.....</i>	<i>36</i>
1.2. Variáveis relevantes.....	38
1.2.1. <i>Características do alvo.....</i>	<i>38</i>
1.2.2. <i>Características da apresentação</i>	<i>40</i>
1.2.3. <i>Características do contexto.....</i>	<i>42</i>
1.2.4. <i>Características do observador</i>	<i>43</i>
1.3. Deslocamento M e deslocamento O: Gravidade representacional	45
1.4. Momento representacional auditivo: Semelhanças e diferenças relativamente ao momento representacional visual.....	48
2. Interação-integração multissensorial	53
2.1. Hipóteses e modelos: Dominância visual, adequação da modalidade, modelo de estimação da máxima verosimilhança, modelos Bayesianos de inferência causal	55
2.1.1. <i>Hipótese da dominância visual.....</i>	<i>55</i>

2.1.2.	<i>Hipótese da adequação da modalidade</i>	58
2.1.3.	<i>Modelos Bayesianos de integração sensorial</i>	60
2.1.3.1.	<i>Modelo da estimação de máxima verosimilhança (MMV)</i>	62
2.1.3.2.	<i>Modelo Bayesiano de inferência causal</i>	65
2.2.	Regras e princípios: Janela de integração temporal, regra espacial, princípio da eficiência inversa.	70
2.2.1.	<i>Regra temporal e janela de integração multissensorial</i>	71
2.2.2.	<i>Regra espacial</i>	74
2.2.3.	<i>Princípio da eficiência inversa</i>	76
2.3.	Integração multissensorial e atenção: Atenção intermodal exógena e endógena.....	77
2.4.	Assunção de unidade e a noção de percepto audiovisual	83
2.4.1.	<i>Problemas com a assunção de unidade</i>	83
2.4.2.	<i>O que é um percepto audiovisual?</i>	86
3.	Modulação da percepção do movimento visual pelas propriedades do som: Momento representacional audiovisual	89
3.1.	Impacto das propriedades espaciais, temporais e semânticas do som no movimento visual.....	89
3.2.	Momento representacional audiovisual: Efeitos de atração temporal e espacial do som.....	95
3.3.	Questões em aberto	100
3.3.1.	<i>Dissociação experimental do efeito da informação temporal e espacial do som</i>	100
3.3.2.	<i>Papel respetivo das disparidades audiovisuais no começo e termo do movimento</i>	102
3.3.3.	<i>Enviesamentos de atração nas condições síncronas versus efeitos de melhoria audiovisual</i>	104
3.3.4.	<i>Relações entre deslocamento M e deslocamento O no momento representacional visual, auditivo e audiovisual</i>	106
PARTE II - CONTRIBUTOS EMPIRÍCOS		109
4.	Processamento audiovisual numa tarefa de momento representacional: Efeitos da duração e da espacialização de um som na localização de um alvo visual	111
4.1.	Hipóteses e previsões	112
4.2.	A tarefa base: Opções operacionais	115
4.3.	Método	117
4.3.1.	<i>Participantes</i>	117
4.3.2.	<i>Estímulos</i>	118
4.3.3.	<i>Desenho experimental</i>	119

4.3.4.	<i>Procedimento experimental</i>	122
4.3.5.	<i>Análise dos dados</i>	123
5.	Deslocamento M	125
5.1.	Resultados	125
5.1.1.	<i>ANOVA global de medidas repetidas (5 vias)</i>	125
5.1.2.	<i>Efeitos das disparidades audiovisuais relativas ao offset</i>	127
5.1.3.	<i>Efeitos da magnitude das disparidades audiovisuais</i>	129
5.1.4.	<i>Efeitos da tarefa: Disparidades AV temporais, espaciais e espaço-temporais</i>	132
5.1.5.	<i>Apresentações audiovisuais e visuais unimodais: Efeitos comparativos</i>	135
5.1.6.	<i>Um teste à localização percebida do termo (offset) do movimento de estímulos auditivos</i>	139
5.2.	Discussão.....	145
5.2.1.	<i>Conclusões gerais</i>	145
5.2.2.	<i>Evidência de mecanismos guiados pelos dados (bottom-up)</i>	147
5.2.3.	<i>Mecanismos explicativos: Proposta de um quadro interpretativo do conjunto dos resultados</i>	148
5.2.4.	<i>Objeções e alternativas possíveis: Discussão argumentada</i>	151
5.2.5.	<i>Sumário provisório</i>	153
6.	Deslocamento O	155
6.1.	Resultados	155
6.1.1.	<i>Efeitos das tarefas (T, E, ET) e das condições modais (V, A e AV)</i>	156
6.1.2.	<i>Efeitos da distância percorrida/duração do trajeto do alvo: Comparação entre condições modais (AV, V, A)</i>	158
6.1.3.	<i>Efeitos do ponto de desaparecimento: Comparação entre condições modais (AV, V, A)</i>	162
6.1.4.	<i>Efeitos do começo (onset) e termo (offset) do estímulo auditivo nas condições audiovisuais (AV)</i>	165
6.1.5.	<i>Correlações entre o deslocamento M e o deslocamento O</i>	169
6.2.	Discussão.....	171
6.2.1.	<i>Semelhanças e diferenças entre a modalidade visual e auditiva no deslocamento O</i>	171
6.2.2.	<i>Mecanismos de integração da informação visual e auditiva no deslocamento O: Avaliação indicativa do modelo de máxima verosimilhança da integração multissensorial</i>	173
6.2.3.	<i>Semelhanças e diferenças entre o deslocamento O e o deslocamento M</i>	178
7.	Discussão geral e conclusão	181
7.1.	Principais resultados.....	181

7.2.	Implicações de natureza aplicada.....	186
7.3.	Limites do estudo.....	190
7.4.	Conclusão provisória.....	193
	Referências bibliográficas	195

INTRODUÇÃO

Vivemos num mundo multissensorial, cercados por múltiplas fontes de estimulação heterogéneas. Os nossos cérebros são continuamente inundados por informação que chega através de diferentes vias sensoriais e deve ser integrada de forma adequada (Shams, 2012; Wallace, Woynaroski, & Stevenson, 2020). Um único evento no ambiente produz na maioria das vezes sinais sensoriais distintos quer no interior de uma modalidade, quer entre as modalidades. De modo a obter uma representação multimodal coerente do mundo externo capaz de suportar a interação organismo-meio e as aprendizagens necessárias, o cérebro procura otimizar a sua perceção (Alais & Burr, 2019) através da exploração das redundâncias e complementaridades proporcionadas pelas diversas modalidades sensoriais (Alais, Newell, & Mamassian, 2010), designadamente a visão e a audição (Ernst & Bühlhoff, 2004).

Apesar da aparente facilidade com que transformamos constantemente subconjuntos de sinais em perceções contínuas (Noppeney, 2021), a interação-integração intermodal não é um problema de resolução trivial para o cérebro. Os estímulos no ambiente mobilizam diferentes formas de energia e ativam múltiplos sistemas sensoriais que processam e codificam aspetos complementares do estímulo, associados a diferentes formas de representação (Alais & Burr, 2019; Wallace et al., 2020). A informação a combinar a nível central procede de diferentes córtices primários, cuja arquitetura neural se tem revelado mais interativa e complexa do que admitido até recentemente, encontrando-se muitas vezes desfasada no tempo e codificada em sistemas de coordenadas heterogéneas (pelo menos inicialmente) (Alais & Burr, 2019). Contudo, apesar da complexidade do processo perceptivo, da multiplicidade dos fluxos sensoriais (veiculando informação incerta, ambígua e incompleta) e do risco dos cálculos e mecanismos envolvidos gerarem eles próprios interpretações ambíguas, o cérebro humano saudável “reconstrói” com surpreendente

sucesso o mundo exterior e as suas propriedades (Ernst & Bühlhoff, 2004; K. E. Jordan, Clark, & Mitroff, 2010; Wallace et al., 2020).

Por exemplo, quando estamos sentados num comboio a olhar pela janela e um comboio na linha ao lado começa a movimentar-se, é gerada uma situação de ambiguidade: qual o comboio que está em movimento? A desambiguação é uma forma de interação sensorial que o cérebro utiliza para formar uma estimativa perceptiva mais robusta. Se uma única modalidade se revelar insuficiente, podem ser combinadas informações de várias modalidades. No exemplo considerado, o cérebro pode recolher informação sobre o evento perceptivo a partir de uma modalidade sensorial distinta, como o sistema vestibular, de modo a desambiguar a situação e reter uma única interpretação final (seja certa, se contrariar no caso a ilusão de auto-movimento, ou errada). Procura dessa forma escolher rápida e eficazmente uma representação coerente entre as representações possíveis, evitando perpetuar a incerteza e viabilizando a interação em tempo real com o ambiente, o principal objetivo do sistema perceptivo (Ernst & Bühlhoff, 2004).

A variabilidade, incluindo uma parcela importante de ruído, é uma característica essencial da informação sensorial e das suas contrapartidas neurais. Reconhecê-lo equivale a aceitar que a perceção é um processo intrinsecamente probabilístico, envolvendo interpretações e inferências sobre a natureza provável dos estímulos externos (Alais & Burr, 2019). Para além da evidência sensorial, o conhecimento prévio, adquirido através da experiência de interação com o mundo, desempenha também um papel nas inferências perceptivas, contribuindo para a geração de estimativas sensoriais e para a resolução de ambiguidades no processo de construção de uma perceção multissensorial coerente (Alais & Burr, 2019; Ernst & Bühlhoff, 2004).

Apesar de nenhum sistema de processamento de informações, seja técnico-artificial ou biológico, ser suficientemente poderoso para “perceber e agir” com precisão e confiabilidade em todas as condições, a chave para uma perceção robusta reside, segundo Ernst & Bühlhoff (2004), na combinação-integração eficiente de múltiplas fontes de informação sensorial. A evidência disponível sugere que, longe de ocorrer de forma rígida, essa combinação obedece a regras flexíveis, dependentes da situação, permitindo maximizar a sua eficiência em função das circunstâncias (Burr & Alais, 2006). Diferentes estratégias de combinação e de integração são assim concebíveis, podendo em geral organizar-se em duas grandes classes: (1) maximizar a informação fornecida pelas diferentes modalidades sensoriais (combinação sensorial); (2) reduzir de forma ótima a variabilidade da estimativa sensorial, aumentando a sua confiabilidade (integração sensorial) (Ernst & Bühlhoff, 2004).

Através destas estratégias, orientadas para a exploração de complementaridades e redundâncias nos sinais sensoriais (e.g., as informações originadas por um evento comum encontram-se amiúde espacial e temporalmente correlacionadas), é possível obter inúmeros benefícios perceptivos e comportamentais. Uma representação multissensorial redundante oferece vantagens óbvias, evitando falhas catastróficas da percepção se as condições do ambiente tornarem uma das modalidades temporariamente ineficaz (e.g., a visão durante a noite; um evento fora do campo de visão; um som relevante mascarado por ruído de fundo) ou se um sentido for degradado/perdido permanentemente (Alais & Burr, 2019; Alais et al., 2010). Por sua vez, a vantagem estatística de dispor de duas amostras do mesmo estímulo, ou de integrar múltiplos estímulos, manifesta-se em benefícios documentados como tempos de reação mais rápidos, melhor detecção e discriminação de estímulos multissensoriais (comparativamente a estímulos unimodais) e aumento da precisão (menor dispersão) das respostas (Alais & Burr, 2019; Alais et al., 2010; Wallace et al., 2020). Combinar os pontos fortes complementares e superar as fragilidades individuais de cada modalidade sensorial (Noppeney, 2021) tem sido encarado neste sentido como uma estratégia evolutiva associada a vantagens substanciais, constituindo no caso de organismos fortemente dependentes de sistemas perceptivos desenvolvidos, um requisito crucial para a sobrevivência e adaptação, quer individual, quer específica (De Sá Teixeira & Oliveira, 2013; Godfroy-Cooper, Sandor, Miller, & Welch, 2015; Soto-Faraco, Lyons, Gazzaniga, Spence, & Kingstone, 2002; Wallace & Stevenson, 2014). Um exemplo saliente destes benefícios, envolvendo a combinação de informação auditiva e visual, é o ganho de inteligibilidade dos sinais de fala, na presença de ruído, quando apresentados num contexto audiovisual – em que, para além da voz, é visível a articulação dos sons pelo falante (Wallace & Stevenson, 2014; Wallace et al., 2020).

Apesar das vantagens assinaladas, a integração multissensorial não produz apenas benefícios. A investigação empírica tem consistentemente documentado dois tipos de resultados: (1) efeitos de melhoria multissensorial do desempenho: por exemplo, respostas mais rápidas e precisas a estímulos apresentados simultaneamente em diferentes modalidades sensoriais (ainda que alguns sejam irrelevantes para a tarefa) do que a cada um apresentado isoladamente (Tang, Wu, & Shen, 2016); (2) ilusões multissensoriais, resultantes da integração de informações incongruentes entre modalidades sensoriais (Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010; Tang et al., 2016). Estas ilusões ilustram uma propensão geral do sistema nervoso para combinar informações de diferentes modalidades. O que, em circunstâncias favoráveis (congruência intermodal) contribui para reduzir a ambiguidade do estímulo pode, em circunstâncias desfavoráveis (conflito

intermodal), ser indutor de ilusões (Talsma et al., 2010). Shams (2011) sugere assim que as interações intermodais correspondem a processos robustos, generalizados e predominantes no processamento perceptivo humano. Mesmo a percepção visual (modalidade que se acreditava dominante e largamente autónoma) pode ser dramaticamente influenciada (modulada ou “iludida”) pelo processamento intermodal.

O grau de integração perceptivo da informação multimodal pode variar, dependendo da consistência verificada entre os estímulos e da experiência anterior, desde uma integração completa (fusão perceptiva), passando por uma integração parcial, até à segregação completa das informações de diferentes modalidades (Shams, 2012). A ocorrência e resultados da interação-integração multissensorial é modulável tanto por variáveis associadas a níveis mais baixo de processamento (*low-level*) – e.g., características físicas dos estímulos, sincronia ou proximidade espacial entre sinais – como por variáveis associadas a níveis mais altos de processamento (*high-level*), envolvendo conhecimento, expectativas, detecção-aprendizagem de correlações ou associações (Wallace et al., 2020). A integração multissensorial apresenta uma tolerância mais reduzida a disparidades espaciais e/ou temporais entre sinais do que a interação intermodal (Schmiedchen, Freigang, Nitsche, & Rübsem, 2012). Quanto mais reduzidas forem essas disparidades, maior a probabilidade de que os diferentes sinais sejam combinados e integrados (Wallace et al., 2020).

Embora os benefícios perceptivos da interação-integração multissensorial se encontram bem estabelecidos, a forma como o cérebro implementa essa integração é atualmente uma das questões mais desafiantes da neurociência sensorial e cognitiva (Alais & Burr, 2019; Burr & Alais, 2006). Nas últimas duas décadas recrudescer o interesse pela investigação dos mecanismos através das quais entradas sensoriais de diferentes modalidades são combinadas e integradas pelo cérebro, modulam a percepção e influenciam o comportamento (Spence, 2015; Wallace & Stevenson, 2014). Uma massa cumulativa de estudos nas áreas da psicologia, psicofísica, neurociência cognitiva, neurofisiologia e imagiologia dedicados à investigação e modelização da interação e integração multissensorial, tanto em humanos como noutros animais, tem contribuído para uma compreensão crescente da generalidade e complexidade destes mecanismos, sem que, contudo, várias das questões fundamentais tenham até à data obtido resposta satisfatória, menos ainda consensual (Calvert, Spence, & Stein, 2004; Spence, 2013; Spence & Driver, 2004; Stein, 2012; Wallace et al., 2020).

Um outro domínio no qual se mantêm desafios importantes, envolvendo a cognição, a memória e a percepção, é o do processamento de estímulos ou eventos dinâmicos (Thornton

& Hubbard, 2002). Uma parte importante da nossa percepção e ação envolve o planeamento e implementação da nossa interação (e.g., evitamento, contacto, interseção) com os objetos, pessoas e outros animais. Estas interações ocorrem no quadro de eventos que se desenrolam no tempo, envolvendo por isso mesmo informações dinâmicas (sobre transformações/alterações) do meio ambiente (De Sá Teixeira & Oliveira, 2013). Do ponto de vista biológico, os eventos espaço-temporais envolvendo mudança e movimento, parecem constituir os estímulos mais importantes para o sistema nervoso: os *onset* (surgimento) e *offset* (extinção) dos estímulos, associados a eventos de transição, originam neste sentido respostas neuronais particularmente intensas. Em consonância com esta importância particular dos eventos dinâmicos (cuja deteção se afigura igualmente crucial para a sobrevivência dos organismos), propostas para o tratamento conjunto, e não separado, das suas dimensões temporais e espaciais (não sem um eco do movimento realizado nesse sentido pela própria física do espaço e do tempo) têm sido avançadas (Nijhawan & Khurana, 2010).

A noção de representação mental dinâmica, cunhada por J. Freyd (1987), refere-se justamente ao modo como o sistema cognitivo representa este tipo de informação sobre transições. De acordo com Freyd (1987), uma representação dinâmica caracteriza-se por incluir *intrinsecamente* uma representação analógica do tempo, incorporando as suas propriedades de continuidade e direcionalidade, de tal modo que a representação se altera ela própria no tempo, de forma regulada e previsível (Nijhawan & Khurana, 2010). Uma possibilidade admitida por Hubbard (1998) é a de que este tipo de representações assegure a internalização na nossa arquitetura cognitiva de regularidades físicas externas (e.g., momento, fricção, gravidade, força centrípeta ou centrífuga, entre outros) (Hubbard, 1995c, 2005), permitindo simular-antecipar internamente as consequências destes “invariantes ambientais” (Amorim, Siegler, Baurès, & Oliveira, 2015; Hubbard, 1995c). Estas simulações corresponderiam, em princípio, ao exercício de um conhecimento incorporado (largamente implícito) da dinâmica das forças atuantes no ambiente (White, 2012). Existem no entanto evidências tanto no sentido do envolvimento de mecanismos implícitos (Hubbard, 1998a) como de mecanismos explícitos, conscientes (Hubbard, 1995c, 2005, 2006). Um debate durável na investigação dos erros de localização de alvos dinâmicos, uma das áreas mais proeminentes do estudo das representações dinâmicas, refere-se exatamente ao papel respetivo nesses erros de mecanismos sensoriomotores de baixo nível (e.g., Kerzel, 2002b) e de processos cognitivos de alto nível (e.g., Hubbard & Favretto, 2003; Munger & Minchew, 2002; Vinson & Reed, 2002). Tem sido argumentado neste contexto que uma explicação

baseada num pequeno número de processos de alto nível seria mais económica do que explicações mobilizando numerosos processos de baixo nível, específicos para cada dimensão, formato ou modalidade de estímulo (Hubbard, 2005). No entanto, para além de se apresentar modulável por um amplo espectro de variáveis de diferente natureza (ver, e.g., Hubbard, 2005, 2014), a localização de alvos dinâmicos tem igualmente resistido a explicações assentes num único princípio ou mecanismo dominante (Hubbard, 2010, 2017).

O objetivo desta dissertação situa-se na confluência destes dois domínios, o da interação-integração multissensorial e o das representações dinâmicas. Especificamente, procura investigar o modo como pistas temporais e/ou espaciais auditivas influenciam a localização da última posição percebida de um alvo visual em movimento que desaparece súbita e imprevistamente. Este tipo de tarefa dá caracteristicamente lugar a um erro na localização do alvo para diante, no sentido do movimento, cunhado por Freyd & Finke (1984) como Momento Representacional (MR) (por analogia com o momento físico ou “quantidade de movimento”, correspondente à multiplicação da massa pela velocidade instantânea) e considerado por Freyd (1987) uma ilustração privilegiada da noção de representação dinâmica.

O estudo da interação-integração audiovisual no quadro de uma tarefa de MR torna possível examinar várias questões em aberto no domínio do processamento intermodal: saber se a audição é capaz de influenciar o processamento visual do movimento em tarefas de localização espacial (designadamente quando o alvo é foveado); em caso de resposta positiva, saber se tais efeitos devem ser temporalmente mediados ou podem decorrer de pistas auditivas espaciais; saber se a informação auditiva pode induzir melhorias audiovisuais da localização de alvos visuais móveis ou apenas efeitos de enviesamento perceptivo (a resposta a esta última questão é particularmente útil de um ponto de vista aplicado, contendo potenciais implicações para a reabilitação de défices visuais). A investigação de uma tarefa de MR audiovisual permite, por suas vez, examinar o fenómeno do MR num contexto distinto daquele em que tem sido quase exclusivamente estudado, o da modalidade visual (Freyd & Finke, 1984; Hubbard, 2010, 2018a; Kelly & Freyd, 1987). Do exame dos mecanismos através dos quais as propriedades do som modulam o MR podem esperar-se, por exemplo, contributos para debates como o da articulação entre processos de baixo e alto nível no MR ou o da extensão do envolvimento de eventuais processos amodais (transversais às modalidades). Aspetos do MR raramente considerados na literatura anterior, como o das relações entre os erros de localização para diante, no sentido do movimento (deslocamento M ou MR) e para baixo (deslocamento O, ortogonal, alinhado com

a direção da gravidade no caso de movimentos horizontais), podem igualmente ser abordados num contexto comparativo, envolvendo apresentações visuais, auditivas e audiovisuais, o mesmo sucedendo com os efeitos de variáveis como a distância percorrida pelo alvo ou o seu ponto de desaparecimento, apenas investigadas até à data por um reduzido número de estudos na modalidade visual.

A dissertação organiza-se em duas partes principais – “estado da arte” e “contribuições empíricas”. O estado da arte incluiu três capítulos (Capítulos 1 a 3). O primeiro capítulo, dedicado ao momento representacional, define-o e situa-o no contexto mais amplo dos erros de localização, caracteriza o paradigma operacional associado, passa em revista as principais teorias e modelos explicativos e identifica as variáveis moduladoras principais. Adicionalmente distingue e caracteriza os erros de localização no sentido do movimento (M) e ortogonal ao movimento (O), introduzindo a noção de gravidade representacional. Conclui com uma secção sobre o momento representacional auditivo, caracterizado nas suas semelhanças e diferenças relativamente ao MR visual. O segundo capítulo ocupa-se da interação-integração multimodal: revê os principais quadros interpretativos, apresenta e discute as regras temporais e espaciais da integração multissensorial, o estado da arte relativamente aos contributos da atenção intermodal (exógena e endógena), a hipótese de *assunção de unidade* (convicção de que dois sinais sensoriais procedem de um mesmo evento ou objeto) e a noção de percepto audiovisual. Por fim, o terceiro capítulo é dedicado à revisão dos resultados conhecidos sobre modulação da perceção visual do movimento pelas diferentes propriedades do som (temporais, espaciais, semânticas) e, em particular, da modulação do MR visual por pistas temporais e espaciais. O capítulo termina com a identificação de um conjunto de questões deixadas em aberto pelos estudos discutidos e retomadas na contribuição empírica da dissertação.

A contribuição empírica compreende por sua vez quatro capítulos (Capítulos 4 a 7). O primeiro desses capítulos explicita as hipóteses de partida e previsões, descreve a tarefa experimental de base, justificando as principais opções operacionais, e apresenta a secção geral de método (participantes, estímulos, desenho experimental, procedimento experimental e racional da análise de dados). O capítulo seguinte (Capítulo 5) apresenta e discute os resultados relativos ao erro de localização no sentido do movimento, deslocamento M, bem como os resultados de uma tarefa suplementar envolvendo a localização de alvos auditivos dinâmicos. No terceiro capítulo da contribuição empírica (Capítulo 6) são apresentados e discutidos os resultados relativos ao erro de localização ortogonal ao movimento (deslocamento O) e analisadas as relações observadas entre os dois

tipos de erro de localização. Finalmente, o último capítulo corresponde a uma discussão geral que articula os dois capítulos anteriores, sistematiza os principais resultados, extrai potenciais implicações aplicadas, discute limites do estudo e sinaliza direções de investigação futura.

PARTE I

ESTADO DA ARTE

CAPÍTULO I

MOMENTO REPRESENTACIONAL: UM ERRO DE LOCALIZAÇÃO DO MOVIMENTO

Num mundo que frequentemente se oferece à experiência sob a forma de eventos dinâmicos, a capacidade de localizar objetos em movimento constitui uma função essencial da percepção (Getzmann & Lewald, 2007). No entanto, o modo como o sistema visual integra sinais de movimento e sinais de posição na localização de um alvo móvel não é ainda claro (Kwon, Tadin, & Knill, 2015; Whitney, 2002). Um aspeto deste problema, objeto de interesse considerável nas últimas décadas, é a ocorrência sistemática de diversos tipos de erro na localização de objetos em movimento (ver, e.g., Hubbard, 2018b; Nijhawan & Khurana, 2010). A investigação destes erros de localização tem mobilizado e articulado questões fundamentais da percepção do movimento, do processamento da informação espacial e temporal, do ciclo percepção-ação, da percepção causal e da física ingénuas e, mais recentemente, da interação-integração multissensorial, entre outras (Hubbard, 2018b; Nijhawan & Khurana, 2010).

Entre os erros de localização mais populares contam-se o *efeito Fröhlich* (EF) (Fröhlich, 1924; Müsseler, Stork, & Kerzel, 2002), o *efeito flash-lag* (EFL) (Baldo & Klein, 2010; MacKay, 1958; Müsseler et al., 2002; Nijhawan, 1994) e o *momento representacional* (MR) (Freyd & Finke, 1984; Hubbard, 2010). O efeito Fröhlich indica a tendência para perceber o ponto de partida de um movimento (*onset*) como deslocado para diante, na direção do movimento (Müsseler et al., 2002). O momento representacional refere-se a uma tendência para localizar a última posição percebida de um alvo em movimento à frente do seu ponto de desaparecimento (*offset*), no sentido do movimento (Freyd & Finke, 1984). Apesar

de ambos os efeitos ilustrarem deslocamentos da posição percebida no sentido do movimento, existem igualmente erros no sentido oposto. Menos estudado do que os anteriores, o *erro de repulsão da origem* (*onset repulsion effect*: ERO) ilustra justamente uma tendência para recordar o ponto de partida do movimento como recuado relativamente à sua localização real (no sentido contrário ao do movimento) (Hubbard & Motes, 2002; Thornton, 2002). O *efeito Fröhlich* e o ERO ocorrem em situações distintas (com diferentes gamas de velocidades e diferentes formas de aparição do alvo), sendo a caracterização das condições diferenciais para a ocorrência de um e de outro um tópico de investigação em aberto. O *efeito flash-lag* envolve por sua vez, na sua versão clássica, a localização de um alvo em movimento num ponto da sua trajetória assinalado temporalmente pela ocorrência de um *flash* espacialmente alinhado com o alvo. Tipicamente, o *flash* é percebido como localizado atrás do alvo (*flash-lag*), ou, o que é equivalente, tratando-se de uma localização relativa, o alvo é perceptivamente localizado à frente do *flash* no sentido do movimento. Menos estudado do que o *flash-lag*, um efeito de arrastamento da posição percebida do *flash* no sentido do movimento do alvo (*flash-drag*: Murai & Murakami, 2016) foi também documentado. Vários destes erros de localização foram entretanto observados e estudados noutras modalidades, como a audição (Alais & Burr, 2003; Getzmann, 2005a; Getzmann, Lewald, & Guski, 2004) ou o tato (Drewing, Hitzel, & Scocchia, 2018; Merz, Meyerhoff, Spence, & Frings, 2019).

Sugestões sobre a possibilidade de unificar vários destes erros, particularmente os que envolvem deslocamentos da posição percebida no sentido do movimento, num quadro explicativo comum têm sido avançadas. Esforços nesse sentido incluem a investigação de Munger & Owens (2004) sobre as relações entre o momento representacional e o efeito *flash-lag*, o estudo comparativo de Müsseler et al. (2002) envolvendo o EF, o EFL e o MR, e estudo de Hubbard (2008) sobre a contribuição do MR para a indução de erros de localização de um objeto estático adjacente (captura pelo movimento). Inventários comparativos sistemáticos das semelhanças e diferenças no modo como o EFL e o MR, em particular, são modulados por diferentes variáveis, foram também publicados por Hubbard (2013; 2014b), servindo de base a propostas de possíveis mecanismos comuns, que se mantêm todavia conjeturais.

Talvez com exceção do efeito *flash-lag*, o MR constitui o erro de localização do movimento mais bem caracterizado do ponto de vista da investigação (ver adiante, secção 1.2). Do ponto de vista explicativo, tem resistido a interpretações em termos de um único mecanismo ou teoria (Hubbard, 2010, 2017). Com base na evidência disponível, é igualmente claro que tanto processos sensoriais e perceptivos de baixo nível (e.g., persistência retiniana,

enviesamento foveal) como processos descendentes, de alto nível (e.g., conhecimento sobre o alvo) modulam a sua ocorrência e magnitude (Hubbard, 2005, 2010, 2014a, 2017, 2018a). Do ponto de vista operacional, a presente investigação assenta numa tarefa de momento representacional, procurando apreciar os efeitos das propriedades espaciais e temporais de estímulos auditivos na modulação do MR visual. O presente capítulo destina-se a apresentar o paradigma do momento representacional, identificar algumas das principais variáveis moduladores deste erro de localização e sumariar resultados obtidos anteriormente com tarefas de MR, com destaque para os que se afiguram mais relevantes para a presente dissertação.

1.1. DEFINIÇÃO, PROPOSTAS DE EXPLICAÇÃO E PARADIGMA OPERACIONAL

1.1.1. Definição e propostas de explicação

O momento representacional (MR) refere-se à tendência sistemática dos observadores para recordarem a última posição percebida de um alvo móvel que imprevisivelmente desaparece como situada à frente do ponto de desaparecimento (Thornton & Hubbard, 2002). Este fenómeno de desfasamento mnésico foi inicialmente documentado por Freyd & Finke (1984), que assinalaram também a sua semelhança com o momento físico. O momento (ou quantidade de movimento) de um objeto é igual ao produto da sua massa pela velocidade, podendo o MR conceber-se em primeira análise, na ótica de Freyd & Finke (1984), como um análogo mental desta quantidade (Hubbard, 2010). Tal como um objeto físico em movimento não se detém instantaneamente quando uma força de resistência é aplicada (e.g., a ação dos travões num carro em movimento), também a representação mental de um objeto em movimento incorporaria um equivalente de momento, não interrompendo instantaneamente o movimento representado (Hubbard, 1995b).

O momento representacional constitui um fenómeno robusto, observável não apenas com movimento real e contínuo aparente (transições de localizações percebidas como contínuas; e.g., animações vídeo), mas também com movimento implícito (sequências de estímulos discretos, percebidos como espacial e temporalmente separados, sugerindo uma trajetória de movimento) e mesmo sugerido por imagens ou fotografias (instantâneos) ilustrando momentos de um acontecimento/ação dinâmica (Freyd, 1983; Freyd & Finke,

1984; Futterweit & Beilin, 1994; Hubbard & Bharucha, 1988). Beneficia também de uma grande generalidade, tendo sido documentado com vídeos de cenas complexas, como o movimento de multidões (Thornton & Hayes, 2004), alterações do ponto de vista de uma cena (Munger et al., 2006), movimentos de orientação de faces (Hudson, Liu, & Jellema, 2009), expressões faciais dinâmicas (Prigent, Amorim, & de Oliveira, 2018; Uono, Sato, & Toichi, 2014), representações de movimento biológico (pontos luminosos) (Verfaillie, De Troy, & Van Rensbergen, 1994), figuras humanas animadas (Verfaillie & Daems, 2002), etc. (para uma revisão, ver Hubbard, 2018a).

Uma ambiguidade da definição do momento representacional envolve a confusão possível entre o erro de localização para diante, no sentido do movimento, e a componente do deslocamento mnésico consistente com o modo como o momento físico influencia um objeto físico. Na realidade, num movimento vertical descendente ou num movimento oblíquo descendente (num espaço submetido à ação da gravidade), por exemplo, uma parte do deslocamento na direção do movimento seria atribuível à ação da gravidade, e não ao momento físico (Hubbard, 1995c, 2010). Este facto ilustra uma confusão entre um uso descritivo da noção de momento representacional, para fazer referência ao deslocamento observado na direção do movimento e o uso explicativo da noção como referindo-se ao impacto na localização percebida de um análogo mental do momento físico. O problema será retomado e discutido a propósito da distinção entre momento representacional e gravidade representacional, na secção 1.3 deste capítulo).

A ideia que o momento representacional associa propriedades do mundo e propriedades das representações rapidamente excedeu a estrita analogia com o momento físico. Um papel geral deste tipo de deslocamentos na compensação dos atrasos neurais (e.g., tempos associados ao processamento neural) no circuito percepção-ação tem sido admitido (Hubbard, 2005, 2006; Nijhawan & Khurana, 2010). No momento em que percepção de um alvo em movimento é gerada, após uma cascata de processamentos a partir da entrada sensorial, a posição ocupada pelo alvo encontra-se já à frente da posição percebida. O momento representacional, envolvendo uma deslocação da posição percebida para diante, poderia por conseguinte repor o alinhamento posicional com o alvo (realinhar a percepção com o presente: Changizi, Hsieh, Nijhawan, Kanai, & Shimojo, 2010) contribuindo desse modo para suportar, por exemplo, ações bem-sucedidas de interseção e captura (Hubbard, 2006). Tratando-se de um erro de localização absoluta, e não relativa, como o *flash-lag* (erro de localização do *flash* relativamente a um alvo móvel, ou vice-versa), o momento representacional poderia mesmo assegurar uma melhor compensação (absoluta) dos atrasos

neurais do que o *efeito flash-lag*, classicamente interpretado como o resultado de um mecanismo extrapolatório de compensação (Hubbard, 2018a; Nijhawan, 1994, 2008).

A consistência desta extrapolação compensatória com o modo como os princípios físicos influenciam um objeto físico (exemplificada na analogia com o momento), foi igualmente objeto de alargamentos. A possibilidade de que o momento representacional refletisse uma propriedade geral de coerência espaço-temporal entre o mundo físico externo e as representações mentais internas (e não uma internalização específica do momento físico) foi cedo admitida por Freyd, (1987; 1993). O traço essencial destas representações, ao qual devem a sua coerência, é serem dinâmicas, implicando, na definição de Freyd (1987), que incorporam analogicamente os aspetos temporais da informação, como a direcionalidade e a continuidade. Longe de se limitar à representação do momento físico e do movimento espacial, o deslocamento mnésico associado ao momento representacional poderia assim ocorrer com qualquer dimensão estimular capaz de variar continuamente (de forma coerente). Algumas dimensões capazes de variação contínua, como a luminância crescente ou decrescente de um estímulo, mostraram no entanto não dar lugar ao deslocamento para diante característico do momento representacional, e sim a um deslocamento de sentido oposto (Brehaut & Tipper, 1996).

Outras formas de expansão da consistência do momento representacional com o modo de ação dos princípios físicos foram avançadas, menos gerais (mais constrangidas) do que a mera consistência espaço-temporal. Hubbard (1998a; 1998b; 2005; 2006) sugeriu neste sentido a incorporação no sistema representacional de conhecimento implícito relativo a diferentes invariantes físicos ambientais (e.g., gravidade, fricção, força centrípeta), para além do momento, todos eles capazes de, através das propriedades dinâmicas da representação, produzir distorções mnésicas (deslocamentos ou erros de localização) alinhadas com a ação do princípio invariante. Uma alteração importante desta proposta relativamente à ideia de coerências espaço-temporal é que estes conhecimentos implícitos não têm que refletir exatamente a ação do invariante, mas as suas consequências subjetivas (Hubbard, 2006). Por exemplo, um invariante físico relativo aos efeitos da massa do objeto é percebido pelas suas consequências percebidas enquanto peso (num ambiente com gravidade): efeitos da massa do objeto no momento representacional são observados com movimentos verticais (alinhados com a direção da gravidade) mas não horizontais, sugerindo a representação de um efeito subjetivamente experimentado do peso e não do efeito objetivo da massa (Cooper & Munger, 1993; Hubbard, 1997).

Hubbard (2005; 2006; 2010; 2018a) recupera neste sentido a noção de isomorfismo de segunda ordem das representações mentais, introduzida por Shepard & Chipman (1970), para caracterizar os mecanismos subjacentes a estes deslocamentos. Enquanto num isomorfismo de primeira ordem as propriedades de um estímulo físico se encontram literalmente na representação (e.g., a cor ou a forma do estímulo são a cor e a forma da representação), num isomorfismo de segunda ordem as propriedades do estímulo são preservadas apenas de forma abstrata e funcional (e.g., uma informação de distância pode ser representada por uma sequência temporal). A correspondência entre representações internas e as consequências subjetivas dos invariantes físicos seria deste segundo tipo, permeável, designadamente, à modulação por expectativas do observador ou circunstâncias contextuais. Uma proposta integrativa de Hubbard (2006), visando dar conta, simultaneamente, da consistência do momento representacional com as consequências subjetivas de invariantes físicos e do papel evidenciado pela investigação de numerosas variáveis estranhas a esses invariantes (e.g., expectativas variadas do observador e informações contextuais: ver abaixo, secção 1.2), envolveu a articulação de dois fatores. O primeiro desses fatores refletiria o isomorfismo de segunda ordem com as consequências subjetivas dos invariantes e, na ausência de outros determinantes contextuais, determinaria a direção e magnitude do deslocamento (o erro de localização). O segundo, correspondendo à presença de elementos variáveis de contexto, modularia os efeitos do primeiro, exprimindo o deslocamento observado nesse caso uma combinação da ação dos princípios físicos invariantes e do contexto.

A principal oposição a uma explicação do deslocamento associado ao MR através da internalização de princípios físicos, conhecimentos implícitos ou extrapolação perceptiva antecipatória, resultou de teorias perceptivas de “baixo nível” (excluindo memória, expectativas, experiência anterior), assentes em propriedades do comportamento oculomotor e formas de processamento inicial da informação retiniana. O ponto de partida empírico para estas propostas foi a verificação de um cancelamento do MR quando os movimentos oculares de seguimento de um alvo em movimento contínuo eram impedidos (Kerzel, 2000, 2002c, 2002b). Com base na observação de que estes movimentos se estendem para além da posição final do alvo (*overshoot*), arrastando a imagem persistente (persistência retiniana) na direção do movimento (Kerzel, 2000; Kerzel, Jordan, & Müsseler, 2001), o desvanecimento do erro de localização foi interpretado como significando que se trata de um resultado do comportamento oculomotor. Para além de alguma evidência contrária (referida na secção 1.2 deste capítulo), um limite importante da proposta é o facto

de se circunscrever a situações de movimento contínuo (a interdição de movimentos oculares não afeta os deslocamentos obtidos a partir de instantâneos fotográficos ou em situação de movimento implícito: Kerzel, 2002a), oferecendo assim quanto muito uma teoria local. Igualmente local é uma segunda proposta explicativa de “baixo nível”, avançada por Bertamini (2002), baseada em efeitos de adaptação perceptiva, que tem apenas aplicação no contexto de um dos métodos de resposta utilizados em paradigmas de MR (o método da sonda mnésica, descrito em 1.1.2 e na secção 1.2 deste capítulo) (ver Hubbard, 2010).

Uma versão mais geral de uma teoria essencialmente perceptiva e motora deste tipo de deslocamento da posição percebida foi igualmente proposta por Kerzel (2005; 2006), articulando três fatores. Kerzel fez convergir para esta teoria outros resultados para além dos relativos aos efeitos do movimento ocular, designadamente evidência sobre a importância da natureza da resposta, perceptiva *versus* motora. Mesmo na ausência de movimentos oculares, Müsseler et al. (2002) encontraram um erro de localização para diante com movimento contínuo. No entanto, diferentemente de outros estudos nos quais o erro se desvaneceu, a resposta neste caso era motora, envolvendo o posicionamento de um cursor. A existência de deslocamento para diante com respostas motoras de localização (tocar no local identificado) e a ausência de erro com julgamentos de localização, expressos verbalmente, foi também documentada por Kerzel & Gegenfurtner (2003), que concluíram por uma base sensoriomotora da extrapolação para diante, associada à conversão das coordenadas retinotópicas do alvo em coordenadas egocêntricas utilizadas para a ação de tocar. Na teoria dos três fatores, de Kerzel, o deslocamento é determinado pela (a) presença ou ausência de movimentos oculares de seguimento, (b) o tipo de movimento (percebido como contínuo ou como descontínuo, implícito) e (c) a natureza da resposta (verbal ou motora). Os movimentos oculares contribuem para o erro de localização para diante no caso de movimentos contínuos, mas não descontínuos, e a resposta motora contribui para o deslocamento em qualquer dos tipos de movimento. Apesar de mais geral do que uma estrita explicação oculomotora, esta proposta continua a não se aplicar, por exemplo, aos deslocamentos para diante com base em instantâneos fotográficos ou a tipos de deslocamento tipicamente associados ao movimento de alvos na horizontal, como um erro de localização para baixo (no eixo ortogonal ao do movimento).

Na sua classificação das diferentes teorias explicativas propostas para o MR, Hubbard (2010) agrupa-as em cinco categorias (teorias de internalização, teorias baseadas em crenças, teorias de neo-internalização, teorias de baixo nível e modelos de redes), cada uma das quais compreendendo diferentes variantes. Através deste espectro largo de propostas, a

oposição entre teorias da internalização ou da correspondência representativa com princípios físicos e as teorias mais estritamente perceptivo-motoras mantém o seu valor organizador, convocando de forma parcial (não existe sobreposição geral entre os dois tipos de oposição) a diferença entre abordagens *top-down* e *bottom-up*. Várias das teorias consideradas em Hubbard (2010) são propostas de implementação (caso dos modelos de rede) e a distinção entre várias outras é relevante apenas de pontos de vista específicos, como o de um interesse pelas eventuais relações entre o MR e a física ingênua (McCloskey, Washburn, & Felch, 1983), periférico no quadro da presente dissertação. O ponto essencial, sublinhado por Hubbard (2010) no termo da sua revisão, é que nenhuma teoria ou combinação de teorias do MR se mostrou até à data capaz de integrar o conjunto dos dados disponíveis.

1.1.2. Paradigma operacional

Historicamente, o primeiro paradigma de momento representacional (Freyd & Finke, 1984) consistiu em apresentar aos participantes uma sucessão de três retângulos concêntricos (estímulos indutores) separados por intervalos de 250 ms, com diferentes orientações, sugerindo um movimento de rotação no sentido horário ou anti-horário. Após a desapareição do terceiro retângulo, um quarto retângulo era apresentado (sonda), cuja orientação poderia corresponder a uma rotação idêntica, um pouco menor, ou um pouco maior do que a do último retângulo indutor. Os participantes avaliavam se a orientação da sonda era a mesma ou diferente da do terceiro retângulo (em variantes posteriores do método, decidir se a sonda se encontrava mais à esquerda ou à direita da última posição do alvo foi também utilizado, embora mais raramente: Kerzel, 2000). Tipicamente, a probabilidade de um participante responder “mesma” era maior para sondas com um ângulo de rotação maior do que o do último indutor.

O movimento envolvido neste primeiro tipo de paradigma, dito de “sonda mnésica”, era assim implícito, induzido por estímulos discretos. A trajetória sugerida era usualmente de rotação ou circular. A quantificação dos resultados realizava-se a partir das percentagens de resposta “mesma” dada a cada sonda para cada uma das orientações/posições do último indutor. Duas formas comuns de obter uma estimativa do erro para diante consistiam em: (1) calcular o ponto de igualdade subjetiva, correspondente (no caso de respostas mesmo-diferente) à orientação da sonda com a mais elevada percentagem de respostas “mesma”, e

medir a diferença (e.g., em graus) relativamente à orientação/posição do indutor (Kerzel, 2003a); (2) calcular a soma dos produtos da percentagem de respostas “mesma” de cada sonda pela diferença entre a posição da sonda e o indutor, e dividi-la pelo somatório das percentagens obtidas no conjunto das sondas (Freyd & Jones, 1994; Hubbard, 1993; Munger, Solberg, & Horrocks, 1999). Este índice corresponde a uma média ponderada dos desfasamentos de cada sonda. Um valor positivo assinala uma distorção mnésica para diante e um valor negativo uma distorção mnésica para trás.

Um paradigma operacional distinto foi introduzido por Hubbard & Bharucha (1988), que contribuiu para um segundo fôlego na investigação do MR. Hubbard & Bharucha (1988) apresentaram aos sujeitos alvos em movimento horizontal ou vertical contínuo (sem que fossem percebidas descontinuidades; e.g., animações vídeo) e desapareciam imprevisivelmente. Os participantes localizavam a última posição percebida do alvo através do posicionamento de um cursor controlado por um rato de computador. A medida do deslocamento era obtida pela diferença, em pixéis, entre a coordenada-x (no caso de movimento horizontal) indicada pelo participante e a coordenada-x correspondente ao ponto de desaparecimento. Podem assim assinalar-se três diferenças importantes relativamente ao paradigma da sonda mnésica: (1) o tipo de movimento utilizado era contínuo; (2) a trajetória do movimento era linear (vertical ou horizontal) e não rotacional; (3) os participantes respondiam através do posicionamento de um cursor no local onde perceberam o alvo pela última vez. Outras alternativas de resposta de localização motora têm sido usadas, como o apontar direto ou mediado por ponteiro num ecrã de toque (Kerzel & Gegenfurtner, 2003).

Entre as vantagens comparativas deste segundo paradigma contam-se: (1) uma medida mais simples do deslocamento, com precisão aumentada (ao nível do pixel, no caso do posicionamento de um cursor), (2) o acesso a uma avaliação direta da variabilidade das respostas, (3) uma redução significativa do número de replicações dos ensaios (requeridas para a obtenção de percentagens de resposta) e da necessidade de trabalho piloto (seleção do número e espaçamento das sondas) e (4) a sensibilidade a outros erros de localização do alvo, por exemplo, na direção ortogonal ao movimento (para baixo, no caso do movimento horizontal). As vantagens do método da sonda incluem, por sua vez, um controlo mais estrito do intervalo temporal entre a desaparecimento do alvo e a resposta e, para alguns, a minimização da contribuição de componentes motores para o deslocamento, preservando a sua natureza perceptiva (Hubbard, 2005). O método da sonda pode ainda revelar-se mais adequado a dimensões estimulares assentes em contínuos de frequência ou de intensidade (e.g., altura

tonal e brilho), nos quais uma medida de localização espacial pode revelar-se pouco adaptada (Hubbard, 2018a).

1.2. VARIÁVEIS RELEVANTES

Enquanto o debate sobre os mecanismos subjacentes ao MR se mantém aberto, a dependência do MR relativamente a um conjunto muito amplo de variáveis encontra-se pelo contrário bem estabelecida (Schmiedchen, Freigang, Rübsamen, & Richter, 2013). Ao longo de várias revisões dos resultados acumulados sobre os efeitos destas variáveis moduladoras, de natureza muito heterogénea, Hubbard (2005; 2010; 2014a; 2018a) optou por organizá-las em quatro categorias: (1) características do alvo (uma das áreas mais investigadas na literatura do MR), (2) características da apresentação, (3) características do contexto e (4) características relacionadas com o observador (classicamente menos exploradas, mas objeto de um recrudescimento de interesse recente). Embora mantendo as mesmas categorias, novas variáveis foram sendo acrescentadas em sucessivas revisões, pelo que nos referiremos, no essencial, à última revisão em data (Hubbard, 2018a). As variáveis consideradas apresentam entretanto um valor muito desigual, quer do ponto de vista da sua contribuição para a literatura do momento representacional, quer do ponto de vista da relevância para a presente dissertação. A estratégia seguida nesta secção consistirá assim, primeiro, em ilustrar genericamente o espectro das variáveis incluídas em cada uma das categorias, limitando em seguida a atenção a um número reduzido delas, em função da sua importância para a compreensão do momento representacional e/ou no contexto do presente trabalho.

1.2.1. Características do alvo

Incluem a velocidade, distância percorrida, direção, forma, identidade e categoria semântica do alvo, tamanho, excentricidade, modalidade, entre outras (ver Hubbard, 2014a, 2018a).

(i) A *velocidade* constitui uma das variáveis moduladoras fundamentais do MR, diretamente envolvida no estabelecimento da analogia com o momento físico (que aumenta com a velocidade) e com efeitos robustamente estabelecidos. A magnitude do MR visual

aumenta com a velocidade do alvo, no caso de velocidades constantes (e.g., Finke & Freyd, 1985; Finke, Freyd, & Shyi, 1986; Freyd, Kelly, & DeKay, 1990; Hubbard, 1990, 1995b; Hubbard & Bharucha, 1988; Hubbard & Ruppel, 2002; Kerzel et al., 2001), reduz-se quando o alvo apresenta velocidade irregular (Getzmann & Lewald, 2009), é modulada pela velocidade instantânea do alvo no final (maior ou menor quando são introduzidas acelerações-desacelerações) mas não no início ou durante a trajetória (Actis-Grosso, Bastianelli, & Stucchi, 2008). A utilização de alvos que se deslocam com velocidades constantes constitui assim uma condição geral favorável à produção de erros de localização do alvo para diante.

(ii) A *distância percorrida pelo alvo* é uma variável manipulada no quadro da presente dissertação. Na maioria das investigações sobre MR não foi incluída como um fator experimental. Quando tal sucedeu, tendeu a ilustrar um decréscimo do MR com distâncias maiores (H. Choi & Scholl, 2006; Hubbard & Ruppel, 2002). No entanto, encontrou-se invariavelmente confundida nesses casos com o ponto de desaparecimento do alvo. Num estudo em que estas duas variáveis (ponto de desaparecimento e distância percorrida) foram ortogonalmente manipuladas, de modo a dissociar experimentalmente os seus efeitos, apenas o ponto de desaparecimento, não a distância percorrida, apresentou efeitos (De Sá Teixeira & Oliveira, 2011). Resultados mistos foram reportados por McGeorge, Beschin, & Della Sala (2006), que observaram uma diminuição do MR com a distância percorrida em pacientes com heminegligência e compromisso do hemisfério direito e efeitos nulos numa amostra de participantes controlo. Em geral, pode concluir-se que o papel da distância percorrida pelo MR se encontra largamente por esclarecer.

(iii) Quanto à *direção do movimento*, o erro de localização para diante é maior para alvos em movimento horizontal do que em movimento vertical (Hubbard, 1990) e, neste último caso, para movimentos descendentes comparativamente a movimentos ascendentes (Hubbard, 1990; Munger & Owens, 2004). Relativamente à direção esquerda-direita *versus* direita-esquerda, não existem resultados consistentes em favor de uma diferença. Tipicamente, não têm sido documentadas diferenças significativas (e.g., Chien, Ono, & Watanabe, 2013; Cooper & Munger, 1993; Hubbard, 1990; Teramoto, Hidaka, Gyoba, & Suzuki, 2010). Nos casos em que uma diferença foi reportada (e.g., Halpern & Kelly, 1993), o movimento esquerda-direita deu lugar a magnitudes de deslocamento superiores. Em geral, a utilização de uma só direção de movimento horizontal (em particular, da esquerda para a direita) em tarefas de MR visual parece configurar uma estratégia aceitável de redução do número de condições experimentais, não acarretando à partida perdas significativas de informação.

As restantes variáveis mencionadas nesta categoria são, no caso vertente, menos relevantes. Os efeitos da forma, identidade e categoria do alvo incluem ilustrações importantes de um papel do conhecimento (incluindo conceptual) na modulação do MR (Kelly & Freyd, 1987; Nagai & Yagi, 2001; Vinson & Reed, 2002) e suportam a sua penetrabilidade cognitiva (Finke & Freyd, 1989); os efeitos da excentricidade do ponto de desaparecimento ocorrem em condições de interdição de movimentos oculares, que impedem a manutenção do alvo em visão foveal, traduzindo-se num aumento do MR em localizações mais periféricas (até um dado valor de excentricidade: Schmiedchen et al., 2013); o tamanho do alvo (associado à sugestão de massa ou peso) deu lugar a resultados inconsistentes, com estudos a documentar quer uma ausência de efeitos no erro de localização para diante (Cooper & Munger, 1993; Hubbard, 1997), quer efeitos crescentes da magnitude do erro com o tamanho (De Sá Teixeira, Oliveira, & Amorim, 2010; De Sá Teixeira, Oliveira, & Viegas, 2008). Esta inconsistência teve um papel na problematização da analogia com o momento físico (massa \times velocidade), que levaria a prever um efeito robusto da massa. Quanto à modalidade, vários estudos documentaram a ocorrência de momento representacional nas modalidades auditiva (Getzmann, 2005a; Getzmann & Lewald, 2009; Getzmann et al., 2004; Schmiedchen et al., 2012) e tátil (Brouwer, Franz, & Thornton, 2004; Merz, Meyerhoff, Frings, & Spence, 2020; Merz et al., 2019). Os estudos de MR auditivo serão separadamente apresentados e discutidos adiante, na secção 1.4 deste capítulo.

1.2.2. Características da apresentação

Incluem o tipo de movimento do alvo, o tipo de resposta dos participantes, o controlo do movimento pelo participante, o intervalo de retenção (atraso imposto entre a desaparecimento do alvo e a resposta), o contraste entre o alvo e o fundo, entre outras (ver Hubbard, 2014a, 2018a).

(i) Dois tipos principais de *movimento do alvo* têm sido utilizados: movimento implícito, utilizando estímulos discretos (claramente separados no espaço e no tempo) indutores cuja sequência sugere um movimento ao longo de uma trajetória (Freyd & Finke, 1984) e movimento contínuo (*smooth continuous*) no qual as alterações de posição do alvo são percebidas sem a aparência de descontinuidades (analogicamente) (Hubbard & Bharucha, 1988). A comparação entre efeitos dos dois tipos de movimento é dificultada pelo facto de as tarefas nas quais são utilizados tipicamente diferirem também noutras

características (Hubbard, 2018a). Resultados apontando seja para a ausência de diferenças (Munger & Owens, 2004), para maior MR com movimento implícito em trajetórias circulares (Kerzel, 2003c) ou ainda para maior MR com movimento contínuo em trajetórias lineares (Faust, 1990), têm sido reportados. Poljanšek (2002) observou por sua vez maiores efeitos da aceleração e desaceleração do alvo com movimento contínuo. A modulação do MR pelo tipo do movimento não parece assim obedecer a um padrão simples. Para além destas duas formas de apresentação de movimento, têm também sido utilizados instantâneos fotográficos que representam um momento particular de uma sequência de movimento (e.g., pessoa que salta ou caminha) (Freyd, 1983; Futterweit & Beilin, 1994). Evidência de que fotos instantâneas de eventos dinâmicos ativam o processamento do movimento tem sido fornecida (e.g., Kourtzi & Kanwisher, 2000; Senior et al., 2000). Esta última forma de apresentação tem contudo pouca expressão nos estudos de MR. Enquanto uma redução ou cancelamento do erro para diante com a exigência de fixação ocular num ponto tem sido reportada com movimento contínuo, o MR obtido com movimento implícito ou sugerido por instantâneos fotográficos não é influenciado pelos movimentos oculares (Hubbard, 2014a).

(ii) Três métodos predominantes de *resposta de localização* têm sido utilizados nos estudos de MR. No método da sonda mnésica os participantes avaliam se um estímulo idêntico ao alvo (sonda), apresentado depois deste último desaparecer, tem a mesma orientação (no caso de movimentos de rotação) ou a mesma localização que o alvo na sua última posição percebida (Freyd & Finke, 1984). No método de posicionamento de cursor, os participantes indicam o local onde o alvo desapareceu por meio de um cursor (controlado, e.g., por um rato de computador) (Hubbard, 1990; Hubbard & Bharucha, 1988). No método do apontar ou tocar, o participante toca diretamente (com o dedo), ou de forma mediada por um ponteiro o local da última posição percebida do alvo (Ashida, 2004; Kerzel & Gegenfurtner, 2003). O método da sonda mnésica, envolvendo um julgamento perceptivo, produz magnitudes de deslocamento inferiores aos restantes dois métodos, que envolvem uma resposta motora (Kerzel, 2003c; Kerzel & Gegenfurtner, 2003). O erro de localização é por sua vez maior com o método do tocar quando as mãos não se encontram visíveis do que com o posicionamento por cursor (Ashida, 2004). Uma vantagem dos métodos de localização motora sobre o método da sonda mnésica é a necessidade de muito menos ensaios (do ponto de vista psicofísico, a localização motora equivale a um método de ajustamento, enquanto a sonda mnésica, implicando uma variação da posição das sondas e a replicação de ensaios com cada posição, corresponde ao método constante). O método de sonda permite por sua vez um melhor controlo sobre o intervalo de tempo entre a desapareição do alvo e a resposta.

Medidas de localização mais precisas (ao nível do pixel) podem ser obtidas pelo posicionamento de cursor. No entanto, a resolução espacial do método do tocar é tipicamente inferior à dos outros dois métodos (Hubbard, 2018a).

As restantes variáveis mencionadas na categoria têm um papel reduzido no contexto da presente dissertação. A possibilidade de o participante controlar a velocidade, a direção e o movimento de desaparecimento do alvo resulta numa redução do MR (J. S. Jordan & Knoblich, 2004; J. S. Jordan, Stork, Knuf, Kerzel, & Musseler, 2002); por outro lado, participantes que puderam previamente controlar estas características do alvo ou observaram alguém a controlá-las, apresentam maior MR em tarefas onde esse controlo deixa de ser possível (J. S. Jordan & Hunsinger, 2008). Estes resultados indicam um papel importante do planeamento da ação na produção do MR, e contribuem reciprocamente para inscrever o MR nos mecanismos de controlo associados ao circuito percepção-ação (Hommel, Müsseler, Aschersleben, & Prinz, 2001; Prinz, 1997). A introdução de um atraso imposto entre o momento do desvanecimento do alvo e a resposta do participante visa examinar o curso temporal do MR e tem revelado resultados variáveis (ver revisão em Hubbard, 2005; 2018a). Em geral, a magnitude do erro para diante aumenta durante as primeiras centenas de milissegundos do intervalo até um máximo (pico), sendo que vários estudos apontam para uma estabilização dos valores do MR depois do pico (assíntota) (Finke & Freyd, 1985; Kerzel, 2000) e outros para uma redução progressiva do MR após o pico (De Sá Teixeira, Hecht, & Oliveira, 2013; Freyd & Johnson, 1987). Quanto ao contraste (de luminância) entre o alvo e o fundo, apesar de alguns resultados divergentes para diferentes tipos de movimento (implícito *versus* contínuo), contrastes mais intensos, ou crescentes ao longo do movimento, entre o alvo e o fundo dão lugar, por regra, a maiores magnitudes de MR (Hubbard & Ruppel, 2014; Maus & Nijhawan, 2009).

1.2.3. Características do contexto

Incluem interações com estímulos não-alvo (designadamente noutras modalidades), o enquadramento circundante, marcas-referências espaciais locais (*landmarks*), sombras e sombreamentos, expectativas quanto a trajetórias possíveis induzidas por elementos da situação (e.g., barreiras ao movimento, padrões de movimento oscilatório/cíclico), atribuições relativas à fonte de movimento (e.g., causalidade *versus* propulsão autónoma, ou animacidade), entre outras (ver Hubbard, 2014a, 2018a).

A *interação intermodal com estímulos não-alvo* constitui aqui, face aos propósitos da dissertação, a principal variável de interesse. Do mesmo modo que estímulos contextuais na mesma modalidade do alvo podem modular a magnitude do erro de localização (e.g., a representação de uma superfície estática sob o alvo em movimento produz reduções do erro para diante: Hubbard, 1995b), estímulos contextuais noutra modalidade podem também fazê-lo. Hubbard & Courtney (2010) observaram neste sentido uma influência de séries de alturas tonais descendentes no aumento do erro de localização “para baixo” (na vertical) da última posição percebida de um alvo visual em movimento. Vários outros estudos ilustrando, por sua vez, a modulação do erro de localização para diante (não para baixo) de um alvo visual por propriedades de sons contextuais são também mencionados nas revisões de Hubbard (2014a; 2018a). Pela sua importância no quadro da presente dissertação, esses estudos são descritos e comparados entre si numa secção específica do capítulo três consagrada ao MR audiovisual (secção 3.2). Como nota Hubbard (2014), a designação destes estímulos auditivos como contextuais envolve alguma convenção, podendo considerar-se, nalgumas situações, que a apresentação conjunta de um estímulo visual e auditivo corresponde na realidade à apresentação de um só estímulo audiovisual. Contudo, em todos os casos referidos a instrução solícita a localização do alvo visual, convertendo a visão em modalidade relevante e a audição numa modalidade não-alvo (contextual, neste sentido) (Hubbard, 2014a).

1.2.4. Características do observador

Incluem a alocação de atenção, movimentos oculares, idade, experiência anterior, psicopatologia, entre outras (ver Hubbard, 2005, 2014a, 2018a).

(i) Os efeitos da *atenção* no MR mantêm-se objeto de debate. Resultados favoráveis a um aumento do MR na ausência de atenção têm sido reportados (Hayes & Freyd, 2002; Joordens, Spalek, Razmy, & Van Duijn, 2004; Munger & Owens, 2004). Na interpretação de Hayes & Freyd (2002) a dinâmica do movimento do objeto (incluindo a tendência a manter o movimento) é representada sem necessidade de atenção, enquanto alterações à dinâmica, como a paragem ou desaparecimento do objeto, requerem atenção (na ausência de atenção ou com atenção diminuída, a magnitude do erro para diante, por conseguinte, aumenta). Pelo contrário, Kerzel (2003a) encontrou uma diminuição do MR na presença de distratores, concluindo pela necessidade de atenção para a manutenção da extrapolação para diante da

posição percebida do alvo. Hubbard, Kumar, & Carp (2009) mostraram que a apresentação de uma pista relativa ao local de desaparecimento do alvo (favorecendo a atenção ao local) reduzia o MR, tanto mais quanto maior a sua relevância (pistas mais relevantes especificavam as coordenadas verticais e horizontais do local, pistas menos relevantes apenas a coordenada horizontal). Apesar da diminuição do erro para diante, este nunca desapareceu completamente, levando os autores a concluir, simultaneamente, pela possibilidade de um controlo atencional do erro de localização, por um lado, e pela existência no MR de uma componente modular (cognitivamente impenetrável) de extrapolação da posição do objeto para diante.

(ii) Vários resultados têm indicado que, com alvos em movimento contínuo, a imposição de fixação ocular (interdição do seguimento oculomotor do alvo) resulta numa diminuição ou mesmo desaparecimento do MR (De Sá Teixeira et al., 2013; Kerzel, 2000, 2002c, 2003b). Estas observações levaram à proposta de que os mecanismos de perseguição oculomotora (“ultrapassagem” do alvo pelo olho na posição final), combinados com a persistência retiniana, poderiam explicar o momento representacional (Kerzel, 2000, 2005; Stork & Müsseler, 2004). No entanto, a interdição de movimentos oculares não afeta o MR obtido com movimento implícito (Kerzel, 2003a) ou com a apresentação de instantâneos fotográficos (Futterweit & Beilin, 1994). Embora a imposição de fixação ocular reduza o MR, quando a resposta é dada pelo método do tocar essa redução não ocorre quando as mãos do participante não se encontram visíveis (Ashida, 2004). Alguns estudos recentes utilizando movimento contínuo reportaram efeitos de deslocamento para diante em situações de interdição dos movimentos oculares (Getzmann & Lewald, 2009; Schmiedchen et al., 2013). A indiferença ao comportamento oculomotor tem sido documentada em vários estudos de momento representacional auditivo (Getzmann, 2005a; Getzmann & Lewald, 2009; Schmiedchen et al., 2012, 2013). No estado atual do debate sobre os mecanismos subjacentes ao MR, incluindo aí o papel a atribuir a mecanismos perceptivos “de baixo nível”, não é possível excluir que a interdição de movimentos oculares condicione na realidade a operação de representações de movimento das quais os próprios movimentos oculares dependem (e.g., controlo antecipatório dos movimentos oculares: e.g., Barnes & Marsden, 2002; Krauzlis, 2005) mais do que causam (ver Hubbard, 2005, 2006, 2010).

As restantes variáveis na categoria (e.g., idade, experiência, psicopatologia) não apresentam relevância direta para o presente trabalho. Sugestões de uma redução da magnitude do MR e mesmo da sua desaparecimento em populações mais idosas (> 65) têm sido fornecidas (Piotrowski & Jakobson, 2011), mas nenhum padrão consistente de alterações ao

longo do desenvolvimento se encontra estabelecido. Ser perito num domínio envolvendo interações com objetos em movimento (e.g., ser um condutor automóvel, piloto de avião, ou batedor de baseball experiente) determina um aumento comparativo (relativamente a não-peritos) do MR na localização desse tipo de objetos, mas não um aumento não específico do MR em todas as situações (Blättler, Ferrari, Didierjean, & Marmèche, 2011; Blättler, Ferrari, Didierjean, van Elslande, & Marmèche, 2010; Nakamoto, Mori, Ikudome, Unenaka, & Imanaka, 2015). Este resultado é consistente com um envolvimento importante no MR de expectativas-previsões relativas ao comportamento futuro dos objetos (Hubbard, 2018a).

1.3. DESLOCAMENTO M E DESLOCAMENTO O: GRAVIDADE REPRESENTACIONAL

O método da *sonda mnésica* revelou a existência de um deslocamento da recordação da última posição percebida de um alvo móvel para diante, na direção do movimento, dando lugar, desse modo, à analogia inicial com o momento físico enquanto tendência de manutenção do movimento (dada pela massa \times velocidade instantânea) (Freyd & Finke, 1984). A utilização posterior de alvos animados de movimento percebido como contínuo (*smooth continuous*) e respostas de localização pelo participante (e.g., Hubbard & Bharucha, 1988) revelou contudo deslocamentos noutras direções que não a do movimento. No caso de um alvo em movimento horizontal, para além do típico erro para diante associado ao momento representacional, os observadores cometem igualmente um erro de localização para baixo, ortogonal à direção do movimento (Hubbard, 1995c, 1995b, 2010). A circunstância de, num movimento horizontal, o erro ortogonal ao movimento se encontrar alinhado com a direção implícita da gravidade (conjugada com a observação de maiores deslocamentos em alvos animados de movimentos verticais descendentes do que ascendentes) levou à proposta de que se trataria de um análogo mental da atração gravitacional (Hubbard & Bharucha, 1988), designado como “gravidade representacional” (Hubbard, 1995c, 1995b). Somando-se ao momento representacional, a gravidade representacional constituiu um passo no sentido do alargamento da proposta inicial de Freyd & Finke (1984) em direção a uma hipótese mais geral de incorporação de invariantes físicos ambientais, sob a forma de conhecimento implícito, no nosso sistema representacional (Hubbard, 1995c). Resultados compatíveis com análogos representacionais da fricção (fricção representacional) (Hubbard, 1998b) ou força centrípeta (força centrípeta representacional) (Hubbard, 1996) foram neste sentido adicionalmente reportados.

A diferença entre momento representacional e gravidade representacional envolve um problema simultaneamente conceptual e de terminologia. A expressão momento representacional tem sido utilizada tanto para designar o deslocamento da última posição percebida de um alvo móvel no sentido do movimento, como para designar o mecanismo subjacente a esse deslocamento (Hubbard, 1995c, 2010). Ora, enquanto para um movimento horizontal o deslocamento no sentido do movimento é consistente com o momento representacional (análogo do momento físico), num movimento vertical envolve tanto o momento representacional, como a gravidade representacional (i.e., mais do que um mecanismo). Por sua vez, num movimento horizontal, o deslocamento ortogonal (para baixo) corresponde à gravidade representacional, mas o mesmo não sucede já num movimento vertical (onde pode exprimir a ação de outras variáveis, por exemplo, a atração por marcas espaciais - *landmarks*: Hubbard, 1998b). Hubbard (1995c, 2005) sugeriu neste sentido (1) utilizar as expressões descritivas *deslocamento M* e *deslocamento O* para referir, respetivamente, os erros de localização no sentido do movimento e ortogonais ao movimento, e (2) reservar as expressões *momento representacional* e *gravidade representacional* para as componentes do deslocamento observado consistentes com a forma como o momento físico e a gravidade, respetivamente, influenciariam um objeto físico. Esta proposta, permitindo dissociar claramente os níveis teórico e descritivo e acomodar a contribuição conjunta de vários mecanismos para um mesmo deslocamento, é seguida na presente dissertação.

Boa parte da investigação em torno da gravidade representacional tem envolvido a tentativa de esclarecer o modo como é afetada por diferentes variáveis (e.g., velocidade, altura do alvo no campo visual, orientação do movimento, tamanho/peso sugerido do alvo, orientação do observador, intervalo imposto até à resposta, modalidade de resposta, movimentos oculares: De Sá Teixeira & Oliveira, 2014; Hubbard, 2020) e a sua relação com outros mecanismos de deslocamento na memória espacial, entre os quais o momento representacional (De Sá Teixeira et al., 2013; Hubbard, 2020). As duas questões têm-se mantido relacionadas através do princípio de que efeitos semelhantes das variáveis moduladoras na gravidade representacional e no momento representacional apontariam para mecanismos comuns e divergências desses efeitos para mecanismos distintos. Na realidade, na ausência de um quadro explicativo capaz de integrar o conjunto da evidência, nenhum catálogo de semelhanças ou diferenças no modo como os dois tipos de deslocamento respondem a diferentes variáveis permite, por si só, decidir sobre a existência ou não de um mecanismo comum. Um tal quadro parece ainda longe de existir e as

evidências disponíveis mantêm até à data um estatuto incerto, assentando maioritariamente em resultados não replicados, frequentemente discordantes (ver revisão de Hubbard, 2020).

O estudo da gravidade representacional e do momento representacional em tarefas separadas, utilizando, respetivamente, movimentos verticais e horizontais, é menos favorável à compreensão das eventuais relações entre os dois tipos de fenómeno do que o seu estudo no quadro duma mesma tarefa de localização (De Sá Teixeira et al., 2013; De Sá Teixeira & Oliveira, 2014). Recentemente, a utilização de movimentos lineares com diferentes direções, associados a formas de decomposição trigonométrica (análise de Fourier) das componentes atribuíveis ao momento e à gravidade no erro de localização (De Sá Teixeira, 2014, 2016), tem apontado para uma dissociação dos mecanismos subjacentes às duas componentes (ver também, para outros argumentos de dissociação, De Sá Teixeira, Kerzel, Hecht, & Lacquaniti, 2019). No entanto, a tradução de evidência relativa aos efeitos diferenciais de certas variáveis (e.g., movimentos oculares) em evidência relativa a uma dissociação de mecanismos não é possível sem a mobilização de vários pressupostos adicionais. No caso particular de movimentos horizontais, onde o deslocamento O coincide com a direção implícita da gravidade e o deslocamento M com a ação implícita do momento representacional, resultados inconsistentes relativos ao efeito da velocidade do alvo, por exemplo, foram reportados. Enquanto De Sá Teixeira et al. (2013) observaram uma redução do deslocamento O com o aumento da velocidade (que induz por sua vez maior deslocamento M), Hubbard (1990) e Hubbard & Bharucha (1988) reportaram uma ausência de efeitos da velocidade no deslocamento O. Uma tendência para o aumento do deslocamento O com distâncias crescentes percorridas pelo alvo foi por sua vez documentada em De Sá Teixeira et al. (2013), genericamente consistente com o princípio de uma queda em trajetória parabólica, com a distância, de um objeto lançado horizontalmente (embora os resultados de detalhe sejam mais próximos de uma heurística de dissipação do ímpeto, capaz, igualmente, de articular os dois deslocamentos).

Numa dissertação de doutoramento não publicada, Martins (2015) examinou a correlação entre a magnitude de deslocamento M e deslocamento O de alvos em movimento horizontal em dois grupos de participantes, com e sem paralisia cerebral (uma condição clínica caracterizada fundamentalmente por problemas na planificação, controlo e execução de movimentos). A hipótese de partida era a de que toda a previsão sensorial ao serviço da ação é estrangida pelo reportório específico de ações do organismo (König, Wilming, Kaspar, Nagel, & Onat, 2013) (ver também König & Krüger, 2006): hábitos de planeamento da interseção ou captura de objetos específicos a cada participante poderiam assim suportar

correlações significativas entre os dois deslocamentos no quadro da execução de uma resposta de localização. Os resultados foram consistentes com essa hipótese nos dois grupos (participantes com maior deslocamento M tendiam igualmente a apresentar maior deslocamento O, sendo as magnitudes de deslocamento significativamente superiores no grupo com paralisia cerebral). Este resultado não foi contudo nunca replicado, mantendo-se por conseguinte incerto quanto à sua generalidade.

Com exceção de um estudo realizado com séries descendentes e ascendentes de alturas tonais (manipulando portanto o contínuo das frequências) (Hubbard & Ruppel, 2013), a gravidade representacional não foi investigada no domínio auditivo, tendo sido exclusivamente abordada até à data com estímulos visuais. Uma influência intermodal de séries tonais descendentes no aumento do deslocamento O de um movimento visual horizontal foi também observada por Hubbard & Courtney (2010). Como assinalado a propósito do estudo anterior, estes efeitos foram mediados por manipulações do som no domínio das frequências, não existindo registo de estudos realizados com sons espacializados em movimento.

1.4. MOMENTO REPRESENTACIONAL AUDITIVO: SEMELHANÇAS E DIFERENÇAS RELATIVAMENTE AO MOMENTO REPRESENTACIONAL VISUAL

A maioria dos estudos de momento representacional foi conduzida na modalidade visual, onde este erro de localização foi originalmente estabelecido (Freyd & Finke, 1984; Hubbard, 1995c, 2010). Porém, a questão de saber se se restringia a transformações visualmente percebidas foi colocada desde cedo (Kelly & Freyd, 1987), levando à realização de alguns estudos com estímulos auditivos (Freyd et al., 1990; Hubbard, 1993, 1995a; Kelly & Freyd, 1987). Todos estes estudos iniciais de momento representacional auditivo foram conduzidos no domínio das frequências, envolvendo transformações ascendentes e descendentes da altura tonal, e não transformações no espaço (movimento de fontes sonoras). Em geral, séries de três ou quatro tons eram apresentadas e os participantes comparavam o último tom da série a um tom subsequente (sonda mnésica). Distorções na memória em direção a sondas com maior frequência (mais agudas) foram encontradas nas séries ascendentes, e em direção a sondas com menor frequência (mais graves) nas séries descendentes, em

conformidade com o deslocamento para diante típico do momento representacional (Kelly & Freyd, 1987).

A comparação com os resultados conhecidos do momento representacional visual revelou ainda outras semelhanças. Freyd et al. (1990) documentaram efeitos da velocidade e da aceleração das transformações de frequência semelhantes aos observados no momento representacional de alvos visuais (magnitude do efeito proporcional à velocidade da transformação e à aceleração final: Finke, Freyd, & Shyi, 1986). Mostraram ainda a importância da coerência das transformações, ilustrando o cancelamento dos efeitos de distorção quando a ordem dos primeiros dois tons era inconsistente com a direção geral da série (Freyd et al., 1990). Idêntico resultado fora obtido com estímulos visuais (Freyd & Finke, 1984), revelando-se a previsibilidade das trajetórias de movimento uma condição indispensável para a ocorrência de momento representacional visual (Kerzel, 2002a). De acordo com a demonstração no MR visual de efeitos do conhecimento dos observadores sobre o movimento típico do objeto (e.g., alvos definidos como representações de um foguetão apresentavam maior deslocação para diante: Reed & Vinson, 1996; Vinson & Reed, 2002), Hubbard (1993a) observou distorções mnésicas do último tom de uma série de três, consistentes com o conhecimento das escalas tonais da música ocidental (e.g., numa série ascendente em que os dois primeiros tons correspondiam à tônica e à dominante de uma escala maior, um terceiro tom um pouco acima da oitava era recordado como mais grave – e não como mais agudo –, aproximando-se da relação de oitava característica da escala). Vários estudos com estímulos visuais mostraram a dependência do momento representacional relativamente a expectativas criadas por padrões de transformação mais complexos (globais) do que o simples movimento linear ou circular – e.g., movimento periódico de vaivém linear (Hubbard & Bharucha, 1988), movimento periódico de rotação (Verfaillie & D'Ydewalle, 1991), movimento em espiral (Freyd & Jones, 1994). Num estudo de momento representacional auditivo com sequências tonais envolvendo transformações lineares e periódicas (cíclicas), Johnston & Jones (2006) mostraram a ocorrência de distorções mnésicas consistentes com os padrões de transformação global (alterações periódicas) e não apenas local. Por fim, Hubbard (1995a) mostrou que a apresentação discreta ou contínua das transformações tonais não influenciava o MR auditivo, tendo replicado parcialmente os resultados anteriores de Freyd et al. (1990) quanto aos efeitos da velocidade (distorções crescentes com a velocidade foram encontradas nas séries descendentes, embora não nas ascendentes).

Este conjunto de observações não apenas sugeria uma natureza supramodal do momento representacional, não restringido à visão, como questionava a sua relação necessária com transformações que especificassem alterações da posição espacial dos objetos. Em geral, variações da altura tonal não se encontram correlacionadas com o movimento dos objetos (o efeito Doppler, um dos poucos exemplos onde uma associação entre variação dinâmica do *pitch* e movimento existe ecologicamente, é pouco prevalente no ambiente perceptivo humano). Em alternativa ao abandono integral da ideia de uma relação do fenómeno do momento representacional com o momento físico – em favor, por exemplo, de um mero princípio de “boa continuação”, sem relação necessária com o movimento –, Freyd et al. (1990) e Kelly & Freyd (1987) consideraram duas possibilidades. (1) O momento representacional poderia ter origem numa internalização do momento físico no sistema visual, sendo em seguida cooptado por outros sistemas e modalidades, como a audição, onde poderia revelar-se útil do ponto de vista preditivo. Kelly & Freyd (1987) evocam neste sentido as descrições de Attneave & Olson (1971) e Shepard & Cooper (1982) do contínuo das alturas tonais como um meio espacial no qual os tons se movimentam, podendo assim beneficiar de um mecanismo antecipatório com origem visuo-espacial. (2) O momento representacional poderia corresponder a uma propriedade geral de um sistema representativo dotado de coerência espaço-temporal (i.e., incorporando o tempo numa representação espacial abstrata das transformações) e utilizando o ritmo ou a velocidade das alterações para antecipar o estado futuro dos eventos. Efeitos semelhantes ao momento representacional visual poderiam, desse modo, encontrar-se em qualquer dimensão de variação contínua. Uma solução deste segundo tipo parece inspirar as propostas recentes de Hubbard (2014a, 2015, 2017) de perspetivar diferentes formas de momento (representacional, operacional, atencional) como instâncias de um mesmo mecanismo de extrapolação de transformações representáveis (abstratamente, funcionalmente) sob a forma de movimento num sistema de coordenadas espaciais.

Apesar do interesse destes primeiros estudos, a interpretação dos resultados dependia de um pressuposto de analogia funcional entre variações do som no domínio das frequências e variações da posição espacial de um objeto visualmente percebido. Comparações diretas entre o movimento representacional visual e o movimento representacional auditivo exigiriam, diferentemente, a utilização de tarefas espaciais, com estímulos auditivos espacializados em movimento. Tendo em conta as diferenças entre pistas de movimento utilizadas por cada modalidade (e.g., alteração de padrões retinianos na visão, alterações das diferenças binaurais de tempo e intensidade na audição) e as prováveis

diferenças na codificação dos espaços visual (baseada numa representação topograficamente organizada) e acústico (dependente de pistas locais) (Wilson & O'Neill, 1998), divergências importantes quanto ao momento representacional entre as duas modalidades poderiam na realidade existir, sem que os estudos anteriores no domínio das frequências as pudessem revelar. Getzmann et al. (2004) realizaram o primeiro estudo sistemático de momento representacional auditivo com uma tarefa espacial. Utilizando um arranjo denso de altifalantes ativados sucessivamente, verificaram um erro para diante na localização da última posição percebida do alvo, semelhante ao momento representacional visual. Este resultado confirmava observações anteriores sobre a localização final de objetos acústicos em movimento, que indicavam erros no sentido do movimento (Mateeff & Hohnsbein, 1988; Perrott & Musicant, 1977), e foi corroborado em vários estudos posteriores (Getzmann, 2005a; Getzmann & Lewald, 2007, 2009; Schmiedchen et al., 2013), apoiando assim a hipótese de uma natureza supramodal do momento representacional.

No entanto diferenças importantes no modo como diferentes variáveis afetavam este erro de localização na visão e na audição foram também documentadas. Por exemplo, enquanto no MR visual uma relação crescente entre a velocidade do movimento e a magnitude do erro é um dado robusto (desde que a posição de desaparecimento do alvo não ultrapasse certos valores de excentricidade: ver Schmiedchen et al., 2013), no MR auditivo espacial essa relação mantém-se controversa. Os resultados reportados incluem ausência de efeitos (Perrott & Musicant, 1977), efeitos crescentes (Feinkohl, Locke, Leung, & Carlile, 2014), decrescentes (Getzmann et al., 2004) e condicionais à direção do movimento (Schmiedchen et al., 2013). Contrariamente a resultados sobre o enfraquecimento ou mesmo eliminação do MR visual obtido com movimentos contínuos na ausência de movimentos oculares (Kerzel, 2000, 2005), a magnitude do MR auditivo espacial não parece ser influenciada pelos movimentos oculares (Getzmann, 2005a; Getzmann & Lewald, 2009; Schmiedchen et al., 2012, 2013). No primeiro estudo envolvendo uma comparação direta da magnitude do MR visual e auditivo em tarefas espaciais idênticas, com alvos movendo-se a velocidades constantes nas duas modalidades, Getzmann & Lewald (2009) encontraram valores de MR semelhantes na visão e na audição. Os autores concluíram que a magnitude do MR é uma dimensão independente da modalidade. No entanto, Schmiedchen et al. (2013) verificaram que esta semelhança apenas se verificava para posições de desaparecimento do alvo relativamente centrais. Ao procederem a comparações em regiões paracentrais e periféricas, importantes diferenças na magnitude do erro surgiram entre modalidades.

Tomados no seu conjunto, os estudos, ainda pouco numerosos, sobre o momento representacional auditivo espacial apoiaram uma semelhança geral deste erro de localização entre as modalidades visual e auditiva, suportando as indicações das primeiras investigações conduzidas no domínio das frequências. Este resultado é compatível com o contributo de mecanismos de extrapolação preditiva supramodais, partilhados por ambas as modalidades. Contudo, diferentemente dos primeiros estudos, documentaram também um impacto no MR do processamento específico a cada modalidade de diferentes parâmetros do movimento (efeitos da velocidade, direção, ponto de desaparecimento, movimentos oculares). À luz das indicações disponíveis, tanto mecanismos partilhados de natureza preditiva (eventualmente de mais alto nível) como mecanismos específicos a cada modalidade parecem assim contribuir para o momento representacional na visão e na audição (Hubbard, 2014a; Schmiedchen et al., 2013).

CAPÍTULO 2

INTERAÇÃO-INTEGRAÇÃO MULTISSENSORIAL

A percepção do mundo real é invariavelmente multissensorial. Não obstante, a ideia de que os humanos são animais predominantemente visuais, combinada com uma concepção estritamente modular das bases neurocorticais do processamento sensorial remeteram durante muito tempo a investigação da percepção multimodal para um plano secundário (Shams, 2012). Esta situação parece ter-se alterado sob a influência de vários fatores convergentes, incluindo a acumulação de evidência contra o pressuposto da dominância visual (Alais et al., 2010; Hidaka, Teramoto, & Sugita, 2015; Shams, 2012; Shams & Kim, 2010), o reconhecimento da extensão do processamento multissensorial no cérebro, evidenciado mesmo em áreas classicamente consideradas unimodais (e.g., Ghazanfar & Schroeder, 2006; Schroeder & Foxe, 2005; Stein & Stanford, 2008), uma compreensão crescente da multiplicidade de níveis a que as interações intermodais podem ter lugar (e.g., de Haas & Rees, 2010; Shams & Kim, 2010; Talsma et al., 2010) e a demonstração recorrente de efeitos de melhoria bimodal/multimodal (relativamente aos níveis de execução unimodais) em diversas tarefas, com potenciais implicações práticas significativas (e.g., Frassinetti, Bolognini, & Làdavas, 2002; Seitz, Kim, & Shams, 2006; Shams & Seitz, 2008).

A investigação sobre interação-integração multissensorial, com destaque para o domínio do processamento audiovisual, desenvolveu-se rapidamente nas últimas duas décadas sob o efeito destas alterações. Apesar do rápido crescimento da base empírica disponível, a complexidade intrínseca da percepção multissensorial continua a suscitar novas questões e direções de investigação (Wallace & Stevenson, 2014), por exemplo, quanto às

articulações do processamento multissensorial com a atenção (e.g., Talsma et al., 2010; Tang et al., 2016) e outros processos cognitivos (e.g., Wallace, Woynaroski, & Stevenson, 2020). Embora a expressão “integração multissensorial” seja frequentemente tomada, num sentido lato, como referindo-se «ao conjunto de processos através dos quais as informações procedentes das modalidades sensoriais individuais (e.g., visão, audição, tato) interagem e influenciam o processamento noutras modalidades, incluindo o modo como estas informações sensoriais são combinadas entre si, de modo a dar lugar a uma experiência perceptiva multissensorial unificada» (Talsma et al., 2010), vários autores acentuam a conveniência de distinguir entre integração multissensorial e interação intermodal (Bertelson & Radeau, 1981; Schmiedchen et al., 2012; Spence, Senkowski, & Röder, 2009). Neste sentido mais estrito, a integração implica uma ligação efetiva (*binding*) dos diferentes sinais sensoriais num só percepto unificado (fusão perceptual), enquanto a interação intermodal se aplica genericamente à modulação/enviesamento do processamento de uma modalidade sensorial por uma outra. Apesar de conceptualmente razoável, esta distinção envolve o problema empírico-operacional da distinção entre as duas situações e, em particular, dos critérios empíricos que definem a ocorrência de integração multissensorial “no sentido próprio” (para a ausência de consenso e dificuldades associadas a este ponto ver, e.g., Bizley, Jones, & Town, 2016; Colonius & Diederich, 2017; Ernst & Bühlhoff, 2004). Por outro lado, não é inteiramente clara a forma como intersecta outras distinções propostas, por exemplo, entre “combinação” e “integração sensorial” (Ernst & Bühlhoff, 2004), ou entre “integração completa” e “integração parcial” (Ernst & Di Luca, 2011). Tanto neste capítulo como no trabalho empírico reportado na dissertação, a diferença entre interação e integração multissensorial não desempenha um papel relevante. À semelhança de muita da literatura anterior (ver e.g., Burr & Alais, 2006; Godfroy-Cooper et al., 2015), a expressão “integração” será por conseguinte aqui utilizada como uma forma abreviada da expressão composta interação-integração multissensorial. Manteremos no entanto, em particular no presente capítulo, o uso frequente da expressão composta (interação-integração) e, ao longo da dissertação, da expressão coordenada “interação e integração multissensoriais”, de modo a sinalizar a potencial relevância desta distinção e a existência de aspetos definicionais em aberto na noção de integração multissensorial.

O objetivo do presente capítulo é o de fornecer um panorama dos principais quadros interpretativos dos fenómenos de interação-integração intermodais, uma caracterização das regras e princípios de integração multissensorial emergentes da investigação anterior, uma

apresentação do problema da articulação entre processamento multissensorial e atenção (exógena e endógena) e uma breve revisão argumentada da noção de percepto audiovisual.

2.1. HIPÓTESES E MODELOS: DOMINÂNCIA VISUAL, ADEQUAÇÃO DA MODALIDADE, MODELO DE ESTIMAÇÃO DA MÁXIMA VEROSIMILHANÇA, MODELOS BAYESIANOS DE INFERÊNCIA CAUSAL

2.1.1. Hipótese da dominância visual

Até ao final do século passado, a maioria dos estudos sensório-perceptivos, cognitivos e neurofisiológicos ocupavam-se de cada uma das modalidades separadamente, não existindo senão um interesse secundário pela interação-integração multissensorial (Alais & Burr, 2019). Parte deste desinteresse devia-se à premissa de que a relação entre modalidades era regida por um princípio hierárquico de predominância da visão sobre as restantes modalidades (Shams & Kim, 2010). Num artigo clássico, Posner, Nissen, & Klein, (1976) propuseram uma explicação da dominância visual, observada tanto em tarefas perceptivas como mnésicas, em termos de alocação atencional. Uma menor capacidade dos estímulos visuais (comparativamente aos estímulos de outras modalidades) para alertar automaticamente o organismo quanto à sua presença determinaria a necessidade de uma atenção acrescida à modalidade visual, originando um enviesamento atencional favorável à visão. Uma explicação mais comum, todavia, assenta num privilégio concedido ao processamento da informação espacial (Hidaka et al., 2015), refletida na definição dos humanos como “animais visuais” (Shams, 2012). Em contraste com a natureza imediatamente espacial da informação visual, as restantes modalidades (tato incluído) apenas poderiam fornecer esse tipo de informação, considerada essencial, através da sua associação com a visão, isto é, de modo indireto e subordinado (Rock, 1966). No caso particular das relações entre a visão e a audição, exemplos frequentemente invocados de dominância da visão incluem o ventriloquismo espacial (a atração/captura da localização percebida de um som pela localização de um estímulo visual: e.g., Bertelson, 1999; Recanzone, 2009; Welch & Warren, 1980), o efeito McGurk observado na percepção da fala (uma tarefa considerada auditiva), no qual a articulação labial visível induz uma alteração da sílaba percebida (McGurk & MacDonald, 1976), ou o efeito Colavita (Colavita, 1974), no qual

a presença de um estímulo auditivo apresentado juntamente com um estímulo visual (apresentação audiovisual), ambos requerendo uma resposta, é aparentemente ignorada num largo número de vezes (para uma discussão, ver Y.-C. Chen & Spence, 2017).

Um segundo fator determinante para esta hipótese foi a concepção de uma organização cortical hierárquica em que o processamento da informação de cada modalidade ocorria inicialmente em áreas e vias específicas, ao abrigo da interação com informação de outras modalidades. Apenas em estádios posteriores, suportados por áreas corticais associativas a informação de diferentes modalidades poderia convergir (Alais et al., 2010; Recanzone, 2009; Shams & Kim, 2010). Este modelo hierárquico simples, pressupondo a independência e modularidade do processamento modal, é essencial à hipótese de uma articulação multissensorial tardia sob a égide da dominância visual. No entanto, é hoje impossível de manter face à evidência disponível (ver e.g., Alais & Burr, 2019; Alais et al., 2010; Ghazanfar & Schroeder, 2006; Schroeder & Foxe, 2005; Shams, 2012). Resultados documentando a existência de conectividade anatômica direta e indireta entre as áreas corticais qualificadas como “unissensoriais” têm-se acumulado (Falchier, Clavagnier, Barone, & Kennedy, 2002; Ghazanfar & Schroeder, 2006; Hackett et al., 2007; Rockland & Ojima, 2003). No caso particular da visão e da audição, há evidência crescente de processamento multissensorial na área cortical V1 em consequência de estimulação auditiva (Burton, Snyder, Diamond, & Raichle, 2002; Röder, Stock, Bien, Neville, & Rösler, 2002; Watkins, Shams, Tanaka, Haynes, & Rees, 2006) e mesmo de convergências multissensoriais envolvendo as duas modalidades a níveis subcorticais, como o colículo superior (Stein & Meredith, 1993). Um outro aspeto que questiona igualmente a ideia de um processamento hierárquico simples é a importância demonstrada da modulação reentrante (*feedback-dependent*) do processamento nas áreas sensoriais primárias, semelhante à observada no processamento sensorial contextual e na modulação atencional do processamento (Fiser, Chiu, & Weliky, 2004; Lamme, 1995; Mehta, 2000; Schroeder & Foxe, 2005). A ideia de que o processamento visual ocorre no quadro e sob a influência de uma rede alargada de processamento cortical multissensorial surge assim, do ponto de vista neuroanatômico e neurofuncional, como mais adequada do que a concepção modular inicial (Driver & Noesselt, 2008; Schroeder & Foxe, 2005; Shams, 2012; Shams & Kim, 2010).

Os resultados da investigação comportamental acompanharam também o sentido das alterações verificadas na investigação das bases neurofisiológicas do processamento multissensorial, produzindo demonstrações recorrentes de modulação, tanto quantitativa como qualitativa, da visão por outras modalidades (Shams & Kim, 2010). Por exemplo, a

visão estereoscópica (Atkins, Jacobs, & Knill, 2003), a recuperação da profundidade a partir do sombreado (Adams, Graf, & Ernst, 2004), a percepção visual do declive (Ernst, Banks, & Bühlhoff, 2000) ou da forma (Atkins, Fiser, & Jacobs, 2001) mostraram ser influenciadas por informação tátil. No caso particular da audição, fenómenos de dominância da audição sobre a visão foram evidenciados no domínio do processamento temporal. O chamado *ventriloquismo temporal* (por contraposição ao *ventriloquismo espacial*), interpretado como uma “captura” de aspetos temporais de um estímulo visual (e.g., o momento de início ou a duração) por aspetos temporais do som, constitui um exemplo frequente invocado (Bertelson & Aschersleben, 2003; Morein-Zamir, Soto-Faraco, & Kingstone, 2003). Ilusões como a do “duplo *flash*”, na qual dois bipes apresentados em rápida sucessão conjuntamente com um *flash* induzem a percepção de dois *flashes* (Shams, Kamitani, & Shimojo, 2000), ou o fenómeno da “liderança do som” (*auditory driving*) (Shipley, 1964), no qual a frequência de cintilação (*flicker*) de um *flash* é atraída pela frequência de oscilação (*flutter*) de um som, sem que um efeito simétrico dos *flashes* no som se verifique (Gebhard & Mowbray, 1959), fornecem outros exemplos claros de uma predominância da audição sobre a visão no domínio temporal. Observações no mesmo sentido, ilustrando uma atração pelo som da localização temporal de eventos visuais têm sido reportadas noutros contextos, como o do estudo do movimento aparente (E. Freeman & Driver, 2008; Getzmann, 2007), ou de fenómenos dinâmicos como o efeito *flash-lag* (Alais & Burr, 2003; Vroomen & de Gelder, 2004).

Longe de constituir um fenómeno transitório e avulso, a modulação da visão por outras modalidades, como no caso do *ventriloquismo temporal*, apresenta pós-efeitos duráveis (ver, e.g., Fujisaki, Shimojo, Kashino, & Nishida, 2004; Vroomen, Keetels, de Gelder, & Bertelson, 2004), à semelhança do *ventriloquismo espacial* (ver, e.g., Mendonça, Escher, van de Par, & Colonius, 2015; Recanzone, 1998; Zwiers, Van Opstal, & Paige, 2003), sugestivos de um papel central da interação-integração multissensorial na aprendizagem, adaptação e (re)calibração do sistema percetivo (L. Chen & Vroomen, 2013). A hipótese de uma dominância geral da visão sobre a audição é assim impossível de manter, apontando estes resultados (bem como resultados adicionais assinalados adiante) para a necessidade de outros quadros de leitura da organização das interações-integrações multissensoriais, designadamente audiovisuais.

2.1.2. Hipótese da adequação da modalidade

Uma outra hipótese, capaz de enquadrar no essencial os resultados mencionados na subsecção anterior, é a da *adequação da modalidade* (*modality appropriateness hypothesis*), ou da *especificidade da modalidade* (*modality specificity hypothesis*) (Welch, 1999; Welch & Warren, 1980, 1986). De acordo com esta hipótese, a influência de uma modalidade perceptiva na interação-integração multissensorial depende do seu grau de adequação à tarefa. Considerando que esta adequação depende fundamentalmente da precisão e acuidade informativa da modalidade relativamente à dimensão ou dimensões crítica(s) da tarefa, a hipótese prevê geralmente que, em situações de estimulação bimodal/multimodal, prevalece a modalidade com melhor resolução nas dimensões de interesse (Talsma et al., 2010). Sendo geralmente aceite que a modalidade visual oferece tipicamente melhor resolução espacial do que a audição (Fisher, 1968), e a audição melhor resolução temporal do que a visão (Welch, DuttonHurt, & Warren, 1986), tanto os efeitos de ventriloquismo espacial como os de ventriloquismo temporal referidos atrás seriam assim explicáveis pela hipótese da adequação da modalidade. Que a informação sonora, ao assistirmos a um programa de televisão, seja percebida no local da ação e não nas saídas de som da TV (ventriloquismo espacial), e que o som e o vídeo da TV pareçam síncronos apesar de atrasos físicos entre os sinais (ventriloquismo temporal) podem compreender-se à luz de uma mesma hipótese (L. Chen & Vroomen, 2013).

Uma previsão frequentemente derivada quanto às relações entre a visão e a audição é a de que os sons devem ser capazes de influenciar a visão em tarefas temporais (e.g., envolvendo propriedades como a frequência, ritmo, *timing* ou duração), mas não em tarefas espaciais (e.g., localização espacial, direção, tamanho). Na realidade, no quadro da hipótese da adequação da modalidade, o privilégio da audição no processamento temporal permite prever efeitos do som no processamento visual mesmo em tarefas espaciais, desde que tais efeitos sejam mediados pelo processamento temporal (Kawachi, 2016). A hipótese pode assim acomodar um conjunto numeroso de resultados que ilustram a modulação pela audição de aspetos espaciais do processamento visual do movimento. Exemplos possíveis incluem a ilusão do “fluxo-ressalto” (*stream-bounce illusion*), na qual a apresentação de um som na vizinhança temporal da interseção de dois alvos visuais em movimento altera qualitativamente a direção e estrutura do movimento percebido (Sekuler, Sekuler, & Lau, 1997), a modulação da posição espacial percebida do “ressalto” de um alvo visual em

movimento contra uma superfície pelo intervalo temporal entre a apresentação de um som e o momento do ressalto (Heron, Whitaker, & McGraw, 2004), a redução do erro na localização espacial de um *flash* estático relativamente a um alvo em movimento (*flash-lag*) induzida por um som não apresentado em simultâneo com o *flash* (Vroomen & de Gelder, 2004), ou a determinação da direção percebida de um movimento visual aparente pela relação temporal entre “bipes” e estímulos visuais estáticos indutores da impressão de movimento (E. Freeman & Driver, 2008).

Apesar da sua abrangência, esta hipótese é no entanto incompatível com outros resultados conhecidos. Por exemplo, no estudo mencionado acima de Heron et al. (2004), o efeito do som na localização percebida do “ressalto” apenas ocorria com sons de muito curta duração (2.33 ms), implicando uma baixa incerteza temporal, e não com sons de maior duração (e.g., 74.56 ms), envolvendo maior incerteza temporal. O efeito, apesar de se exercer no domínio temporal, depende assim não de um privilégio da modalidade auditiva *em si mesma*, mas de a informação temporal fornecida pelo som ser, na circunstância, mais fiável (menos incerta) do que a fornecida por outras entradas sensoriais. Do mesmo modo, a dominância da visão, longe de se apresentar geral no processamento de informação espacial, depende da fiabilidade da informação que pode fornecer na circunstância. Alais & Burr (2004) demonstraram neste sentido um efeito de ventriloquismo espacial inverso (atração/captura da localização visualmente percebida pela localização aparente de sons espacializados) numa tarefa de localização de estímulos visuais estáticos degradados por “desfocagem” (*blur*). Vários outros estudos ilustraram a modulação de aspetos espaciais do processamento visual por informação espacial auditiva na periferia do campo visual, onde a resolução espacial da visão (a fiabilidade da informação visual) se encontra significativamente reduzida (Hidaka, Manaka, et al., 2009; Hidaka et al., 2011). De modo sistemático, Godfroy-Cooper et al. (2015) documentaram a dependência da combinação de informação espacial visual e auditiva em tarefas de localização espacial relativamente aos efeitos da direção e excentricidade na precisão e acuidade de cada uma das modalidades.

Estes resultados, contrários às previsões deriváveis da hipótese da adequação da modalidade, tornam clara a necessidade da sua reconversão no sentido de admitir que as interações intermodais não dependem apenas ou primariamente das propriedades características de cada modalidade sensorial, mas da certeza relativa/fiabilidade dos sinais de entrada (Hidaka et al., 2015), sendo os conflitos intermodais aparentemente resolvidos por meio da subordinação do sentido menos fiável e de recalibrações sensoriais (Ernst & Banks, 2002; Ernst & Bühlhoff, 2004; Soto-Faraco, Spence, & Kingstone, 2004). Ernst &

Bülthoff (2004) assinalam neste sentido que a expressão “precisão da estimativa” é factualmente mais apropriada do que a de “adequação da modalidade” ou “precisão da modalidade” (por vezes utilizada como um equivalente). A ideia de que os graus de dominância dependem da fiabilidade relativa das estimativas dos sinais sensoriais e não da precisão da modalidade em geral tem sido concretizada e formalizada largamente no quadro dos modelos Bayesianos de interação-integração multissensorial (Alais & Burr, 2004; Burr & Alais, 2006). Para além de reinterpretar toda a anterior evidência consistente com a hipótese da “adequação da modalidade” em termos da precisão das estimativas fornecidas por cada modalidade, o quadro Bayesiano acomoda também flexivelmente, através dos mesmos mecanismos, as situações contrárias à “adequação da modalidade” (e.g., ventriloquismo temporal inverso: Alais & Burr, 2004) e a observação comum de que raramente a dominância de uma modalidade sobre outra é total, mantendo a modalidade dominada pelo menos uma contribuição residual (Rock & Victor, 1964). Enquanto do ponto de vista de um pressuposto rígido de dominância de uma modalidade sobre outra esta componente residual variável se mantém inexplicada, no quadro de um modelo flexível de ponderação da informação sensorial pela sua fiabilidade, a presença da informação sensorial dominada, com um peso reduzido (num grau que depende da sua fiabilidade relativa), é esperável na maioria das situações (Alais & Burr, 2019).

2.1.3. Modelos Bayesianos de integração sensorial

Os modelos Bayesianos de integração são modelos matemáticos que representam sob a forma de relações probabilísticas a incerteza associada ao conjunto das variáveis (tanto de entrada como de saída) e parâmetros de interesse de um sistema. Têm sido amplamente utilizados na modelização de conhecimento em biologia computacional e bioinformática e na modelização de funções probabilísticas do sistema nervoso, incluindo aí a integração multissensorial (Shams & Beierholm, 2010; Shangari, Falahi, & Bakouie, 2015). Muita da investigação de inspiração Bayesiana no domínio da combinação de pistas sensoriais tem sido realizada no domínio da visão e da integração multissensorial, com sucesso assinalável. No entanto, os princípios utilizados têm uma natureza geral e podem aplicar-se igualmente à combinação intramodal de pistas noutras modalidades, com um significativo potencial de contribuição, por exemplo, para os domínios menos investigados da perceção auditiva e da psicoacústica (Alais & Burr, 2019). Comparativamente a outras abordagens, os modelos

Bayesianos apresentam várias vantagens, com destaque para a capacidade de operar em situações com dados de entrada em falta, de modelizar relações causais que suportam uma compreensão do sistema e a previsão de consequências de intervenções, e de combinar conhecimento *a priori* e evidência sensorial de forma ótima/ideal (Shangari et al., 2015).

Na sua lógica mais geral, a abordagem Bayesiana assenta num processo formal de revisão de probabilidades, possibilitando a passagem de uma distribuição de probabilidades inicial (*a priori*), através da sua combinação com nova evidência, a uma distribuição de probabilidades final (*a posteriori*) que reflete a combinação ótima das duas fontes de informação. No domínio do estudo da perceção, oferece um quadro formal adequado à ideia da perceção como “inferência inconsciente” (Helmholtz, 1866) ou do mecanismo percetivo como um gerador de hipóteses prováveis sobre os objetos e eventos do mundo (Gregory, 1997) com base em conhecimento prévio e evidência sensorial incerta. O conhecimento (resultante da aprendizagem, expectativas, experiência prévia) é representado no modelo pela probabilidade *a priori* dos eventos, sendo frequentemente concebido como um modelo interno das propriedades estatísticas do meio. A evidência sensorial encontra-se por sua vez associada a uma função de verosimilhança, que descreve a sua probabilidade condicionalmente a diferentes eventos. A distribuição de probabilidades final, em conformidade com o teorema de Bayes, é uma combinação multiplicativa da probabilidade *a priori* e da verosimilhança, correspondendo à probabilidade de um evento condicionalmente ao conhecimento e evidência sensorial disponíveis (Alais & Burr, 2019; Feldman, 2015).

No caso de estimulação multimodal, haverá sinais sensoriais em duas ou mais modalidades, cada um dos quais associado a uma função de verosimilhança. A probabilidade de combinar as diferentes estimativas modais numa estimativa multissensorial integrada pode modelizar-se, nos termos da equação de Bayes, como a multiplicação da probabilidade *a priori* de estimulação multimodal por uma fonte comum (probabilidade de coocorrência) pela verosimilhança de cada uma dos componentes modais, normalizada (por quociente) ao produto das probabilidades *a priori* de cada componente (Alais & Burr, 2019). Este modo de formalização da interação-integração multissensorial tem sido aplicado com sucesso numa ampla variedade de domínios (e.g., Deneve & Pouget, 2004; Ernst & Banks, 2002; Feldman, 2015; Fetsch, DeAngelis, & Angelaki, 2013; Kayser & Shams, 2015; Kilteni, Maselli, Kording, & Slater, 2015), designadamente à perceção de movimento audiovisual e ao estudo de efeitos audiovisuais em tarefas de localização (Burr & Alais, 2006).

Apesar da lógica de base comum, existem diferentes classes de modelos Bayesianos de interação-integração sensorial, com implicações e alcances diversos. Uma distinção

particularmente importante envolve a diferença entre modelos de combinação de sinais sensoriais que pressupõem a sua origem comum num mesmo objeto/evento (i.e., sinais sensoriais sem disparidades sistemáticas, apenas afetados por fontes de ruído independentes), e modelos que admitem origens distintas para diferentes sinais, envolvendo inferências sobre a estrutura causal do ambiente. Enquanto os primeiros têm a sua aplicação essencialmente limitada às situações de “fusão multissensorial” os segundos acomodam todo o espectro de interações possíveis, da fusão à segregação, passando por diferentes graus de integração parcial (Ernst & Di Luca, 2011; Kayser & Shams, 2015; Mendonça, Mandelli, & Pulkki, 2016; Roach, Heron, & McGraw, 2006; Shams, 2012; Shangari et al., 2015). A mesma distinção vigora no quadro de modelos Bayesianos computacionais dinâmicos (i.e., incorporando as características temporais dos processos de interação-integração sensorial), por exemplo, entre o Filtro de Kalman (pressupondo entradas sensoriais sem enviesamentos, afetados unicamente por ruído) e as Redes Bayesianas dinâmicas (integrando inferência causal) (Shangari et al., 2015). As duas subsecções seguintes detalham aspetos relevantes desta distinção.

2.1.3.1. *Modelo da estimação de máxima verosimilhança (MMV)*

O modelo MMV (Alais & Burr, 2004; Ernst & Banks, 2002) da integração multissensorial é um modelo Bayesiano simplificado de combinação de pistas sensoriais, que considera unicamente a verosimilhança dos sinais, ignorando as probabilidades *a priori* da sua coocorrência. Isto significa que os sinais são considerados redundantes (isto é, veículos de informação não enviesada sobre uma fonte comum), diferendo apenas na sua fiabilidade, tanto maior quanto menor a incerteza devida ao ruído que os afeta. Tipicamente, o ruído é assumido como gaussiano e independente (não correlacionado) entre sinais. Com base nestes pressupostos, o MMV preconiza uma forma de combinação ótima das pistas sensoriais, sendo o critério do “ótimo” neste quadro a obtenção da estimativa combinada com menor variância, isto é, que minimiza a incerteza devida ao ruído nos sinais componentes. O carácter ótimo da combinação refere-se assim à precisão da estimativa final, não à exatidão dessa estimativa, pressuposta pela admissão do carácter não enviesado da informação unimodal (Alais & Burr, 2019; Burr & Alais, 2006; Ernst & Di Luca, 2011; Godfroy-Cooper et al., 2015).

Do ponto de vista matemático, a estimativa fornecida por cada sinal sensorial relativamente a uma dada propriedade do estímulo (e.g., localização) é descrita pela média da sua distribuição de probabilidade (associada à distribuição do ruído) e a incerteza dessa estimativa pela variância da distribuição. O MMV combina as estimativas de sinais relativos à mesma propriedade em duas modalidades distintas (e.g., S_A e S_B) através de uma média ponderada:

$$\hat{S}_{AB} = W_A \hat{S}_A + W_B \hat{S}_B ,$$

com \hat{S}_{AB} correspondente à estimativa bimodal resultante da combinação das estimativas \hat{S}_A e \hat{S}_B , e W_A e W_B correspondentes, respetivamente, à ponderação atribuída às modalidades A e B, assumindo valores entre 0 e 1 e mantendo entre si uma relação complementar ($W_A = 1 - W_B$). Para que a estimativa combinada seja ótima, no sentido definido acima de maximização da precisão, o peso de cada modalidade deve ser função da sua fiabilidade relativa, calculada como o inverso da sua variância (normalizada à soma da fiabilidade de todas as componentes, neste caso, $1/\sigma_A^2 + 1/\sigma_B^2$):

$$W_A = \frac{1/\sigma_A^2}{1/\sigma_A^2 + 1/\sigma_B^2} \quad e \quad W_B = 1 - W_A$$

Nas condições pressupostas pelo modelo, este esquema de ponderação assegura que a variância de estimativa combinada é inferior à menor das variâncias unimodais (ou igual, se o peso dessa componente for 1, e o das restantes zero, o que equivale a uma não integração), em virtude da seguinte relação entre variâncias:

$$\sigma_{AB}^2 = \frac{\sigma_A^2 \sigma_B^2}{\sigma_A^2 + \sigma_B^2} \leq \min(\sigma_A^2, \sigma_B^2)$$

A situação em que as duas variâncias componentes (no caso bimodal) são iguais é a mais favorável, assegurando o maior ganho de precisão com a estimativa combinada (redução da variância por um fator de $\sqrt{2}$). Divergências entre as variâncias componentes, por exemplo, no caso em que uma das fontes tem uma fiabilidade muito reduzida, diminuem a magnitude deste ganho de precisão, mas não o anulam (Alais & Burr, 2019) continuando a

indicar um benefício da integração multissensorial (Shams, 2012). Este facto poderia explicar, por exemplo, a ocorrência de integrações audiovisuais em tarefas de localização espacial onde a vantagem de juntar às estimativas visuais estimativas auditivas muito menos fiáveis não é óbvia e pode parecer contraintuitiva. Mesmo no caso de uma correlação entre as distribuições de ruído dos diferentes sinais (violando um dos pressupostos do modelo) este modo de combinação de sinais redundantes resulta ainda numa precisão aumentada da estimativa final (Oruç, Maloney, & Landy, 2003).

O MMV produz assim dois tipos característicos de previsões quantitativas, uma relativa à variância combinada, que deve ser inferior à menor das variâncias componentes, a segunda relativamente à média da estimativa combinada, correspondente a uma média ponderada das estimativas unimodais. Numerosos estudos, envolvendo diferentes tarefas e combinações de modalidades sensoriais, têm apoiado a ideia de que o MMV descreve uma estratégia de integração largamente utilizada pelo sistema perceptivo para minimizar os efeitos do ruído (e.g., Ernst & Banks, 2002; Ghahramani, Wolpert, & Jordan, 1997; Meijer, Veselič, Calafiore, & Noppeney, 2019; van Beers, Sittig, & Gon, 1999). Vários desses estudos têm mostrado, em particular, a adequação do modelo na combinação de informação visual e auditiva (Alais & Burr, 2004; Hillis, Ernst, Banks, & Landy, 2002; Landy & Kojima, 2001). O pressuposto de redundância e não enviesamento dos sinais limita no entanto a aplicabilidade do modelo às situações de ausência de conflito (ou envolvendo apenas pequenas discrepâncias) entre sinais, onde a fusão sensorial se justifica.

Apesar de constituir um modelo frequentemente ajustado à integração sensorial em adultos, o MMV tem-se revelado inadequado em crianças com idades inferiores a 10 anos. Abaixo desta idade, as crianças manifestam uma dominância unissensorial, não parecendo explorar os benefícios estatísticos da integração multissensorial, mesmo quando o sentido dominante é menos preciso do que outros. Uma interpretação possível é a de que, no curso do desenvolvimento, cada sistema sensorial apresenta uma especialização inerente, utilizada no afinamento e calibração de outros sentidos (Alais & Burr, 2019): por exemplo, a visão calibraria a audição para o espaço e a audição a visão para o tempo. Até que este processo intermodal de afinação e calibração recíprocas estivesse concluído no desenvolvimento, a integração multissensorial não teria lugar (ver, e.g., Barutchu, Crewther, & Crewther, 2009; Gori, Del Viva, Sandini, & Burr, 2008) Esta possibilidade coloca questões interessantes quando a modalidade dominante para efeitos de afinamento-calibração se encontra comprometida por défices sensoriais. Por exemplo, quais os efeitos na calibração da audição para o espaço de um défice na modalidade visual? A previsão seria a de um défice da audição

em tarefas espaciais (de forma simétrica, um déficit sensorial na audição deveria refletir-se numa calibração deficitária da visão em tarefas temporais). Vários resultados favoráveis a esta hipótese têm sido reportados (Gori, Sandini, Martinoli, & Burr, 2010; King, 2009). Os resultados documentam igualmente a importância de processamento unissensorial adequado nas fases iniciais do desenvolvimento, mostrando importantes diferenças entre participantes com cegueira congénita e com cegueira adquirida. Num estudo com crianças com surdez submetidas a implante coclear, apenas crianças implantadas abaixo dos 30 meses apresentavam os benefícios típicos da fusão audiovisual na perceção de fonemas (Schorr, Fox, Van Wassenhove, & Knudsen, 2005).

Para além da limitação já notada do MMV a situações nas quais a fusão sensorial se justifica (ausência de conflito entre os sinais), ponderações não ótimas ou subótimas das componentes unissensoriais, inconsistentes com o modelo, têm sido empiricamente observadas (e.g., Arnold, Petrie, Murray, & Johnston, 2019; Meijer et al., 2019).

No estudo de Meijer et al. (2019), por exemplo, envolvendo a integração audiovisual em tarefas de localização espacial, a ponderação dos sinais visuais foi significativamente superior à prevista pelo MMV. Segundo Meijer et al. (2019), este tipo de desvios ao esquema de ponderação do MMV poderão dever-se à incerteza dos sujeitos sobre a estrutura causal do mundo. Battaglia, Jacobs, & Aslin (2003) tinham já sugerido, no mesmo sentido, que a fiabilidade dos sinais sensoriais parece basear-se na experiência dos participantes (na maioria das situações, o sinal visual espacial é muito mais preciso do que o sinal espacial auditivo), algo que não se encontra incorporado no modelo MMV. Meijer et al. (2019) notam ainda, a este propósito, que no estudo de Alais & Burr (2004), favorável ao modelo MMV, os participantes treinaram intensivamente a tarefa de localização auditiva, potenciando através da aprendizagem a fiabilidade dos sinais auditivos. Meijer et al. (2019) sugerem por isso, com base nos seus resultados, que a integração espacial audiovisual para participantes não treinados não é adequadamente descrita pelo modelo MMV, apresentando-se antes consistente com modelos Bayesianos capazes de acomodar a inferência causal.

2.1.3.2. *Modelo Bayesiano de inferência causal*

Um *modelo Bayesiano de inferência causal* da perceção multissensorial corresponde a uma estratégia para arbitrar entre fusão sensorial e segregação da informação, tendo em conta as probabilidades de diferentes cenários causais possíveis (Noppeney, 2021; Rohe, Ehlis, &

Noppeney, 2019). Esta arbitragem torna-se necessária a partir do momento em que se abandona o pressuposto de uma fonte única, utilizado pelo MMV, e se admite a possibilidade de múltiplas causas para múltiplos sinais (Shams & Beierholm, 2010). Como notam Kayser & Shams (2015), o sistema perceptivo enfrenta dois problemas no que se refere ao melhor uso da informação de diferentes modalidades: (1) o problema de como integrá-las, algo a que o MMV se propõe responder normativamente (combinação ótima de pistas incertas); (2) o problema de decidir se deve ou não integrá-las, correspondendo à questão da inferência causal. Dois sinais devem ser integrados quando exprimem uma causa comum (facultam informação sobre um mesmo evento), mas não necessariamente quando exprimem dois eventos causais distintos. Neste contexto alargado, não se trata apenas de estimar uma variável oculta relativa às propriedades de um objeto, mas também às propriedades da estrutura causal do ambiente (Vilares & Körding, 2011).

A introdução do problema da inferência causal equivale a reconhecer que, para além dos benefícios, há também custos potenciais associados à integração (Ernst & Di Luca, 2011). A integração de sinais com enviesamentos sistemáticos relativamente à propriedade de interesse ou relativamente uns aos outros resulta inevitavelmente no enviesamento da estimativa final integrada (ainda que possa aumentar a sua precisão, minimizando a incerteza) (Ernst & Di Luca, 2011). Num contexto caracterizado pela ocorrência de múltiplos sinais sensoriais e potencialmente adverso, a questão de saber se dois ou mais estímulos de modalidades diferentes têm ou não uma causa comum pode ser crítica para um organismo (Colonius & Diederich, 2010). Por exemplo, ao atravessar a rua num dia de nevoeiro seria vantajoso combinar informações auditivas e visuais para estimar a velocidade e direção de um carro que se aproxima. No entanto, combinar a informação auditiva procedente de um carro com a informação visual relativa a um outro carro pode constituir um erro potencialmente fatal (Shams, 2012). Para além dos custos potencialmente incorridos, uma combinação rotineira de pistas sensoriais independente da compreensão das causas da estimulação sensorial, rapidamente conduziria, em ambientes dinâmicos, ao esgotamento dos recursos disponíveis para a integração (Colonius & Diederich, 2010). A ideia de que a primeira função das pistas sensoriais é a de inferir as causas subjacentes dos eventos perceptivos, raramente sendo ecologicamente relevantes em si mesmas, constitui a motivação base dos modelos Bayesianos de inferência causal (Körding et al., 2007).

Uma manifestação de que o sistema perceptivo procura balancear os benefícios e custos da integração multissensorial é o cancelamento da integração observado quando o grau de conflito entre as modalidades se torna importante (Ernst & Di Luca, 2011; Knill, 2007; Roach

et al., 2006). Discrepâncias significativas na localização espacial (e.g., Jack & Thurlow, 1973; Warren & Cleaves, 1971) ou assincronias significativas entre os sinais (e.g., Bresciani et al., 2005; Radeau & Bertelson, 1987; van Wassenhove, Grant, & Poeppel, 2007) impedem a interação-integração intermodal. Este facto assinala que o sistema perceptivo procura minimizar o risco de integrar informação não relacionada tendo em conta pistas como a sincronia temporal ou a coincidência espacial dos sinais (Rohe et al., 2019). No entanto, decidir pela integração *versus* não integração apenas considera dois casos limites (fusão perceptiva *versus* segregação), deixando de lado um espectro de combinações multissensoriais designáveis como integrações parciais (Shams, 2012; Shams & Beierholm, 2010). Nestes casos, documentados em situações em que se verificam discrepâncias moderadas entre sinais (e.g., Shams & Beierholm, 2010, para discrepâncias temporais; Körding et al., 2007, para discrepâncias espaciais), observam-se efeitos de integração graduados pela consistência temporal e/ou espacial dos sinais.

Os modelos Bayesianos de inferência causal procuram acomodar o espectro integral das combinações multissensoriais (fusão, integração parcial e segregação) num quadro de inferência probabilística ótima, associado a um observador ideal (termo de comparação para os resultados empíricos). Trata-se de um problema de estimação difícil por várias razões. As relações causais não são diretamente observáveis e devem ser inferidas a partir da informação sensorial (Vilares & Körding, 2011), mas as estimativas sensoriais produzidas não informam sobre a sua exatidão (ou enviesamento) relativamente à propriedade de interesse (Ernst & Di Luca, 2011). A discrepância entre estimativas sensoriais não indica, do mesmo modo, qual, ou se alguma, está correta. Na realidade, mesmo que dois sinais tenham origem no mesmo objeto/evento, o ruído no ambiente ou no sistema nervoso, combinado com importantes diferenças de precisão entre modalidades (e.g., quanto às dimensões espacial e temporal), origina na maioria das vezes algum grau de inconsistência entre sinais sensoriais (Shams, 2012). Para ultrapassar estas dificuldades, um modelo flexível de inferência causal necessita de considerar não apenas o grau de consistência espacial, temporal e estrutural entre sinais, como o conhecimento prévio disponível sobre a sua probabilidade de coocorrência (construído durante o neurodesenvolvimento ou derivado da experiência, por aprendizagem, a diferentes escalas de tempo) (Mihalik & Noppeney, 2020; Rohe et al., 2019; Shams, 2012). Formalmente trata-se de representar probabilisticamente e articular entre si diferentes tipos de incerteza (sensorial, causal e prévia) num modelo flexível de interação-integração multissensorial (Noppeney, 2021).

Um modelo Bayesiano de inferência causal frequentemente referido como “não-hierárquico” consiste em representar por uma distribuição de probabilidade conjunta *a priori* – $P(S_A, S_B)$, no caso de dois sinais S_A e S_B , associados às modalidades A e B – o conhecimento acumulado sobre a sua coocorrência no meio, combinando-a multiplicativamente com as verossimilhanças de cada um dos sinais. Na medida em que admite valores entre 0 e 1, esta distribuição conjunta *a priori* reflete o grau de interação reconhecido às modalidades no processo de inferência, permitindo graduar o espectro contínuo das “integrações parciais” (para além de obter também os casos limite, fusão e segregação, quando esta probabilidade se torna 1 ou 0, respetivamente) (ver Bresciani, Dammeier, & Ernst, 2006; Roach et al., 2006; Shams, 2012). Esta estratégia de utilização de uma distribuição conjunta *a priori* (generalizável a mais de dois sinais) não estabelece previsões causais explícitas, isto é, que detalhem cada um dos cenários causais possíveis, probabilizem cada um deles e descrevam o modo como suporta a inferência perceptiva final. Não envolve por isso, também, a questão das diferentes estratégias de decisão possíveis perante vários cenários causais. Realizar previsões sobre a estrutura causal percebida requer uma outra classe de modelos de inferência causal, designados como hierárquicos. O traço característico destes modelos é a existência de uma variável adicional (frequentemente denotada C) que especifica os diferentes cenários causais e possibilita, através da sua probabilização, diferentes estratégias de decisão.

No caso mais simples, envolvendo apenas dois sinais e dois cenários causais (causa comum, $C = 1$, ou causas distintas, $C = 2$) a probabilidade de cada cenário pode representar-se como uma probabilidade condicional à semelhança (numa métrica de semelhança especificada) das representações sensoriais dos dois sinais, x_A e x_B , $P(C = 1 | x_A, x_B)$ e $P(C = 2 | x_A, x_B)$. Quando $P(C = 2) = 1$, as estimativas dos sinais deverão ser segregadas, isto é, $\hat{S}_{A,C=2} = x_A$ e $\hat{S}_{B,C=2} = x_B$. Quando $P(C = 1) = 1$ os sinais deverão ser integrados nos termos do MMV (média das estimativas ponderadas pelas sua fiabilidade relativa). Quando $P(C = 1)$ e $P(C = 2)$ são ambos diferentes de zero ($\neq 0$) e inferiores a 1 (< 1) várias estratégias são possíveis, com destaque para as seguintes: (1) se as estruturas causais têm probabilidades diferentes (i.e., com dois cenários, ambos têm uma probabilidade $\neq 0.5$), é possível selecionar a estrutura causal mais provável (estratégia de *seleção do modelo*); (2) combinar as estimativas deriváveis a partir de cada uma das duas estruturas causais (utilizando o MMV no caso $C = 1$ e a segregação no caso $C = 2$) ponderando-as pela probabilidade de cada um dos cenários causais (estratégia de média ou *model averaging*); (3) variação na escolha da estrutura causal ao longo de diferentes ensaios, em que a frequência de escolha de um ou

outro cenário causal é proporcional à sua probabilidade (estratégia de ajustamento às probabilidades) (Mendonça et al., 2016; Rohe et al., 2019; Shams & Beierholm, 2010; Wozny, Beierholm, & Shams, 2010). A investigação com modelos hierárquicos, em particular visando esclarecer as estratégias de decisão baseadas na estrutura causal percebida, é ainda relativamente recente, oferecendo no entanto resultados promissores e várias direções de investigação em aberto (ver, e.g., Körding et al., 2007; Mendonça et al., 2016; Noppeney, 2021; Shams & Beierholm, 2010).

Independentemente do tipo de modelo, os processos de inferência causal associados à produção de estimativas multissensoriais não são concebidas na abordagem Bayesiana como dependendo da consciência e de processos cognitivos de alto nível, e sim como decorrendo em contínuo e sem esforço na percepção (Körding et al., 2007). Shams & Beierholm (2010) assinalam neste sentido a existência de evidência comparativa favorável a uma comunidade evolucionária de estratégias de inferência causal e à contribuição de formas simples de aprendizagem por reforço, que partilhamos com numerosas espécies, para o desenvolvimento da inferência causal na integração multissensorial.

Apesar de várias propostas relativas à implementação neural da inferência causal Bayesiana (Kayser & Shams, 2015; Mihalik & Noppeney, 2020; Rohe et al., 2019), não existe uma compreensão clara dos processos envolvidos. O modo de obtenção das estimativas de fiabilidade (medida e representação da incerteza sensorial) (Burr & Alais, 2006), o modo de geração dos *a priori*, a relação entre *a priori* subjetivos (refletidos em enviesamentos de resposta) e objetivos (refletindo as estatísticas do meio) (Shams & Beierholm, 2010), dúvidas sobre a representação independente dos *a priori* e das verosimilhanças (Beierholm, Quartz, & Shams, 2009; Shams, 2012), os potenciais limites (por exemplo, do ponto de vista do custo computacional e da implementação neural) das estratégias de otimização de um observador ideal (Shams & Beierholm, 2010), a modelização do papel da atenção na integração multissensorial (van Ee, van Boxtel, Parker, & Alais, 2009) constituem outros tantos desafios à abordagem Bayesiana, que se estabeleceu, entretanto, como o quadro interpretativo mais influente e promissor no estudo da interação-integração multissensorial.

2.2. REGRAS E PRINCÍPIOS: JANELA DE INTEGRAÇÃO TEMPORAL, REGRA ESPACIAL, PRINCÍPIO DA EFICIÊNCIA INVERSA

Os estudos pioneiros sobre a integração multissensorial em neurónios do colículo superior (CS) de gatos anestesiados (Meredith, Nemitz, & Stein, 1987; Meredith & Stein, 1983, 1986), depois replicados em primatas não humanos (Wallace & Stein, 1996; Wallace, Wilkinson, & Stein, 1996) estabeleceram, para além das diferenças entre a resposta multissensorial dos neurónios e a soma das suas respostas a cada componente unimodal (revelando frequentemente um efeito de sobreaditividade: mas ver Holmes & Spence, 2006; Stanford & Stein, 2007), um conjunto de princípios relativos a características da estimulação que regulam a integração multissensorial. As respostas neurais integradas aos estímulos sensoriais mostraram ser mais prováveis e/ou intensas quando os estímulos são próximos no tempo (“regra temporal”), próximos no espaço (“regra espacial”) e/ou quando pelo menos um deles apresenta uma reduzida eficácia na indução de uma resposta (“princípio da eficiência inversa”). Dito de outro modo, à medida que aumenta a separação dos estímulos no tempo e no espaço e/ou aumenta a sua capacidade individual de elicitar uma resposta (e.g., através de um aumento de intensidade) a probabilidade e/ou magnitude da resposta multissensorial diminui.

Estes princípios estabelecidos com registos de célula única em animais anestesiados são frequentemente aceites como base para a compreensão da integração multissensorial e dos seus efeitos percetivos e comportamentais em humanos (e.g., Calvert & Thesen, 2004; Stein, Stanford, & Rowland, 2014). No entanto, a transição entre o nível celular e o do organismo inteiro, bem como a complexidade dos contextos de integração com que se deparam os organismos ativos, em particular humanos, aponta para a necessidade de avaliar em cada caso os limites desta generalização, as condições em que pode validamente assumir-se e o detalhe dos processos subjacentes (Sarko et al., 2012; Spence, 2013). Muita evidência favorável à operação destes princípios na combinação de informação multissensorial em humanos tem sido reportada (e.g., Bolognini, Frassinetti, Serino, & Làdavas, 2005; Colonius & Diederich, 2010; Hairston, Laurienti, Mishra, Burdette, & Wallace, 2003; Lewald & Guski, 2003; Sarko et al., 2012; Slutsky & Recanzone, 2001; Soto-Faraco et al., 2002; Stekelenburg & Vroomen, 2009). Contudo, outros resultados têm destacado interações complexas entre os três princípios (Sarko et al., 2012), a sua modulação por processos *top-down*,

designadamente a atenção (e.g., Spence, Shore, & Klein, 2001; Zampini, Guest, Shore, & Spence, 2005), e ainda limites e violações das previsões (e.g., Holmes, 2007; Spence, 2013).

2.2.1. Regra temporal e janela de integração multissensorial

Uma marca durável da origem neurofisiológica da “regra temporal” é a noção de que a coincidência ou estrita proximidade temporal dos sinais sensoriais suporta a sua integração automática através de processos *bottom-up*, guiados pelo estímulo (Lippert, Logothetis, & Kayser, 2007).

A ideia de janelas temporais de integração sensorial (*temporal integration windows: TIWs*) não se prende especificamente com a integração multissensorial. Perceber objetos, movimentos ou eventos requer uma integração de informação durante um intervalo de tempo (Melcher, Wutz, Drewes, & Fairhall, 2014). Dois estímulos visuais separados temporalmente por menos de 40 ms são percebidos como simultâneos; em tarefas de mascaramento retroativo ou de inibição de retorno a interação eficiente entre estímulos é regulada por intervalos temporais mais amplos (200 ms - 300 ms); e janelas temporais mais longas ainda, da ordem dos segundos (2 - 3 s) (Melcher et al., 2014; Pöppel, 2009), ou mesmo minutos (Hasson, Yang, Vallines, Heeger, & Rubin, 2008), podem encontrar-se no contexto da integração de sequências informativas complexas, por exemplo, no domínio da linguagem ou da produção motora. A existência de janelas de integração a diferentes escalas temporais é assim um dado essencial do funcionamento perceptivo, plausivelmente ao serviço do objetivo de segregar o fluxo da informação sensorial em unidades significativas (reportáveis a objetos e eventos).

A noção de janela temporal de integração multissensorial, associada à “regra temporal”, é concebível, analogamente, como um princípio de coordenação entre modalidades na construção de representações coerentes (Körding et al., 2007; Vroomen & Keetels, 2010). Informação intermodal no interior desta janela teria uma alta probabilidade de ser integrada e dar lugar à percepção de um evento multissensorial unificado, enquanto informação fora da janela não é integrada (Colonius & Diederich, 2010). Benefícios obtidos pela apresentação simultânea ou temporalmente próxima de pares de estímulos de diferentes modalidades têm sido amplamente documentados, envolvendo, por exemplo, ganhos de deteção e discriminação perceptivas (e.g., Frassinetti et al., 2002; Staufenbiel, van der Lubbe, & Talsma, 2011; Stein, London, Wilkinson, & Price, 1996; Van der Burg, Olivers,

Bronkhorst, & Theeuwes, 2008), reduções da latência da resposta (Diederich & Colonius, 1987, 2004; Frens, van Opstal, & Van der Willigen, 1995), ou ganhos de exatidão na percepção da fala num contexto de ruído (van Wassenhove et al., 2007).

Uma questão natural a propósito desta janela temporal é a da sua extensão. Uma janela temporal demasiado estreita (no limite zero, limitada à simultaneidade) limitaria drasticamente os benefícios alcançáveis através da integração e uma janela com uma extensão ilimitada tornaria obrigatória a integração e incorreria nos custos associados à integração de eventos não relacionados (Colonius & Diederich, 2010). A ideia de uma extensão característica e invariável desta janela foi proposta por alguns autores (Munhall, Gribble, Sacco, & Ward, 1996), mas é contrária à evidência disponível. Por um lado, as estimativas produzidas empiricamente com diferentes métodos e com diferentes tarefas têm variado significativamente, dos 40 ms aos 600 ms e mesmo até intervalos acima do segundo (Colonius & Diederich, 2010). Por exemplo, no domínio da interação-integração audiovisual, enquanto a assincronia entre um tom simples e um *flash* é notada a partir dos 25 ms ou 50 ms (Zampini, Shore, & Spence, 2003), atrasos de 200 ms a 300 ms entre a informação auditiva e visual na percepção audiovisual da fala, da música ou de eventos físicos complexos, não são frequentemente notados (Eg & Behne, 2015; van Eijk, Kohlrausch, Juola, & van de Par, 2008; van Wassenhove et al., 2007). Por outro lado, tendo em conta a multiplicidade das diferenças de tempo de propagação física das energias (que diferem por várias ordens de magnitude, como no caso da luz e do som) e das latências características de diferentes modalidades, uma janela temporal ajustável a diferentes combinações intermodais seria funcional e ecologicamente mais adequada de que uma janela constante (Sarko et al., 2012). Um exemplo de “flexibilidade ecológica” da janela temporal é a assimetria característica da janela de integração audiovisual, mais estreita quando o som precede o estímulo visual, e cerca de duas vezes maior quando o som ocorre depois do estímulo visual (Cecere, Gross, & Thut, 2016; Conrey & Pisoni, 2006; Dixon & Spitz, 1980; van Wassenhove et al., 2007; Zampini et al., 2003). A direção desta assimetria é consistente com um ajuste da janela ao facto de que a propagação da luz é consideravelmente mais rápida do que a do som, pelo que atrasos na emissão do som implicariam as maiores assincronias (Dixon & Spitz, 1980).

Mais geralmente, a amplitude da janela de integração multissensorial apresenta importantes diferenças interindividuais (Stevenson, Zemtsov, & Wallace, 2012) e mostrou variar de forma significativa, entre outros fatores, com a atenção (Donohue, Green, & Woldorff, 2015), a natureza da tarefa (Mégevand, Molholm, Nayak, & Foxe, 2013), a complexidade dos estímulos (van Wassenhove, Buonomano, Shimojo, & Shams, 2008), a

distância entre os estímulos e o observador (Sugita & Suzuki, 2003), a matriz de consequências de diferentes decisões perceptivas (Colonius & Diederich, 2010), ou a natureza do conflito intersensorial (e.g., requerendo recalibração e/ou adaptação: Di Luca, Machulla, & Ernst, 2009; Ernst & Bühlhoff, 2004; Vroomen et al., 2004). Zhou, Cheung & Chan (2020) resumiram assim algumas das evidências disponíveis sobre a janela temporal de integração audiovisual (JTIAV): (1) a sua extensão pode ser modulada por uma vasta gama de processos descendentes (*top-down*) e ascendentes (*bottom-up*); (2) a JTIAV mobiliza uma extensa rede cerebral que funciona em coordenação para garantir a detecção bem-sucedida das (as)sincronias audiovisuais; (3) o padrão de evolução da sua amplitude ao longo da vida segue uma forma em U (alargamento até ao final da adolescência, estreitamento em idades intermédias e de novo alargamento com o envelhecimento); (4) apresenta-se alterada em quadros de défice sensorial (e.g., visual, auditivo), perturbações do neurodesenvolvimento (e.g., do espectro do autismo, da linguagem, de hiperatividade/défice de atenção, dislexia), neuropsiquiátricas e mentais (e.g., esquizofrenia, comprometimento cognitivo ligeiro) neurodegenerativas (e.g., doença de Parkinson, doença de Alzheimer), e ainda noutras condições clínicas (e.g., obesidade, traumatismo craniano, eventos psicogénicos não epiléticos); (5) é altamente flexível, diminuindo significativamente a sua largura com o treino perceptivo e musical (com eventuais implicações para a intervenção em populações clínicas).

Um modelo promissor da flexibilidade da janela temporal de integração no domínio audiovisual foi proposto por Colonius & Diederich (2010). Este modelo estende a lógica Bayesiana de inferência causal, combinando conhecimento *a priori*, verosimilhanças associadas à evidência sensorial (no caso, diferenças entre o tempo de chegada dos sinais) e regras de decisão perceptiva, à previsão de uma janela temporal ótima (adaptada às circunstâncias) de integração audiovisual. Propostas sobre a relação entre atividade oscilatória dos córtices sensoriais e a modulação da janela ótima de integração audiovisual têm sido formuladas (Schroeder, Lakatos, Kajikawa, Partan, & Puce, 2008; ver também, para alguma evidência favorável, Rohe et al., 2019), oferecendo suporte adicional ao modelo de Colonius & Diederich (2010). Por outro lado, um número apreciável de estudos tem sido dedicado nos últimos anos à compreensão do modo como o cérebro lida com assincronias temporais entre os sentidos. Baseados em várias evidências, Vroomen & Keetels (2010) descreveram quatro mecanismos possíveis, que consideram não exclusivos: (1) ignorar pequenos atrasos (janela de integração temporal); (2) utilizar conhecimento prévio para compensar variações previsíveis; (3) adaptar a longo termo o critério de sincronia percebida (recalibração temporal); (4) deslocar o *timing* percebido de um dos sinais em direção a outro

(“captura” temporal). Embora de forma tentativa, o modelo de janela ótima de integração temporal (Colonius & Diederich, 2010) abre perspectivas para uma articulação possível do conjunto destes mecanismos num quadro unificado.

2.2.2. Regra espacial

De acordo com a regra espacial, a interação-integração multissensorial é mais forte ou prevalente quando os estímulos de diferentes modalidades sensoriais coincidem espacialmente, ou são apresentados em locais próximos (Holmes, 2009; Spence, 2013). Na realidade, no domínio neurofisiológico, onde começou por surgir (Meredith & Stein, 1986), a regra espacial não se refere exatamente à coincidência dos estímulos num mesmo ponto do espaço, mas sim numa região onde os campos recetores das células estimuladas se sobrepõem (Meredith & Stein, 1996). Uma complexidade adicional, ainda associada à definição da regra espacial em termos neurofisiológicos, envolve o facto de que os neurónios multissensoriais codificam estímulos de diferentes modalidades em diferentes quadros de referência, implicando diferentes alterações dos campos recetores em resposta à alteração da direção do olhar/cabeça/membros (Wallace, Perrault Jr, Hairston, & Stein, 2004). Holmes & Spence (2005) definem neste sentido a regra espacial como o resultado de um alinhamento dos mapas espaciais das diferentes modalidades sensoriais no colículo superior (CS), uma estrutura essencial na geração e controlo dos movimentos oculares e da cabeça. Representar, em função de uma dada orientação do olhar, uma mesma região do espaço para diferentes modalidades numa pequena região do colículo tornaria possível a integração de informação associada ao local do espaço mais estimulado e, potencialmente mais importante num dado momento (Holmes & Spence, 2005). Esta definição suscita naturalmente a dúvida sobre se a regra espacial tem aplicabilidade em situações que não envolvam respostas de orientação ou localização, ou dependentes, do ponto de vista neuroanatômico, de outras estruturas de processamento multissensorial que não o colículo superior (Holmes & Spence, 2005).

A evidência disponível parece justificar esta dúvida. Um número considerável de estudos suporta, de acordo com a regra espacial, a necessidade de co-localização dos diferentes sinais sensoriais para que ocorra integração multissensorial (e.g., Feng, Störmer, Martinez, McDonald, & Hillyard, 2014; Frassinetti et al., 2002; Meyer, Wuerger, Röhrbein, & Zetzsche, 2005; Sambo & Forster, 2009) e, em vários casos, a previsão de um declínio da

interação multissensorial com o aumento da separação espacial entre os estímulos (e.g., Harrington & Peck, 1998; Lewald, Ehrenstein, & Guski, 2001). No entanto, muitos outros estudos produziram resultados contrários à regra espacial, apontando para a irrelevância do alinhamento espacial entre sinais (e.g., Fiebelkorn, Foxe, Butler, & Molholm, 2011; Van der Burg et al., 2008; Vroomen & Keetels, 2006). De um modo geral, a investigação envolvendo o efeito McGurk e a percepção audiovisual da fala, o ventriloquismo temporal, a recalibração temporal audiovisual, ou os benefícios multissensoriais na percepção do brilho, tem revelado a ausência de diferenças devidas à co-localização *versus* separação espacial dos sinais (L. Chen & Vroomen, 2013; Spence, 2013).

Com base no exame de várias dezenas de estudos favoráveis e desfavoráveis à regra espacial, Spence (2013) encontrou as seguintes regularidades. A larga maioria dos estudos favoráveis requerem ou produzem uma resposta de orientação, mobilizam a atenção espacial (por orientação exógena ou por seleção endógena) e/ou envolvem tarefas nas quais o espaço constitui uma dimensão relevante. Por sua vez, a maioria dos estudos que não requerem respostas de orientação, envolvem a detecção ou identificação de um estímulo-alvo, ou se dirigem a aspetos temporais dos estímulos, apresentam resultados inconsistentes com a regra espacial. O contraste entre as janelas espaciais e temporais de integração audiovisual no ventriloquismo espacial (onde o espaço constitui o atributo relevante para a tarefa) e no ventriloquismo temporal (envolvendo o tempo como dimensão relevante para a tarefa) fornecem uma ilustração clara desta divisão. O ventriloquismo espacial apresenta uma janela temporal de integração assimétrica entre os - 100 ms (som precede) e os + 300 ms (a visão precede) e uma tolerância a separações espaciais horizontais (janela espacial de integração) de $\approx 15^\circ$. No ventriloquismo temporal a janela temporal é algo mais estreita (< 200 ms) e a disparidade espacial não parece ter efeito, isto é, a janela espacial é muito mais ampla ou mesmo inexistente (L. Chen & Vroomen, 2013).

O conjunto dos fatores que determinam a aplicabilidade de regra espacial não está completamente esclarecido, como mostram a existência de exceções ao princípio de que tarefas espaciais suportam a regra e tarefas não-espaciais não (Spence, 2013). Orientações exógenas ou endógenas da atenção espacial, por exemplo, podem converter tarefas não-espaciais em tarefas implicitamente espaciais (Spence, 2013). A natureza das modalidades envolvidas poderá constituir um outro fator (e.g., interações audiovisuais poderão ter um carácter intrinsecamente menos espacial do que interações entre a visão e o tato: Girard, Pelland, Lepore, & Collignon, 2013). Alguma evidência indica ainda que a regra espacial tem maior probabilidade de vigorar em regiões do espaço associadas a respostas de orientação

defensivas (como a região atrás da nuca no caso de estimulação áudio-tátil: Occelli, Spence, & Zampini, 2011). Em qualquer caso, o conjunto dos resultados sugere robustamente que, longe de constituir um princípio geral da integração multissensorial, a regra espacial é fortemente dependente da tarefa, sendo suportada em situações em que o espaço é de alguma forma um atributo relevante para a tarefa (Spence, 2013).

2.2.3. Princípio da eficiência inversa

O princípio da eficiência inversa estipula que à medida que a capacidade de resposta aos estímulos sensoriais individuais (unissensoriais) diminui, a probabilidade e/ou a magnitude da integração multissensorial aumenta proporcionalmente (Holmes, 2009). Por outras palavras, a magnitude da interação-integração intermodal e o benefício multissensorial são maximizados quando as respostas às entradas unimodais são mais fracas (Stein & Meredith, 1993). Por exemplo, no caso da integração audiovisual, se existe um declínio no desempenho unimodal (visual ou auditivo) a integração torna-se mais provável e o benefício multissensorial é superior (Tye-Murray, Sommers, Spehar, Myerson, & Hale, 2010). Diferentemente das duas regras anteriores, que se referem às propriedades espaço-temporais da associação entre estímulos (envolvendo a noção de janelas de tolerância a discrepâncias temporais e espaciais), a eficiência inversa relaciona a interação-integração multissensorial com as propriedades de intensidade ou a razão sinal-ruído dos estímulos unimodais (Bolognini, Senna, Maravita, Pascual-Leone, & Merabet, 2010). O sentido “compensatório” deste princípio, apontando para uma maior probabilidade de utilização de outras fontes de informação no caso da degradação da qualidade de uma das fontes disponíveis, torna-o particularmente significativo para a abordagem das interações multissensoriais em contextos de déficit ou declínio sensorial (Tye-Murray et al., 2010).

À semelhança do sucedido com a regra espacial, a generalidade da “eficiência inversa” enquanto princípio da integração tem sido questionada. Empiricamente, a par de estudos favoráveis (e.g., Bolognini, Frassinetti, et al., 2005; Hairston, Laurienti, et al., 2003; Lippert et al., 2007; van de Rijt, Roye, Mylanus, van Opstal, & van Wanrooij, 2019), alguns outros têm reportado resultados contrários, com os maiores benefícios em níveis intermédios (e não nos mais baixos) da razão sinal-ruído (Ma, Zhou, Ross, Foxe, & Parra, 2009; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007), ou limites ao seu domínio de aplicação (e.g., a gamas de intensidade estimular próximas do limiar: Stanford, Quessy, & Stein, 2005). Interações

importantes entre a localização espacial do estímulo e o princípio da eficiência inversa foram também documentadas, mediadas por diferenças da resposta neural unissensorial em função da localização do estímulo no campo recetor (Sarko et al., 2012). Estes resultados sugerem que a zona de integração multissensorial ótima no caso da intensidade ou razão sinal-ruído dos sinais poderá depender de características da tarefa, do estímulo e da resposta, e não do nível mais baixo de eficiência do estímulo.

Mais problemática tem sido a dependência demonstrada da verificação da “eficiência inversa” relativamente aos métodos analíticos e estatísticos utilizados para avaliar o ganho multissensorial (Holmes, 2007). Em geral, quando a integração multissensorial é expressa como uma percentagem relativa às linhas de base multissensoriais, o princípio da eficiência inversa tende a ser suportado. No entanto, quando é quantificada em termos de melhoria absoluta (e.g., ganho absoluto de exatidão, ou número de disparos – *spikes* – neuronais), a integração não ilustra a eficiência inversa (Holmes, 2007; ver também, Stanford & Stein, 2007; Stein, Stanford, Ramachandran, Perrault, & Rowland, 2009). Este facto lança dúvidas sobre a própria consistência da noção.

Uma tendência crescente nos últimos anos, associada ao objetivo de detalhar os mecanismos (mais do que os princípios) da integração multissensorial, tem sido a de acomodar a eficiência inversa, não como uma regra geral, mas como uma configuração de resultados possíveis, no quadro de modelos Bayesianos de ponderação ótima das pistas (Angelaki, Gu, & DeAngelis, 2009; Ma et al., 2009). Estes modelos, assentes em combinações lineares ótimas, têm-se revelado capazes de arbitrar as transições entre os padrões de sobreaditividade (resposta multissensorial superior à soma das respostas dos componentes unissensoriais) classicamente associados à eficiência inversa e os padrões de aditividade e subaditividade também frequentemente observados, oferecendo assim uma interpretação possível para o conflito entre resultados (Angelaki et al., 2009).

2.3. INTEGRAÇÃO MULTISSENSORIAL E ATENÇÃO: ATENÇÃO INTERMODAL EXÓGENA E ENDÓGENA

As relações entre a integração multissensorial e a atenção constituem um tópico controverso. São vários os fatores que contribuem para isso, a começar pela convicção difundida de que a integração multissensorial constitui um processo automático, dispensando o envolvimento

da atenção. Para além das demonstrações de integração multissensorial em animais anestesiados (Stanford et al., 2005), diversos resultados parecem suportar esta convicção. Por exemplo, o ventriloquismo audiovisual (deslocação da localização percebida de um som em direção ao local de apresentação de um estímulo visual simultâneo) tem-se revelado imune a efeitos da atenção espacial (Bertelson, Vroomen, de Gelder, & Driver, 2000; Vroomen, Bertelson, & De Gelder, 2001) e induz, do ponto de vista neurofisiológico, uma componente pré-atencional do potencial evocado (*mismatch negativity*) comparável à de um som efetivamente deslocado (Stekelenburg, Vroomen, & de Gelder, 2004). Utilizando um quadro Bayesiano para quantificar a probabilidade de uma causa comum no ventriloquismo audiovisual, Odegaard, Wozny, & Shams (2016) mostraram que a tendência para a integração multissensorial não variou entre condições de atenção focal dirigida a uma das modalidades ou dividida entre as modalidades. No entanto, muitos outros resultados suportam, em sentido inverso, um papel da atenção na integração multissensorial. Por exemplo, o efeito McGurk desvanece-se quando a atenção é desviada para uma tarefa secundária (Alsius, Navarra, & Soto-Faraco, 2007). Talsma & Woldorff (2005) documentaram efeitos da atenção seletiva em potenciais evocados cerebrais relacionados com a integração multissensorial. Modulações atencionais da resposta do colículo superior e dos córtices sensoriais primários por estímulos audiovisuais foram registadas em estudos imagiológicos (Fairhall & Macaluso, 2009). Contrariando estudos anteriores, Talsma, Doty, & Woldorff (2007) reportaram efeitos de facilitação da integração pela atenção intermodal, em contraste com a atenção seletiva a uma das modalidades (ver também Mozolic, Hugenschmidt, Peiffer, & Laurienti, 2008; Mozolic, Joyner, et al., 2008; Oruc et al., 2008). Fatores dependentes da natureza da tarefa e instruções mostraram também influenciar a integração audiovisual através da mediação da atenção (Donohue, Green, & Woldorff, 2015; ver Talsma et al., 2010; Tang et al., 2016). Tomados em conjunto, estes resultados à primeira vista inconsistentes sugerem uma relação complexa entre a atenção e a integração sensorial (ver Macaluso et al., 2016).

Um segundo fator adverso ao esclarecimento da relação entre os dois processos é o facto de cada um deles mobilizar múltiplos mecanismos, associados a diferentes estádios de processamento e níveis heterogéneos da hierarquia cortical (I. Choi, Lee, & Lee, 2018; Koelewijn, Bronkhorst, & Theeuwes, 2010; Macaluso & Driver, 2005; Talsma et al., 2010). As interações entre ambos podem assim ocorrer a diversos níveis e assumir diferentes formas, dependendo do nível de processamento a que tem lugar a integração (Koelewijn et al., 2010; Werner & Noppeney, 2010). A ausência de uma compreensão detalhada dos numerosos

subprocessos envolvidos no controle atencional e na integração intermodal dificulta neste sentido a compreensão da natureza das suas interações, contribuindo para a perpetuação do debate (Macaluso et al., 2016).

A questão da interação entre a integração multissensorial e a atenção tem sido colocada no quadro do problema das relações entre o processamento *bottom-up*, guiado pelos dados, e *top-down*, dependente da tarefa e de objetivos (Macaluso et al., 2016). A tese de uma integração multissensorial automática (pré-atencional) é compatível com uma relação unidirecional entre integração e atenção. Com efeito, eventos multissensoriais integrados têm-se revelado mais eficientes na captura exógena (i.e., guiada de forma *bottom-up* pela saliência dos dados) da atenção do que as suas componentes unimodais, mesmo em situações complexas (Tang et al., 2016). Estímulos de diferentes modalidades apresentados em alinhamento temporal e espacial são mais facilmente selecionados para processamento posterior (Stein, Stanford, Wallace, Vaughan, & Jiang, 2004; Van der Burg et al., 2008; Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2009), assinalando uma orientação preferencial da atenção para entradas com propriedades multissensoriais. Nesta perspetiva, a integração multissensorial constitui uma fonte de controle atencional dirigido-pelo-estímulo, participando ativamente na seleção da informação e, dessa forma, no controle posterior do comportamento (duas das funções usualmente cometidas à atenção).

Um primeiro nível do debate entre diferentes campos surge quando se considera a questão simétrica de saber se a atenção endógena (também dita voluntária, dependente de objetivos, ou *top-down*) pode, por sua vez, desempenhar um papel na seleção da informação a integrar, limitando ou favorecendo a integração multissensorial. Muita da evidência recolhida no estudo da atenção unimodal, em particular visual e auditiva, e da atenção seletiva espacial, sugere essa possibilidade. Kastner, Pinsk, De Weerd, Desimone, & Ungerleider (1999) mostraram, por exemplo, que dirigir a atenção para uma zona do espaço na ausência de estimulação visual, aumenta a atividade do córtex visual, indicando uma modulação do sinal neural pela atenção endógena (ver também Reynolds & Chelazzi, 2004). Na medida em que a estimulação multissensorial implica a ativação dos córtices sensoriais das modalidades envolvidas, este resultado permite conceber que, além da contribuição do sinal sensorial, a atividade nessas áreas possa igualmente refletir contributos atencionais endógenos e veiculá-los para a integração. Num sentido convergente, Vercillo & Gori (2015) observaram que atender seletivamente a um som no quadro de uma estimulação áudio-tátil modula a integração intermodal, aumentando o peso do sinal auditivo na integração. A atenção espacial, por sua vez, parece tender a operar de forma supramodal ou, pelo menos,

articulada entre modalidades. Atender a um local específico numa dada modalidade não apenas aumenta a eficiência do processamento de um estímulo dessa modalidade, como de estímulos de outras modalidades (irrelevantes para a tarefa) apresentados nesse local (e.g., Koelewijn, Bronkhorst, & Theeuwes, 2009; Santangelo, Belardinelli, Spence, & Macaluso, 2009). Numerosos estudos comportamentais e neurofisiológicos ilustram atualmente a possibilidade da atenção *top-down* modular a integração multissensorial, seja através do controlo do ganho dos sinais sensoriais a diferentes níveis da análise sensorial (incluindo os mais precoces), seja através duma articulação espacial “supramodal” das modalidades (e.g., Eimer & Driver, 2001; Fairhall & Macaluso, 2009; Macaluso, Frith, & Driver, 2002; Talsma et al., 2007; Talsma & Woldorff, 2005). O debate diz assim menos respeito, presentemente, à existência ou não de influências bidirecionais entre a integração multissensorial e a atenção, do que à forma de conceber essas influências, tendo em conta os resultados conhecidos e as incertezas sobre os mecanismos de detalhe subjacentes a ambos os processos. Um modelo integrativo, de natureza Bayesiana, foi proposto por Rohe & Noppeney (2018), baseado largamente em resultados obtidos no domínio audiovisual. O modelo assenta na ideia de um processo continuado de combinação de sinais derivados da estimulação sensorial, por um lado, e endógenos (internos: e.g., conhecimento *a priori*, objetivos), por outro, com vista à seleção de informação relevante e controlo do comportamento. O papel da atenção neste processo depende da complexidade dos sinais sensoriais, do contexto da apresentação e da sua forma de representação, associadas por sua vez ao envolvimento de diferentes estruturas cerebrais. Enquanto o córtex superior e áreas sensoriais primárias podem integrar automaticamente sinais sensoriais (e.g., visual e auditivo) representados localmente, de modo a amplificar a sua saliência *bottom-up*, fundamental em tarefas de deteção, o sulco intraparietal produz representações multissensoriais espaciais ou semânticas, fundamentais em tarefas de estimação, discriminação ou categorização, a partir da integração dos sinais ponderados pela sua fiabilidade (Macaluso et al., 2016; Rohe & Noppeney, 2018; Werner & Noppeney, 2010). Embora a atenção endógena possa ter um papel mais prevalente neste último caso, o modelo prevê a possibilidade de uma modulação atencional descendente (*top-down*) do processo em qualquer dos seus níveis. A atenção pode assim: (1) suportar a segregação de diferentes sinais (e.g., um sinal auditivo) a partir de um fundo confuso e ruidoso de estimulação, uma condição necessária para a sua integração com sinais de outras modalidades. A maior ou menor necessidade de atenção a este nível depende da complexidade dos estímulos e do contexto da apresentação (em situações de segregação simples, propriedades como a coincidência temporal podem suportar a integração de forma

mais automática); (2) a atenção a uma modalidade ou a atenção espacial podem ampliar a saliência e fiabilidade dos sinais, modulando a sua ponderação na integração multissensorial; (3) no quadro de um modelo de inferência causal, a relevância para a tarefa de uma dada modalidade (traduzida numa prioridade atencional) influencia a decisão de combinar ou não as estimativas segregadas de cada sinal (e.g., auditivo e visual) com uma estimativa (audiovisual) por fusão forçada; (4) por fim, a atenção específica a uma modalidade (e.g., dirigida à localização de um som no quadro de uma apresentação audiovisual) pode contribuir para reduzir a assunção de uma fonte comum aos dois sinais, reduzindo o peso da estimativa resultante da fusão multissensorial e aumentando o peso das estimativas unissensoriais na inferência causal (Lewis & Noppeney, 2010; Rohe & Noppeney, 2018).

Uma segunda proposta integrativa foi avançada por Talsma et al. (2010). Neste quadro interpretativo, o fator determinante da prevalência de mecanismos *bottom-up* ou *top-down* na integração multissensorial é o grau de competição entre estímulos em cada uma das modalidades. Quando o grau de competição na modalidade de estimulação irrelevante é baixo (e.g., um estímulo auditivo isolado é apresentado numa tarefa em que a visão é a modalidade relevante) a integração com estímulos temporal e/ou espacialmente alinhados na modalidade relevante tenderá a ocorrer pré-atencionalmente (orientada pelos dados). Quando múltiplos estímulos em ambas as modalidades competem por recursos de processamento (alta competição), a atenção *top-down* torna-se necessária para que a integração entre estímulos possa ocorrer (Talsma et al., 2010). Esta proposta foi entretanto inscrita num modelo de codificação preditiva (Talsma, 2015) que situa a atenção num processo continuado de revisão de um modelo estocástico endógeno do ambiente (gerador de previsões) com base na informação sensorial (informação sobre o erro/ajuste das previsões), parcialmente consistente com o espírito da proposta de Rohe & Noppeney (2018).

Talsma et al. (2010) detalham três formas principais da interação entre a integração multissensorial e a atenção. (1) *Influências guiadas pelos dados da integração multissensorial na atenção*: um estímulo saliente (baixa competição) numa modalidade fortalece a representação de um estímulo coocorrente noutra modalidade, contribuindo para que se segregue do fundo e assim capture a atenção. Para além da baixa competição por recursos, o fator determinante aqui é a relação temporal entre os estímulos associados. (2) *Influência da atenção top-down no processamento multissensorial*: quando padrões de estimulação complexos ocorrem em duas modalidades, a atenção endógena é necessária; a atenção espacial tem aqui um papel de relevo, sendo a co-localização espacial de estímulos de

diferentes modalidades um fator determinante. (3) *Difusão intermodal da atenção*: quando um estímulo visual é atendido (mas não quando não é atendido) o processamento de um estímulo auditivo irrelevante apresentado simultaneamente num outro local é aumentado. Este resultado, comparável ao da difusão da atenção dedicada a uma característica de um objeto visual a outras características irrelevantes do mesmo objeto (difusão intramodal), aponta para uma difusão da atenção através das modalidades a um estímulo irrelevante temporalmente coincidente (ver Busse, Roberts, Crist, Weissman, & Woldorff, 2005; Talsma et al., 2007).

A difusão intermodal da atenção sugere assim que, para além da atenção espacial, à modalidade, ou a dimensões específicas de uma modalidade, uma forma de atenção baseada no objeto pode também interagir com a integração multissensorial (Busse et al., 2005; Fiebelkorn, Foxe, & Molholm, 2010; Molholm, Martinez, Shpaner, & Foxe, 2007). Fiebelkorn et al. (2010) propuseram neste contexto uma distinção entre difusão intermodal da atenção guiada-pelos-dados e guiada-por-representações. A primeira, também designada como “baseada-no-objeto”, assenta numa associação automática (*bottom-up*) dos estímulos determinada pela sua coocorrência temporal (a difusão ocorre através do objeto intermodal emergente dessa associação). Esta modalidade de difusão atencional é consistente com a proposta de Tang et al. (2016) de que, num objeto multissensorial, a atenção endógena (e.g., atenção prestada a uma característica sensorial relevante numa modalidade) pode difundir-se duma modalidade para outra de maneira exógena. A difusão guiada por representações, por sua vez, assenta em associações de longo prazo entre propriedades multissensoriais de objetos/eventos conhecidos. A integração dos sinais, tal como a difusão da atenção entre modalidades, depende assim menos das propriedades espaço-temporais da estimulação do que destas representações (mecanismo *top-down*). Um aspeto importante da difusão intermodal da atenção, em qualquer das formas distinguidas por Fiebelkorn et al. (2010), é a possibilidade oferecida de calibrar atencionalmente um estímulo relativo a uma propriedade determinada (e.g., localização) numa dada modalidade, por uma propriedade endogenamente selecionada como relevante (e.g., *timing*) numa outra modalidade (ver, para uma ilustração possível, van Ee et al., 2009; os autores sugerem uma relação com neurónios nos córtices parietal posterior, pré-frontal superior e temporal superior que parecem coordenar funções atencionais voluntárias através das modalidades sensoriais: Degerman et al., 2007).

Apesar da ausência de consenso sobre o detalhe das interações entre a integração multissensorial e a atenção, a existência de relações bidirecionais complexas entre os dois

processos, combinando variadamente mecanismos exógenos e endógenos através de diferentes estádios de processamento e circuitos corticais, não parece oferecer dúvidas. A dependência da extensão do envolvimento da atenção relativamente a um conjunto amplo de fatores (e.g., natureza e objetivos da tarefa, modalidades envolvidas, características do estímulo e do contexto de apresentação, tipo de representação, região cortical envolvida) parece igualmente estabelecida. Os modelos de inspiração Bayesiana ou similar (e.g., codificação preditiva) parecem também aqui oferecer, presentemente, os quadros de interpretação mais promissores (Macaluso et al., 2016).

2.4. ASSUNÇÃO DE UNIDADE E A NOÇÃO DE PERCEPTO AUDIOVISUAL

2.4.1. Problemas com a assunção de unidade

O potencial papel modulador de fatores *top-down* na interação-integração multimodal em humanos tem sido alvo de interesse prolongado (Y.-C. Chen & Spence, 2017; Vatakis & Spence, 2007). Um desses fatores, referido na literatura como *assunção de unidade* (Welch & Warren, 1980), refere-se à suposição ou convicção do observador de que dois ou mais sinais unissensoriais têm origem no mesmo objeto ou evento, favorecendo a sua integração num mesmo evento perceptivo (efeito de unidade).

No seu modelo da integração multisensorial, Welch & Warren (1980) distinguem entre fatores associados ao estímulo (com destaque para as relações temporais e espaciais entre os sinais, mas contemplando também, por exemplo, as propriedades de movimento, orientação, forma ou textura capazes de suportar agrupamentos perceptivos), às modalidades (e.g., exatidão e precisão, latências, limites de capacidade) e ao observador, situando a assunção de unidade entre estes últimos (a par da experiência prévia e da alocação da atenção). Esta distinção reflete o reconhecimento comum de que a decisão de integrar ou segregar estímulos de diferentes modalidades depende não apenas de fatores de baixo nível (*bottom-up*), como as contingências espaciais e temporais (Calvert et al., 2004), mas também de fatores cognitivos de alto nível (*top-down*), de que a assunção de unidade seria um exemplo. Enquanto durante várias décadas a investigação se concentrou nos fatores de baixo nível (Stein, 2012), em particular a coincidência espaço-temporal, a última década e meia assistiu a um interesse crescente por fatores de natureza *top-down* (Y.-C. Chen & Spence,

2017), como a congruência semântica (e.g., imagem de um cão acompanhada de um latido) (Doehrmann & Naumer, 2008), as correspondências intermodais (e.g., associações entre variações de altura tonal de um som e a direção ascendente/descendente de um movimento) (Spence, 2011) e a assunção de unidade (Vatakis & Spence, 2007).

Uma questão relevante para a discussão do contributo da assunção de unidade para a integração multissensorial é a da sua relação com os restantes fatores, sejam de baixo ou de alto nível. Com efeito, pelo menos parte da decisão de que nos encontramos perante um evento multimodal (e não diferentes eventos unimodais) deve basear-se na consistência entre informações fornecidas por cada modalidade sensorial (Welch, 1999; Welch & Warren, 1980). O alinhamento temporal e espacial dos sinais em diferentes modalidades, por exemplo, constitui um importante determinante da integração sensorial (Spence, 2007). Ao mesmo tempo, é também uma fonte reconhecida da assunção de unidade, contribuindo para a convicção de que diferentes sinais sensoriais têm uma origem comum. Do mesmo modo, a congruência semântica dos sinais constitui um fator reconhecido de integração intermodal: estímulos semanticamente congruentes são mais facilmente integrados do que estímulos semanticamente incongruentes (Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004). Ao mesmo tempo, porém, é também reconhecida como uma fonte plausível da assunção de unidade (Y.-C. Chen & Spence, 2017). Esta possibilidade de uma assunção de unidade induzida por características dos estímulos foi traduzida por Welch & Warren (1980) na ideia de “situação estimular persuasiva” (*compelling*), definida, algo circularmente, como uma situação em que a assunção de unidade é forte (Warren, Welch, & McCarthy, 1981).

Não é neste sentido claro o que a assunção de unidade acrescenta ao conjunto dos outros fatores de integração multissensorial (Y.-C. Chen & Spence, 2017). Conceitualmente é possível manter a distinção entre uma assunção de origem comum e uma mera associação entre sinais. Neste sentido, a correlação espaço-temporal entre dois sinais não implica necessariamente uma origem comum, devendo o sistema percetivo ser capaz de distinguir entre as duas situações (Vatakis & Spence, 2007). Essa distinção poderia contudo obter-se, sem necessidade de invocar uma assunção de unidade independente, através do contributo conjugado de fatores como a congruência semântica ou as correspondências intermodais, capazes também de modular aspetos espaciais e temporais da integração (e.g., Laurienti et al., 2004; Parise & Spence, 2009). Pode argumentar-se que a coerência semântica (em especial nas situações artificiais de laboratório) não acarreta também ela, por si só, uma assunção de unidade, que se manteria assim como uma componente independente. No

entanto, continua a não ser claro de que modo a assunção de unidade se distingue funcionalmente da operação dos seus fatores determinantes (fontes de *compelligness*).

Y.-C. Chen & Spence (2017) propõem por isso que a definição da assunção de unidade passe a incluir qualquer fator conducente a uma crença do observador de que dois ou mais estímulos têm uma origem comum. Avaliar empiricamente o seu papel na integração, nos termos desta definição, consistiria em especificar claramente os fatores geradores da impressão de unidade e quantificar a influência de cada um na formação dessa impressão. Da revisão que conduzem, nesta perspetiva, dos resultados disponíveis na literatura, Y.-C. Chen & Spence (2017) retiram conclusões cautelosas. Nenhum panorama claro, permitindo identificar, e menos ainda prever, as condições e formas de emparelhamento estimular que originam uma assunção de unidade (em contraste com as que o não fazem), parece possível. A indução explícita da assunção por meio de instruções do experimentador (uma forma clássica de manipulação desta assunção) não parece suficiente, mostrando-se dependente, por exemplo, de propriedades como a sincronia dos sinais (Warren et al., 1981). Em estudos no domínio das interações visuo-táteis, a integração multissensorial mostrou depender mais da possibilidade de os participantes observarem os movimentos exploratórios da sua própria mão do que do conhecimento de que viam e tocavam o mesmo objeto (assunção de unidade) (Misceo, Jackson, & Perdue, 2014). Evidência favorável a um papel da assunção de unidade na integração multissensorial foi encontrada em estudos de ventriloquismo espacial (e.g., Kanaya & Yokosawa, 2011), mas não em estudos de ventriloquismo temporal (encontrando-se para além disso limitada, no primeiro caso, a estímulos audiovisuais de fala) (Y.-C. Chen & Spence, 2017). O efeito Colavita de dominância visual mostrou obedecer às regras temporal e espacial da integração multissensorial e ser virtualmente imune a efeitos da assunção de unidade (Stekelenburg & Keetels, 2016). A dúvida sobre a necessidade de admitir a assunção de unidade como um fator independente, a par de outros, da integração multissensorial, mantém-se assim em aberto.

Uma possibilidade favorecida por abordagens recentes, no quadro dos modelos Bayesianos de inferência causal, consiste em reconverter e inscrever a assunção de unidade num parâmetro de probabilidade *a priori* que formaliza o conhecimento do observador sobre a probabilidade de os sinais terem uma origem comum (P_{comum}) (Körding et al., 2007; Odegaard & Shams, 2016; Rohe & Noppeney, 2018). Esta proposta apresenta vantagens. Em primeiro lugar, a de graduar e quantificar os efeitos classicamente atribuídos à assunção de unidade (possibilitando também uma avaliação empírica do impacto de diferentes fatores na probabilidade de causa comum) (Y.-C. Chen & Spence, 2017). Em segundo lugar, e mais

importante, a de incluir os efeitos de unidade num processo continuado de revisão do conhecimento prévio (obtido por diferentes meios, a diferentes escalas temporais e com diferentes graus de generalidade) à luz da evidência sensorial, assente unicamente na assunção de um mundo externo dotado de estruturas causais e estatísticas incertas, mas especificáveis.

2.4.2. O que é um percepto audiovisual?

As dificuldades ilustradas pela noção de assunção de unidade são acima de tudo um sintoma de dificuldades na definição do que é um objeto multissensorial (Y.-C. Chen & Spence, 2017). A distinção anteriormente referida (e.g., Bertelson & Radeau, 1981; Schmiedchen et al., 2012; Spence et al., 2009) entre interação intermodal (influência do processamento de um modalidade sensorial no processamento de uma outra, não implicando fusão perceptiva dos sinais) e integração multissensorial (envolvendo um percepto multissensorial unificado, correspondente a um evento externo comum) mobiliza neste sentido tanto a assunção de unidade como uma conceção implícita do objeto perceptivo. Kubovy & Van Valkenburg (2001) assinalaram a natureza visuocêntrica da conceção usual de objeto perceptivo, marcada pela noção de corpos materiais compactos, localizados e opacos (ver também, e.g., O'Callaghan, 2008; Turatto, Mazza, & Umiltà, 2005). Esta noção seria inadequada, por exemplo, à de objeto auditivo, associado à percepção de fontes sonoras (e.g., fala, música) e não de superfícies materiais (Kubovy & Van Valkenburg, 2001) e, mais geralmente, à de um objeto multimodal independente da dominância visual (ver também, e.g., O'Callaghan, 2008).

Uma das insuficiências da conceção comum (visuocêntrica) de percepto multissensorial é a impossibilidade de admitir diferentes graus de objetualidade, dando lugar, em alternativa, a um percepto multimodal unificado (assente na assunção de unidade) ou a perceptos unimodais dissociados (sem assunção de unidade). Diferentemente, a noção de integração parcial, prevista nos modelos Bayesianos de inferência causal (ver secção 2.1.3.2 neste capítulo), é compatível com uma graduação da consistência dos perceptos multissensoriais. Sendo primariamente matemáticos e computacionais (não psicológicos), os modelos Bayesianos não são explícitos sobre as possíveis contrapartidas fenomenológicas dessas integrações parciais (e.g., o que significa, na experiência perceptiva, um grau de certeza intermédio sobre a unidade de duas ou mais entradas sensoriais?). No entanto, um vínculo pode ser conjecturado com uma objetualidade perceptiva que admite variações em grau.

Noções como a de *ficheiro de objeto* (Kahneman, Treisman, & Gibbs, 1992), *proto-objeto* (von der Heydt, 2015) ou *força da representação do objeto* (Shomstein & Behrmann, 2008) têm servido para indexar diferentes graus de objetualidade perceptiva, não apenas no domínio da visão como de outras modalidades e ainda da percepção multimodal, particularmente audiovisual. Zmigrod & Hommel (2009), por exemplo, utilizam a noção de *ficheiro de objeto* na audição, estendendo-a adicionalmente de modo a integrar informação relativa às respostas (planeamento da ação) e não apenas aos estímulos. K. E. Jordan et al. (2010) exploram a noção de *ficheiro de objeto* intermodal no domínio audiovisual, assinalando que estas representações perceptivas (de natureza episódica; ver Mitroff, Scholl, & Noies, 2007) podem variar no seu grau de robustez em função da quantidade e qualidade da informação integrada. Uma noção de percepto multissensorial, e em especial audiovisual, baseada não no modelo do objeto material da visão, mas admitindo, por exemplo, a formação de perceptos audiovisuais definido unicamente pela sua estrutura temporal (Heron, Roach, Whitaker, & Hanson, 2010), parece assim não apenas concebível, como beneficiar do suporte de algumas indicações convergentes.

A admissão de diferentes graus de objetualidade perceptiva torna também possível conceber formas de representação multissensorial menos constrangidas pelas regras espacial (coincidência ou proximidade) e, em especial, temporal (sincronia ou janela temporal estreita). Num contexto onde a representação perceptiva flexível de estruturas causais e estatísticas do ambiente (não necessariamente corpos materiais localizados) constitui o objetivo geral do processamento sensorial, a extensão em que sinais de diferentes modalidades devem ser combinados não depende necessariamente da quase-sincronia (Fiebelkorn et al., 2010; Heron et al., 2010; Parise, Spence, & Ernst, 2012). No domínio da visão, por exemplo, a existência de relações temporais consistentes entre estímulos independentes da sincronia temporal (e.g., alterações correlacionadas no tempo) mostrou ser um poderoso fator de agrupamento (Guttman, Gilroy, & Blake, 2007) e dar lugar a uma variedade considerável de objetos perceptivos (Blake & Lee, 2005). Do mesmo modo, no domínio do processamento multissensorial, a importância das correlações cruzadas entre a estrutura temporal de estímulos visuais e auditivos (i.e., o grau de semelhança dos seus padrões temporais) revelou-se um importante fator de integração multissensorial (Parise, Harrar, Ernst, & Spence, 2013; Parise et al., 2012). Como assinalam os autores, estas correlações são independentes da sincronia audiovisual (condições correlacionadas e não correlacionadas encontravam-se equiparadas quanto à sincronia-assincronia dos sinais) (Parise et al., 2012). A interpretação que fornecem dos resultados, num quadro Bayesiano, é

a de que a correlação entre estruturas temporais é utilizada como um índice de causa comum e representada na distribuição *a priori* da probabilidade de coocorrência dos sinais (Parise et al., 2013, 2012).

O estudo da organização perceptiva intermodal (e.g., fenômenos de agrupamento, segregação figura-fundo, *Gestalts* multissensoriais) constitui um segundo exemplo de um domínio onde correspondências de estrutura entre sinais multissensoriais podem determinar a sua organização em unidades perceptivas significativas, não necessariamente dependentes da coincidência espaço-temporal (Spence, 2015). Influências do agrupamento intramodal na integração multissensorial foram já demonstradas. Por exemplo, quando um som é parte de uma melodia ou outro padrão auditivo, a sua integração com um sinal visual simultâneo é dificultada ou impedida (Bruns & Getzmann, 2008; Cook & Van Valkenburg, 2009; Vroomen & de Gelder, 2000). Embora relativamente ao agrupamento intermodal a evidência seja ainda escassa (e mista, no caso particular das *Gestalts* espaciais: ver, e.g., Huddleston, Lewis, Phinney, & DeYoe, 2008), a organização perceptiva multissensorial parece obedecer aos mesmos princípios de forma que vigoram na percepção intramodal (e.g., proximidade, destino comum, similaridade) (Spence, 2015). A possibilidade de um papel na percepção multimodal de padrões de contingências entre sinais não estritamente dependentes da regra temporal converge não apenas com distinções como a de Fiebelkorn et al. (2011) entre processos multissensoriais guiados pelos dados (dependentes do alinhamento espaço-temporal) e por representações (dependentes de associações aprendidas, de longo termo, entre propriedades dos eventos), como ainda com o reconhecimento crescente da importância de associações aprendidas “de alto nível” (*high level*) na adaptação e recalibração intermodais (Heron et al., 2010; Hidaka et al., 2015; Li, Yang, Sun, & Wu, 2015; Seitz et al., 2006; Shams, Wozny, Kim, & Seitz, 2011).

CAPÍTULO 3

MODULAÇÃO DA PERCEÇÃO DO MOVIMENTO VISUAL PELAS PROPRIEDADES DO SOM: MOMENTO REPRESENTACIONAL AUDIOVISUAL

O presente capítulo revê os principais resultados disponíveis sobre a modulação da percepção do movimento visual por diferentes propriedades do som (secção 3.1) e, enquanto caso particular em foco neste trabalho, sobre a modulação do momento representacional visual por pistas auditivas (secção 3.2). Na secção 3.3 identifica ainda um conjunto de questões irresolvidas na literatura que assinalam a transição entre o estado da arte e os objetivos e racional orientador da investigação apresentada na parte empírica da dissertação (Parte II).

3.1. IMPACTO DAS PROPRIEDADES ESPACIAIS, TEMPORAIS E SEMÂNTICAS DO SOM NO MOVIMENTO VISUAL

A percepção visual do movimento suporta um conjunto de respostas comportamentais com um elevado valor adaptativo, como o evitamento de colisões, a interseção ou a captura de alvos móveis. Constitui simultaneamente uma componente nuclear na interpretação e representação de cenas, promovendo a integração espaço-temporal da informação em unidades dinâmicas significativas (e.g., trajetórias de movimento) (Albright & Stoner, 1995; Hadad, Schwartz, Maurer, & Lewis, 2015). Considerando que o movimento dos objetos e animais gera tipicamente informação noutras modalidades para além da visão,

designadamente a audição (e.g., um carro em movimento produz sons que se alteram dinamicamente), a questão de saber se a visão e a audição podem influenciar-se reciprocamente na percepção do movimento parece natural. No entanto, apenas foi colocada, de forma sistemática, relativamente tarde. Para isso concorreram diversos fatores. A hipótese da dominância visual em domínios com uma forte componente de processamento espacial previa à partida uma escassa influência da audição (Teramoto, Hidaka, & Sugita, 2018). Na modalidade visual, os sinais de movimento beneficiam de maior saliência perceptiva do que os sinais estáticos (Dick, Ullman, & Sagi, 1987) reduzindo mais ainda a plausibilidade de uma modulação pelo som. Do ponto de vista neurológico, evidência convincente sobre a existência de detetores neurais do movimento, seletivos à direção e velocidade, existe para a visão (ver, e.g., Borst, 2000), mas não para a audição (Carlile & Leung, 2016). Convergindo com estas indicações, os resultados dos primeiros estudos audiovisuais de percepção do movimento suportaram a hipótese de uma dominância da visão. Estímulos auditivos estáticos eram percebidos como em movimento na mesma direção que um movimento visual simultâneo (captura visual dinâmica: Mateeff & Hohnsbein, 1988), sem que o contrário se verificasse. O mesmo sucedia com estímulos auditivos dinâmicos, cuja direção percebida (qualquer que fosse a direção real: horizontal, vertical, ou em profundidade) era fortemente modulada por um movimento visual simultâneo (Kitajima & Yamashita, 1999), sem que uma influência simétrica fosse documentada (Kitagawa & Ichihara, 2002).

A situação alterou-se nos estudos mais recentes, após a demonstração de efeitos de ventriloquismo temporal, equivalentes aos documentados em apresentações audiovisuais estáticas (e.g., Morein-Zamir et al., 2003; Scheier, Nijhawan, & Shimojo, 1999), no domínio da percepção do movimento (e.g., Heron et al., 2004; Vroomen & de Gelder, 2004). Estes resultados, evidenciando a capacidade das propriedades temporais do som para modular aspetos temporais do processamento visual do movimento, mostraram a necessidade de distinguir entre o tipo de pistas auditivas manipuladas (e.g., temporais, espaciais, semânticas: ver Teramoto et al. (2012), estendendo de alguma forma as implicações da hipótese de adequação da modalidade (dominância da audição na dimensão temporal, da visão na dimensão espacial) às interações audiovisuais no processamento da informação de movimento. Por outro lado, contrariando a investigação inicial, vários estudos reportaram também um impacto das características espaciais do som na percepção visual de movimentos apresentados na periferia da retina, onde a resolução espacial do sistema visual se encontra diminuída (e.g., Hidaka, Kawachi, & Gyoba, 2009; Teramoto et al., 2012). Estes resultados

replicaram resultados já antes documentados com estímulos estáticos (ventriloquismo espacial inverso: Alais & Burr, 2004), chamando a atenção para a importância decisiva da fiabilidade relativa dos sinais sensoriais também no domínio da percepção do movimento.

Uma revisão recente de Teramoto et al. (2018) permite caracterizar globalmente o conjunto destes estudos. Adequadamente, a revisão encontra-se organizada em função das propriedades acústicas manipuladas, dedicando secções separadas aos efeitos no processamento visual do movimento de pistas auditivas temporais, espaciais e semânticas.

Na secção consagrada às pistas auditivas temporais (e.g., *onset*, *offset*, duração do som) é possível distinguir três tipos de estudos. (i) Um primeiro tipo ocupa-se *dos efeitos do momento de apresentação de um som na localização relativa de um movimento contínuo* (“*smooth continuous motion*”), de que são exemplo os estudos de Vroomen & de Gelder (2004) com o efeito *flash-lag* e de Heron et al. (2004) com o paradigma do ressalto (*bounce*). No efeito *flash-lag*, um *flash* estático e um estímulo visual em movimento apresentados em alinhamento físico são percebidos como desalinhados (estímulo em movimento mais à frente). Vroomen & de Gelder (2004) mostraram que um som apresentado antes ou depois do *flash* modulava o desalinhamento espacial percebido, diminuindo-o no primeiro caso e aumentando-o no segundo. No efeito de ressalto, um alvo visual em movimento inverte a sua direção (ressalta) ao atingir um estímulo de referência estático. Heron et al. (2004) mostraram que um som apresentado antes do ressalto atrai a localização percebida do ressalto na direção oposta à do movimento inicial. (ii) Um segundo tipo de estudos ocupa-se *dos efeitos do momento da apresentação de sons na qualidade ou na direção percebida de movimentos visuais aparentes*. Getzmann (2007) mostrou neste sentido que um som localizado temporalmente entre dois estímulos visuais apresentados em sucessão melhorava a qualidade percebida do movimento visual aparente, enquanto sons localizados temporalmente antes do primeiro ou depois do segundo estímulo visual a prejudicavam. E. Freeman & Driver (2008) mostraram que a relação temporal entre dois estímulos auditivos apresentados conjuntamente com dois indutores visuais de movimento aparente modulava a direção percebida desse movimento. (iii) Um terceiro tipo de estudos investiga *a modulação da posição final percebida de um movimento visual contínuo* pela localização temporal de um som antes, após, ou em simultâneo com o *offset* do movimento. Em geral, os resultados documentaram uma atração do *offset* temporal do movimento pela localização temporal do som (Chien et al., 2013; Teramoto, Hidaka, et al., 2010). Este tipo de estudos, assentando em tarefas de momento representacional, tem um particular interesse no contexto da dissertação e será descrito mais detalhadamente na secção seguinte. Duas condições

destacadas por Teramoto et al. (2018) como necessárias para a modulação da percepção visual do movimento pelas propriedades temporais do som são um balanço de precisão temporal favorável à informação auditiva (relativamente à visual) e valores de assincronia audiovisual que não excedam uma janela temporal crítica situada em torno dos 100 ms a 150 ms.

O número de estudos que na revisão de Teramoto et al. (2018) suportam efeitos moduladores de pistas auditivas espaciais na percepção visual do movimento é consideravelmente inferior aos reportados no caso das pistas auditivas temporais. Hidaka, Kawachi, et al. (2009) demonstraram a possibilidade de *induzir movimento percebido num estímulo visual estático* intermitente, apresentado na periferia da retina, *através* de um *movimento aparente auditivo* horizontal, gerado por sinais apresentados alternadamente a cada um dos ouvidos por auscultadores. Teramoto, Manaka, et al. (2010) replicaram este efeito (Movimento Visual Induzido pelo Som: MVIS) e estenderam-no a movimentos auditivos aparentes verticais gerados por meio de altifalantes separados verticalmente. Em ambos os casos, a espacialização do som foi obtida pela localização física da fonte (auscultador direito ou esquerdo, altifalante em cima-em baixo). No entanto, na disposição horizontal as pistas espaciais auditivas consistiram em diferenças binaurais (diferença interaural de tempo: ITD) enquanto na disposição vertical dos altifalantes dependeram de alterações espectrais associadas à filtragem direcional do pavilhão auricular, cabeça e ombros (Teramoto, Manaka, et al., 2010). Para além de uma nova replicação do efeito MVIS com a utilização de movimento auditivo contínuo, Hidaka et al. (2011) observaram a *desambiguação da direção percebida de movimento visual global* num ecrã dinâmico composto pela aleatorização de vetores locais de movimento. Utilizando altifalantes em alinhamento horizontal e vertical, Teramoto et al. (2012) mostraram a capacidade do *movimento aparente auditivo* para influenciar a *direção percebida de um movimento visual* aparente (a partir de um certo grau de excentricidade retiniana do movimento visual) (Hidaka et al., 2011). Por fim, um estudo de Schmiedchen et al. (2012) investigou o efeito de um movimento auditivo contínuo gerado por um arranjo denso de altifalantes na localização percebida do *offset* de um movimento visual, cujo termo podia ocorrer em regiões mais ou menos periféricas do campo visual. Este estudo, que não admite conclusões simples, assentou numa tarefa de momento representacional (localização da última posição percebida de um alvo em movimento), sendo por isso detalhado e discutido na secção seguinte. As condições gerais destacadas por Teramoto et al. (2018) como necessárias para a modulação da percepção visual do movimento por características espaciais do som são uma

elevada incerteza posicional do estímulo visual (obtida nos estudos anteriores pela sua localização periférica) e/ou uma alta qualidade das pistas espaciais auditivas.

A consideração, para além de pistas acústicas temporais e espaciais, de pistas semânticas (incluindo latamente correspondências intermodais, como a que se verifica entre sons agudos-graves e localização espacial vertical elevada-baixa) justifica-se em primeiro lugar face ao desafio comumente colocado à audição de segregar e localizar fontes sonoras relevantes em ambientes complexos (com múltiplas fontes, reverberações e interferências auditivas). A combinação de pistas auditivas espaciais (e.g., diferenças interaurais de tempo e intensidade e pistas espectrais monaurais) oferece uma base suficiente para atribuir uma localização espacial subjetiva a uma fonte sonora isolada, num contexto auditivo simples (e.g., sons isolados), mas não resolve a incerteza sobre as pistas a combinar numa cena auditiva envolvendo numerosas fontes (Darwin, 2008; Yost, 1997). Uma forma de resolver este problema é começar por uma análise da cena auditiva (*auditory scene analysis*: Bregman, 1990) em termos de objetos perceptivos suscetíveis de representar fontes físicas ou padrões auditivos temporais (e.g., melodias), limitando depois as combinações a pistas procedentes de uma fonte comum (Hill & Darwin, 1996).

A análise de cenas auditivas pode tirar partido de uma ampla variedade de informações, tanto de baixo nível – e.g., relações harmónicas entre sons (Culling & Stone, 2017), coerência temporal (Shamma, Elhilali, & Micheyl, 2011) consistência no tempo de pistas espaciais (Maddox & Shinn-Cunningham, 2012) – como de alto nível, envolvendo a alocação endógena da atenção, esquemas de conhecimento e expectativas, na qual encontram lugar as pistas de coerência semântica (Lee, Maddox, & Bizley, 2019; Lee & Wallace, 2019). Pode igualmente beneficiar da informação de outras modalidades, como a visão (Lee et al., 2019), sendo a congruência semântica audiovisual (ou formas similares de congruência sinestética e/ou correspondência intermodal: ver Spence, 2011) um fator importante de interação-integração sensorial. Numerosos estudos documentam, em geral, este papel da congruência semântica audiovisual (ver, e.g., Y.-C. Chen & Spence, 2010, 2017). Os estudos reportados na secção que Teramoto et al. (2018) dedica às pistas semânticas auditivas investigam especificamente o seu efeito no processamento visual do movimento. Maeda, Kanai, & Shimojo (2004) mostraram a *possibilidade de desambiguar a direção percebida* de um movimento visual composto por vetores de movimento ascendentes e descendentes através da apresentação simultânea de séries crescentes ou decrescentes de alturas tonais. Hidaka, Teramoto, Keetels, & Vroomen (2013), no entanto, não lograram *induzir movimento percebido num estímulo visual estático* apresentado na periferia (o efeito

MVIS, obtido por Hidaka, Kawachi, et al., 2009) com sons de frequência alternadamente alta e baixa. Resultados mistos foram por sua vez reportados por Hubbard & Courtney (2010), que combinaram séries tonais ascendentes e descendentes (no domínio das frequências) com uma tarefa de momento representacional. A localização do termo dos movimentos visuais verticais não foi afetada pelos sons, enquanto nos movimentos visuais horizontais, um pequeno aumento do desvio para baixo (na direção da gravidade) da localização final percebida foi observada quando conjugados com séries tonais descendentes. Num estudo envolvendo a percepção de movimento biológico, a detecção de uma sequência de dança de sapateado (ilustrada pelo movimento de pontos luminosos) embebida em ruído visual foi beneficiada pela apresentação sincronizada de sons de sapateado, semanticamente congruentes (Arrighi, Marini, & Burr, 2009). Vários outros estudos têm suportado um efeito de informação auditiva semanticamente congruente (ou mantendo correspondências “ecologicamente” significativas com a informação visual) na percepção do movimento biológico (e.g., Saygin, Driver, & de Sa, 2008; Schouten, Troje, Vroomen, & Verfaillie, 2011). Teramoto et al. (2018) concluem a sua revisão do efeito das pistas semânticas auditivas assinalando a relativa inconstância dos resultados entre estudos. Um dos problemas na origem desta inconsistência poderá residir no facto de, em muitas circunstâncias, a congruência semântica ser operacionalizada por emparelhamentos de pistas pouco realistas e desprovidos de valor ecológico (Y.-C. Chen & Spence, 2017; Vatakis & Spence, 2007, 2008). Em qualquer caso, resultados como o de Gohara, Yamada, & Miura (2016), ilustrando uma modulação do momento representacional visual por palavras japonesas miméticas (alusivas a um movimento rápido ou a uma travagem) sugerem um papel das correspondências audiovisuais no processamento visual do movimento.

É possível assim destacar as seguintes tendências a partir da revisão de Teramoto et al. (2018). A maioria dos estudos do efeito de pistas auditivas na percepção do movimento visual utilizou movimento visual aparente, sendo em número bastante inferior os estudos que utilizaram movimento visual contínuo. A maioria dos estudos interessou-se por efeitos de modulação da direção e qualidade percebida de movimento visual aparente, de indução de movimento percebido em estímulos visuais estáticos, de desambiguação da direção de movimentos globais, ou de facilitação da detecção de movimento visual, sendo em número bem menor os que investigaram efeitos na localização percebida do movimento. Poucos dos estudos que manipularam as propriedades espaciais do som utilizaram movimento auditivo contínuo. A proximidade temporal dos sinais visuais e auditivos (regra temporal) e sua

fiabilidade relativa (em conformidade com as propostas do modelo da máxima verossimilhança) condicionaram em geral a ocorrência dos efeitos.

3.2. MOMENTO REPRESENTACIONAL AUDIOVISUAL: EFEITOS DE ATRAÇÃO TEMPORAL E ESPACIAL DO SOM

Como resulta da revisão de Teramoto et al. (2018), poucos estudos avaliaram o impacto de informações auditivas na localização percebida de alvos visuais em movimento, privilegiando em vez disso aspectos como a ocorrência, direção percebida, coerência ou qualidade fenomenológica do movimento. Isto é especialmente verdade no caso das pistas auditivas espaciais, onde apenas um estudo entre os citados na revisão de Teramoto et al. (2018) se ocupou da localização do movimento visual (Schmiedchen et al., 2012). Dado que a localização instantânea de um alvo móvel corresponde ao aspecto mais puramente espacial do movimento (classicamente concebido como a alteração no tempo da localização espacial de um objeto), este estado de coisas poderá refletir a convicção difundida de uma incapacidade das propriedades espaciais do som para afetar componentes intrinsecamente espaciais do processamento visual do movimento.

No entanto, mesmo no caso de pistas auditivas temporais (sendo em geral admitido um primado da audição sobre a visão no processamento temporal), a avaliação do seu impacto na localização visual de alvos móveis envolveu frequentemente respostas de localização relativa (separação relativamente a um segundo estímulo, em geral estático, como nos efeitos de ressalto e *flash-lag*: Heron et al., 2004; Vroomen & de Gelder, 2004) e não absoluta. Esta opção pela localização relativa exprime porventura ainda uma reserva sobre a possibilidade da audição influenciar a posição absoluta percebida de um alvo visual em movimento. Neste sentido, mais importante do que distinguir descritivamente entre estudos focalizados na modulação auditiva do *onset* temporal de eventos visuais dinâmicos e estudos focalizados na modulação do seu *offset* (Teramoto et al., 2018), parece ser notar que os primeiros se mantiveram no âmbito da localização relativa de eventos visuais e apenas os segundos investigaram os efeitos do som na localização absoluta de um alvo em movimento. Todos os exemplos do segundo tipo fornecidos por Teramoto et al. (2018) envolveram a localização do termo de um movimento visual, correspondendo assim a tarefas de momento representacional.

Nesta secção ocupamo-nos, em particular, dos estudos de momento representacional audiovisual, o paradigma-base da componente empírica da presente dissertação. Do nosso melhor conhecimento, apenas três estudos utilizaram uma tarefa de momento representacional intermodal para avaliar a modulação do MR visual por pistas auditivas: Teramoto, Hidaka, et al. (2010), Chien et al. (2013) e Schmiedchen et al. (2012). Para além destes, só o estudo de Merz, Meyerhoff, Frings, & Spence (2020), ilustrando uma influência da informação visual na localização do *offset* de um movimento tátil, parece ter sido conduzido para outros pares de modalidades. Os três estudos de MR audiovisual mencionados são descritos e caracterizados em seguida.

Teramoto, Hidaka, et al. (2010) apresentaram em auscultadores sons contínuos não lateralizados (desprovidos de pistas de localização) que terminaram antes, em simultâneo com, ou depois do termo (*offset*) do movimento visual. As assincronias utilizadas, em número de seis, variaram entre os ± 53.2 ms e os ± 159.6 ms, a passos de cerca de 53 ms. Os participantes deveriam manter o olhar num ponto de fixação durante os ensaios, embora o cumprimento desta instrução não tenha sido controlado. A localização percebida do ponto de desaparecimento do alvo visual foi avaliada através dum método de julgamento perceptivo, consistindo em apresentar um estímulo-sonda subsequente em posições variáveis em torno do ponto de desaparecimento do alvo, e solicitando aos participantes que decidissem (resposta de escolha forçada) se a sonda se encontrava à esquerda ou à direita desse ponto. Quando o som e o movimento tinham início simultaneamente (*onset* síncrono), o erro de localização para diante (Momento Representacional) aumentava com sons de duração superior à do movimento visual e diminuía com sons de duração inferior à do movimento (comparativamente a uma condição de referência sem som). Estes efeitos atingiam o seu máximo para assincronias em torno dos ± 100 ms.

Teramoto, Hidaka, et al. (2010) interpretaram estes resultados como a demonstração de uma modulação temporalmente mediada do MR visual pelo *offset* do som. A informação auditiva temporal contribuiria para uma recodificação do momento (*timing*) do *offset* do movimento visual, daí resultando uma alteração da última posição percebida do alvo. Esta modulação estaria sujeita a duas condições: uma estreita associação entre os sinais auditivos e visuais desde o início, implicando sincronia no *onset*, e uma estreita proximidade temporal no *offset*, implicando assincronias não muito distante dos ± 100 ms. Ambas foram interpretadas por Teramoto, Hidaka, et al. (2010) como condições de emergência de um percepto multissensorial unificado: a assunção de que os sinais têm uma fonte externa comum (“assunção de unidade”: Welch, 1999; Welch & Warren, 1980) e a submissão a uma

janela temporal crítica de integração multissensorial (Meredith, 2002). Nesta explicação, a modulação temporalmente mediada do MR visual por pequenas assincronias audiovisuais ocorre no quadro de um evento audiovisual integrado, sendo portanto condicional à sua emergência.

Chien et al. (2013, experiência 1) apresentaram através de auscultadores sons transientes (duração de 10 ms) antes, ao mesmo tempo, ou após o termo do movimento visual. Tal como Teramoto, Hidaka, et al. (2010), utilizaram seis assincronias, variando entre os ± 40 ms e ± 140 ms, a passos de 40 ms. Os participantes foram solicitados a manter o olhar numa cruz de fixação, sem que, contudo, fossem registados os seus movimentos oculares. A resposta dos participantes consistiu em indicar por meio de um dispositivo apontador (rato) o local onde o alvo visual subitamente desaparecera. Apenas erros de localização “para trás”, isto é, no sentido contrário ao do movimento, foram registados. Os sons simultâneos com ou subsequentes ao *offset* do movimento não apresentaram efeitos significativos, enquanto os sons que precediam esse *offset* induziram erros de localização para trás. Os autores puderam excluir que este resultado se devesse à exigência de manter a fixação ocular que, no caso de movimentos contínuos do alvo, reduz sensivelmente a magnitude do MR (Kerzel, 2000). Através duma replicação da experiência sem instruções de fixação os resultados obtidos na replicação foram idênticos. Uma explicação possível para a predominância dos erros de localização para trás no estudo de Chien et al. (2013), não considerada pelos autores envolve o modo de apresentação do movimento visual. Chien et al. (2013) apresentaram um disco preto que se movia na horizontal a meia altura de um ecrã com a metade inferior preta e a metade superior (fornecendo o contraste de luminância necessário à deteção do alvo) cinzenta. A impressão resultante é a de um disco que se move sobre e em contacto com uma superfície subjacente. Este tipo de apresentação é tipicamente utilizado para sugerir fricção, originando uma redução do erro de localização para diante conhecida como “fricção representacional” (Hubbard, 1995b, 1995c).

Independentemente da ausência de erros para diante estatisticamente significativos, Chien et al. (2013) avançaram uma explicação semelhante à de Teramoto, Hidaka, et al. (2010). Sinais auditivos precedendo temporalmente o *offset* do movimento visual atraíram o *timing* percebido da desaparecimento do alvo, resultando em deslocamentos espaciais mais pronunciados para trás. Os dois estudos surgem por isso listados na recente revisão de Teramoto et al. (2018) entre as demonstrações de efeitos de pistas auditivas temporais na percepção visual do movimento. Embora divergindo quanto à necessidade de associação dos sinais auditivo e visual desde o início do movimento, Chien et al. (2013) partilharam

igualmente com Teramoto, Hidaka, et al. (2010) a interpretação de que a modulação observada dependia da emergência de um evento audiovisual, regido por uma janela temporal de integração em torno dos ± 100 ms.

Numa segunda experiência reportada, Chien et al. (2013, experiência 2) procuraram ainda avaliar a possibilidade de um efeito direto de pistas auditivas espaciais. Para o efeito, apresentaram o som monauralmente, ao ouvido direito ou ao ouvido esquerdo. Observaram enviesamentos da localização para trás mais pronunciados quando o ouvido se encontrava do mesmo lado onde o movimento do alvo tinha origem, sugerindo que o lado de apresentação do som (uma informação espacial) modulava o MR visual. No entanto, o efeito podia ainda assim ser de natureza temporal. Se o lado onde se origina o movimento beneficiasse de um privilégio atencional, um som apresentado desse lado anteriormente ao *offset* do movimento poderia ser detetado mais rapidamente (mais cedo) e induzir um enviesamento mais forte “para trás”. De modo a decidir entre as duas interpretações (espacial e temporal), realizaram uma terceira experiência de julgamento de ordem temporal (JOT), destinada a perceber se o lado de apresentação do som (o ouvido de apresentação) alterava as relações temporais percebidas entre o *onset* do som e o *offset* visual. Os resultados foram negativos, levando os autores a concluir por um efeito da informação espacial do som, não mediado temporalmente, no MR visual. O valor destes resultados é contudo limitado e insuficiente para suportar a conclusão pretendida. Por um lado, assenta em evidência indireta, recolhida numa tarefa muito diferente (paradigma JOT) da tarefa base, não isenta, além disso, de problemas quanto à sua capacidade para avaliar relações temporais (García-Pérez & Alcalá-Quintana, 2012; Matthews, Welch, Achtman, Fenton, & FitzGerald, 2016). Por outro lado, a apresentação de sons a um único ouvido dificilmente pode descrever-se como envolvendo pistas auditivas de localização (e.g., diferenças interaurais de tempo ou intensidade, funções relacionadas com a transferência de cabeça, padrões de reverberação, etc.). Efeitos gerais de compatibilidade estímulo-resposta (Hommel, 2011) ou enviesamentos inespecíficos de lateralidade para o hemiespaço estimulado (Dupierriex, Alleysson, Ohlmann, & Chokron, 2008) poderiam, por si só, explicar estes resultados.

Schmiedchen et al. (2012) geraram sons em movimento em campo livre através da ativação sucessiva de altifalantes adjacentes (organizados num arranjo denso que correspondia fisicamente ao trajeto da fonte sonora). Esta estratégia de espacialização do som assegura pistas auditivas de localização de boa qualidade (ver Meyer et al., 2005). Os movimentos auditivos eram apresentados em simultâneo com um movimento visual,

podendo terminar espacialmente atrás, no mesmo local, ou à frente do *offset* do movimento visual. As disparidades espaciais utilizadas foram em número de seis, variando entre os $\pm 5^\circ$ e os $\pm 15^\circ$ de ângulo visual (a.v.), em passos de 5° . Os participantes deviam olhar em frente durante as apresentações, embora nenhum ponto de fixação fosse apresentado, de modo a excluir que pudesse ser utilizado como uma referência espacial pelos participantes. A resposta consistia em identificar o local percebido do *offset* do movimento visual por meio de um apontador de infravermelhos. O termo percebido do movimento visual (com uma duração de 0.5 s) apresentou enviesamentos em direção ao *offset* espacial dos movimentos auditivos (para trás quando este ocorria espacialmente antes, para a frente quando este ocorria espacialmente adiante). Estes efeitos foram mais fortes quando o movimento visual terminava na periferia da retina ($\pm 60^\circ$ a.v.) do que na zona central ($\pm 8^\circ$ a.v.), mas ocorreram em todas as regiões do campo visual (central, paracentral e lateral). Um problema com a interpretação destes resultados, no entanto, é que as disparidades visuais se encontravam confundidas no desenho experimental com disparidades temporais (assincronias), de tal modo que cada acréscimo de 5° a.v. de disparidade espacial implicava um aumento da assincronia entre os *offsets* visual e auditivo de 65 ms. Deste modo, é impossível decidir se a modulação do MR visual observada se deveu a informação auditiva temporal, espacial, ou a ambas. Apesar de incluído na revisão de Teramoto et al. (2018) entre os estudos demonstrando efeitos de pistas auditivas espaciais na percepção visual do movimento, o estudo de Schmiedchen et al. (2012) é insuficiente para suportar conclusivamente esses efeitos.

A interpretação avançada por Schmiedchen et al. (2012) para os efeitos registados difere da de Teramoto, Hidaka, et al. (2010) e Chien et al. (2013), por recusarem a sua dependência relativamente à emergência de eventos audiovisuais integrados. Baseados numa distinção entre interações intermodais (pré-perceptivas) e integração multissensorial (entendida como fusão perceptual) (ver Bertelson & Radeau, 1981; Spence et al., 2009) e nos resultados de uma experiência suplementar em que avaliaram o grau de unidade percebida dos movimentos visual e auditivo, optaram por uma explicação pré-perceptiva (dispensando a percepção audiovisual). Contudo, tal como os estudos anteriores, assinalaram o papel determinante das pequenas assincronias (disparidades temporais) no *offset*, que não poderiam exceder uma amplitude crítica (± 195 ms) sem que os efeitos moduladores das disparidades espaço-temporais desaparecessem. Apesar da opção por outros mecanismos explicativos, Schmiedchen et al. (2012) preservaram no entanto a subordinação desses mecanismos a uma estrita “regra temporal”.

Entre os estudos discutidos existem múltiplas variações de procedimento. Chien et al. (2013) e Schmiedchen et al. (2012) usaram respostas motoras de apontar (diferindo embora no dispositivo), Teramoto, Hidaka, et al. (2010) julgamentos perceptivos. A extensão dos trajetos do movimento visual, semelhante em Teramoto, Hidaka, et al. (2010) e Chien et al. (2013) foi consideravelmente superior em Schmiedchen et al. (2012). Teramoto, Hidaka, et al. (2010) e Schmiedchen et al. (2012) utilizaram sons contínuos, Chien et al. (2013) sons transientes. Teramoto, Hidaka, et al. (2010) e Chien et al. (2013) apresentaram sons desprovidos de pistas de localização em auscultadores, Schmiedchen et al. (2012) sons espacializados em campo livre. Outras diferenças relativas à localização dos pontos de desaparecimento do alvo visual ou aos detalhes da apresentação dos estímulos visuais poderiam ainda acrescentar-se. O ponto importante a destacar é que, apesar destas diferenças, o resultado comum a todos os estudos foi a observação de modulações do MR visual correspondentes a enviesamento de atração para o som. Como sugeriremos adiante, o número de disparidades utilizadas (seis, em todos os estudos) e a utilização de pequenas magnitudes de assincronia (< 200 ms em todos os estudos) poderão ter constituído os fatores determinantes para este resultado, sobrepondo-se a eventuais efeitos das variações de procedimento.

3.3. QUESTÕES EM ABERTO

3.3.1. Dissociação experimental do efeito da informação temporal e espacial do som

Como resulta da análise dos estudos apresentados na secção anterior, a questão relativa à possibilidade de uma modulação direta (não mediada temporalmente) do MR visual por informações espaciais não obteve uma resposta conclusiva. Do ponto de vista teórico, a investigação nesta matéria tem sido orientada essencialmente por duas hipóteses gerais: a *hipótese da adequação da modalidade* (Welch & Warren, 1980, 1986) e o *modelo da máxima verosimilhança da integração multissensorial* (MMV: Alais & Burr, 2004; Ernst & Banks, 2002), de inspiração Bayesiana.

A hipótese da adequação da modalidade afirma uma vantagem da audição sobre a visão no processamento temporal e uma vantagem simétrica da visão sobre a audição no

processamento espacial: uma previsão inicialmente derivada desta hipótese foi assim a de que a audição poderia apenas influenciar a visão em tarefas temporais, mas não em tarefas espaciais. Na realidade, na medida em que os aspetos temporais e espaciais da visão interagem entre si (Krekelberg & Lappe, 2001; Whitney, 2002), seria possível admitir, de forma ainda compatível com a hipótese de adequação da modalidade, uma modulação da localização de um evento visual pelo som, desde que mediada por alterações no *timing* desse evento (Kawachi, 2016; Teramoto et al., 2018). Por outras palavras, a audição poderia afetar a visão tanto em tarefas temporais como espaciais, mas, em ambos os casos, apenas através de pistas auditivas temporais, não espaciais. Vários resultados favoráveis a um efeito temporalmente mediado da audição no processamento visual de dimensões espaciais foram entretanto produzidos (e.g., E. Freeman & Driver, 2008; Sekuler, Sekuler, & Lau, 1997; Vroomen & de Gelder, 2004), entre os quais se podem incluir também os estudos de Teramoto, Hidaka, et al. (2010) e Chien et al. (2013) discutidos acima (secção anterior). No entanto, demonstrar que apenas efeitos temporalmente mediados são possíveis (segunda parte, essencial, das previsões baseadas na adequação da modalidade) requer ainda demonstrar que, quando apresentadas isoladamente de pistas temporais, as pistas auditivas espaciais não têm impacto no processamento visual da localização. Este ponto implica resolver o difícil problema metodológico da dissociação experimental (por manipulação controlada) da informação temporal e espacial veiculada pelo som (Kawachi, 2016), algo que nenhum dos estudos discutidos logrou realizar adequadamente.

O modelo da máxima verosimilhança postula que a informação de cada modalidade é ponderada pela sua fiabilidade (definida como o inverso da sua variância), podendo assim prevalecer mais ou menos sobre os sinais de outra modalidade em função da fiabilidade relativa (comparativa) desses sinais. Deste ponto de vista, a predominância usual da visão sobre a audição em tarefas espaciais não reflete um privilégio da modalidade visual, mas a circunstância de a informação visual espacial ser usualmente mais fiável do que a informação auditiva espacial. Em condições de informação visual degradada, ou de fiabilidade relativa aumentada da informação espacial auditiva, nada impediria assim esta última de exercer uma influência direta (não temporalmente mediada) no processamento visual. Alais & Burr (2004) forneceram uma ilustração clara desta possibilidade mostrando como a localização aparente de sons atrai a localização percebida de alvos visuais estáticos degradados através de processos de desfocagem (*blur*) (ver também Battaglia et al., 2003). No domínio dos alvos visuais dinâmicos, Schmiedchen et al. (2012) procuraram no mesmo sentido averiguar a capacidade de pistas espaciais auditivas para influenciar a localização espacial em visão

periférica, onde a acuidade posicional na percepção visual do movimento decresce significativamente (Mckee & Nakayama, 1984). Tal como antes, no entanto, a possibilidade de interpretar conclusivamente os resultados depende criticamente da capacidade de dissociar experimentalmente o efeito das pistas auditivas espaciais e temporais, que permaneceram confundidas no desenho utilizado por Schmiedchen et al. (2012).

Uma das questões em aberto que motivou a investigação empírica conduzida no âmbito da presente dissertação, a do potencial impacto de propriedades espaciais do som na modulação do MR visual, envolve assim um problema metodológico central. Propondo um novo quadro operatório para a dissociação das pistas auditivas temporais e espaciais, assente na conjugação de um desenho fatorial específico, por um lado, com uma variação ortogonal das distâncias percorridas por sons espacializados e da sua duração (obtida através da manipulação da velocidade), o presente estudo propõe-se fornecer-lhe uma resposta conclusiva.

3.3.2. Papel respetivo das disparidades audiovisuais no começo e termo do movimento

Os estudos de Teramoto, Hidaka, et al. (2010) e Chien et al. (2013) discutidos em 3.2 divergiram quanto à necessidade de uma forte associação entre o som e o movimento visual desde o início (*onset*). Para Teramoto, Hidaka, et al. (2010), que utilizaram sons contínuos, isto exprimia a importância da *assunção de unidade*, isto é, da formação (não necessariamente consciente) da convicção de que os sinais auditivo e visual tinham uma fonte comum (Vatakis & Spence, 2007; Warren et al., 1981; Welch & Warren, 1980), para que o MR pudesse ser auditivamente modulado. Sendo a coincidência temporal uma das fontes mais poderosas da assunção de unidade, a sincronia dos sinais no *onset* operaria como uma condição necessária na produção da assunção. Chien et al. (2013) obtiveram pelo contrário os seus resultados com sons transientes apresentados na vizinhança temporal do *offset* do movimento, sem, portanto, que houvesse lugar a uma associação com o movimento visual desde o início.

Esta discrepância empírica não implica, em si mesmo, uma refutação da importância da assunção de unidade. Por um lado, há numerosas fontes da assunção de unidade para além da proximidade temporal entre sinais (para uma revisão, ver Y.-C. Chen e Spence, 2017). Por outro lado, como reconhecem Chien et al. (2013), há importantes diferenças de

procedimento entre o seu estudo e o de Teramoto, Hidaka, et al. (2010), que poderiam estar na origem de fontes alternativas de unificação. A opção por uma resposta de localização motora é uma delas. A sugestão de que respostas motoras visualmente guiadas recrutam processos distintos dos envolvidos em julgamentos perceptivos tem uma tradição consolidada (e.g., Milner & Goodale, 1995). Tem igualmente sido sugerido que a extrapolação da posição de um objeto em movimento ocorre na via sensoriomotora do sistema visual (Kerzel, 2003c; Kerzel & Gegenfurtner, 2003). A coocorrência de aspetos significativos do planeamento/preparação da resposta na ocasião das apresentações temporalmente próximas do *onset* do som e do *offset* do movimento, por exemplo, poderia assim facultar uma fonte de unificação alternativa dos sinais visual e auditivo.

O estudo de Schmiedchen et al. (2012) mostrou-se por sua vez desfavorável a um papel da assunção de unidade, apontando para uma modulação do MR independente duma percepção unificada dos sinais auditivo e visual. Esta conclusão foi baseada numa experiência suplementar que avaliou diretamente o grau de unidade percebido pelos participantes entre os dois tipos de sinais. Os efeitos de modulação do MR e os julgamentos de unidade mostraram ser independentes (Schmiedchen et al., 2012). Convém, todavia, notar que se trata de resultados obtidos numa tarefa distinta (julgamentos de unidade percebida), envolvendo uma alocação necessária de atenção às duas modalidades sensoriais (sendo que no MR visual a visão é a modalidade relevante) e pressupondo que a assunção de unidade é de natureza explícita (acessível a um julgamento declarado). Nesse sentido, deve considerar-se esta evidência desfavorável como apenas sugestiva e não conclusiva.

Diferentemente da questão teórica da assunção de unidade, de difícil ou impossível resolução neste quadro, a questão independente da necessidade de sincronia audiovisual no *onset* quando se utilizam sons contínuos pode ser empiricamente esclarecida. Schmiedchen et al. (2012) utilizaram unicamente sons contínuos sincronizados com o início dos movimentos visuais, o que não permite replicar ou contrariar o resultado Teramoto, Hidaka, et al. (2010). Na investigação empírica que conduzimos utilizamos sons contínuos, mas variando sistematicamente os valores de assincronia audiovisual tanto no *onset* como no *offset* do movimento. Diferentemente de Teramoto, Hidaka, et al. (2010), estas variações foram combinadas fatorialmente (*onset* × *offset*) num único bloco de ensaios e a gama de valores de assincronia foi a mesma no *onset* e no *offset* (as assincronias no *onset* utilizadas por Teramoto e colaboradores envolveram sempre um atraso do som, + 400 ms e + 800 ms, e valores várias vezes superiores aos das assincronias no *offset*, + 106.4 ms e - 53.2 ms). Uma das questões em aberto até à data à qual o presente estudo procura dar resposta é assim a do

papel respetivo das disparidades audiovisuais no início e no termo do movimento – ou, o que é praticamente equivalente, a da generalidade do resultado reportado por Teramoto, Hidaka, et al. (2010).

3.3.3. Enviesamentos de atração nas condições síncronas versus efeitos de melhoria audiovisual

Em todos os estudos de MR audiovisual discutidos no ponto 3.2. foram observados efeitos de atração para o *timing* e, no caso de Schmiedchen et al. (2012), para o *timing* e/ou posição espacial do som, sem relação sistemática com uma melhoria no desempenho da tarefa de localização. No entanto, a ausência de benefícios audiovisuais não se verificou apenas nas condições assíncronas. Em nenhum dos estudos a sincronia (disparidade temporal nula) entre os sinais auditivo e visual no *offset* do movimento resultou em melhorias comparativamente a apresentações visuais unimodais. Este resultado é contrário a um número considerável de estudos que documentaram efeitos de facilitação do processamento visual induzidos por sons síncronos (e.g., Frassinetti et al., 2002; Leone & McCourt, 2013; Noesselt et al., 2008; Seitz et al., 2006; Staufenbiel et al., 2011; Stein, London, Wilkinson, & Price, 1996; Van der Burg, Cass, Olivers, Theeuwes, & Alais, 2010; Van der Burg et al., 2008; Vroomen & de Gelder, 2000, 2004). Utilizando uma tarefa de *flash-lag* (que avalia a localização percebida de um *flash* estático relativamente a um estímulo móvel com o qual se encontra espacialmente alinhado) Vroomen & de Gelder (2004) verificaram efeitos distintos dos sons apresentados antes ou após o *flash*, por um lado, e em sincronia com o *flash*, por outro. Enquanto os primeiros atraíam a ocorrência temporal do *flash* (enviesamentos de atração), sons síncronos produziam um efeito de melhoria audiovisual, isto é, uma redução do erro característico da tarefa (*flash-lag error*: perceber o *flash* como atrás do estímulo móvel com o qual se encontrava alinhado). Os autores atribuíram esta melhoria a uma definição mais precisa das fronteiras temporais do *flash* devida ao som simultâneo, beneficiando assim o processamento visual do *flash*. Um resultado semelhante poderia esperar-se no caso da modulação do MR visual pelo *offset* do som quando os sinais auditivo e visual eram síncronos. Dada a maior resolução temporal da audição relativamente à visão, uma melhoria na localização poderia, como no estudo de Vroomen & de Gelder (2004), resultar de um refinamento temporal do *timing* do *offset* do som, reduzindo a magnitude do erro para diante. Por que não ocorreu tal melhoria nos estudos mencionados é assim uma

questão relevante, coincidente com a de saber se, e em que condições, essa melhoria seria possível.

Dois tipos de fatores moduladores do efeito de sons síncronos no processamento visual têm sido destacados na literatura. Um deles diz respeito à consistência das contingências audiovisuais incluídas na tarefa. Lippert et al. (2007) mostraram neste sentido que o efeito benéfico de um som síncrono numa tarefa de detecção de contraste visual é impedido se se acrescentarem à tarefa, de forma aleatória, algumas assincronias audiovisuais (± 150 ms, ± 400 ms). Li et al. (2015) mostraram de forma convergente que num contexto de menor consistência espaço-temporal entre os sinais auditivos e visuais (inclusão na tarefa de assincronias e desfasamentos de lateralidade), um som síncrono cessa de produzir efeitos de facilitação em tarefas de discriminação visual. Targher, Micciolo, Occelli, & Zampini (2017) reportaram que a passagem de assincronias unidirecionais (e.g., som precede sempre) a assincronias bidirecionais (som precede às vezes, sucede noutras) numa tarefa reduz significativamente a melhoria audiovisual obtida nos ensaios com sons síncronos. A importância do número de assincronias utilizadas na tarefa tem sido igualmente suportada noutros contextos, como o do estudo de algumas ilusões audiovisuais (ver, e.g., Chan, Connolly, & Setti, 2018). No conjunto, estes e outros estudos similares (e.g., Busse et al., 2005; Liu, Stevens, & Carrasco, 2007) ilustram a capacidade do padrão de contingências audiovisuais (assincronias, em particular) incluído na tarefa para promover ou cancelar os efeitos benéficos de sons simultâneos com eventos visuais. Uma explicação conjeturável para a ausência de melhorias nas condições síncronas dos estudos de MR intermodal poderia assim residir no padrão específico (número, consistência) das contingências temporais audiovisuais manipuladas.

Um segundo tipo de fator modulador envolve a saliência perceptiva do som. Lippert et al. (2007) concluíram pela necessidade de associações intermodais perceptíveis (para além de consistentes) para a emergência de benefícios audiovisuais induzidos pelo som. Vroomen & de Gelder (2000) mostraram que, para beneficiar a detecção de um alvo visual, um som síncrono tinha, adicionalmente que ser claramente segregável de uma sequência de sons, não produzindo qualquer facilitação quando integrado numa melodia ou tornado menos abrupto. Resultados similares foram reportados por Bruns & Getzmann (2008) e por Watanabe & Shimojo (2001). A importância de aspetos de saliência perceptiva foi igualmente destacada no estudo da recalibração temporal audiovisual, da percepção de ordem temporal intermodal e da aprendizagem de associações audiovisuais (Heron et al., 2010; Hidaka et al., 2015; Ikumi & Soto-Faraco, 2014). Uma segunda explicação conjeturável para a ausência de melhorias na

localização nas condições síncronas dos estudos de MR é assim a uma insuficiente saliência perceptiva das contingências audiovisuais, plausivelmente associada à pequena magnitude das assincronias utilizadas. Em concordância com esta hipótese, apenas um de quatro participantes (isto é, 25%) na experiência 1 de Teramoto, Hidaka, et al. (2010) declararam ter percebido (por vezes) as assincronias.

O grau em que os resultados de enviesamento de atração para o som e de ausência de melhoria audiovisual nas condições síncronas, apesar de transversais a todos os estudos anteriores de MR audiovisual, são inevitáveis e independentes do padrão específico de contingências audiovisuais deve assim considerar-se como uma questão em aberto na literatura. Através da manipulação da consistência das associações audiovisuais e do seu grau de saliência perceptiva (fatores moduladores da promoção-cancelamento de benefícios audiovisuais), o presente estudo empírico procura examiná-la e oferecer-lhe uma resposta.

3.3.4. Relações entre deslocamento M e deslocamento O no momento representacional visual, auditivo e audiovisual

As relações entre o deslocamento M, no sentido do movimento, e do deslocamento O, ortogonal ao movimento, não foi praticamente considerada até à data em estudos anteriores. No caso de movimentos horizontais (que se verifica em todos os estudos de MR intermodal reportados acima) o deslocamento ortogonal coincide com a direção da gravidade, apresentando caracteristicamente o erro de localização para baixo por vezes designado como gravidade representacional. O estudo das relações entre momento representacional (MR) e gravidade representacional (GR) tem passado sobretudo pela tentativa de apreender diferenças e semelhanças na forma como os dois fenómenos são modulados por diferentes variáveis (e.g., os movimentos oculares, a modalidade de resposta, o atraso imposto entre o movimento de desaparecimento do alvo e a resposta – dinâmica temporal do erro, etc.: ver, para uma revisão, Hubbard, 2020) no domínio exclusivamente visual. Todas as conclusões sugeridas se mantêm provisórias, apoiadas em estudos maioritariamente não replicados, e assentes numa extrapolação dos efeitos diferenciais das variáveis moduladoras para diferentes mecanismos subjacentes (diferentes efeitos implicariam diferentes mecanismos). Para além disso, vários dos resultados utilizados nas comparações entre MR e GR referem-se a manipulações separadas em tarefas com movimento horizontal e com movimento vertical, não a resultados obtidos numa mesma tarefa de localização (ver Hubbard, 2020).

A avaliação descritiva da associação entre medidas de deslocamento M e de deslocamento O recolhidas numa mesma resposta a um movimento horizontal através dos participantes, expressa através da sua correlação, tem sido raramente reportada. Embora não tenha implicações diretas na demonstração de uma similaridade ou diferença de mecanismos (por definição, correlações não admitem interpretações causais), constitui uma outra forma de obter informações potencialmente úteis e suscetíveis de contribuir, indiretamente, para o debate sobre os mecanismos subjacentes. A dissertação não publicada de Martins (2015) constitui, no nosso melhor conhecimento, o único exemplo de consideração sistemática das correlações entre deslocamento M e deslocamento O numa tarefa de MR visual com movimento horizontal e resposta de localização motora. Correlações negativas significativas foram encontradas por esta autora em duas amostras de participantes (com e sem paralisia cerebral), significando que participantes que produziam maior deslocamento M apresentavam também maior deslocamento O (sendo medido relativamente à coordenada vertical zero, o aumento do deslocamento O é assinalado por valores cada vez mais negativos). Este resultado seria, por exemplo, compatível como assinala Martins (2015), com a hipótese de König et al. (2013) (ver também König & Krüger, 2006) de que toda a previsão sensorial é constringida pelo reportório específico de ações do organismo – hábitos de planeamento da interseção de objetos específicos de cada participante exprimir-se-iam assim de forma correlacionada na antecipação dos deslocamentos para diante e para baixo. Seria também consistente com a proposta de Munger et al. (1999) de um envolvimento da imaginaria mental na representação do movimento e na sua extrapolação, assumindo adicionalmente que a velocidade de transformação das representações é uma característica estável de cada participante. Ou ainda com variantes desta última proposta em que a vivacidade da imaginaria, e não a velocidade de transformação, determinaria a extrapolação do movimento (Senior, Barnes, & David, 2001; Willems, 2009). Apesar de se tratar de uma simples medida de associação, o exame da correlação entre os dois tipos de deslocamento pode neste sentido favorecer ou desfavorecer seletivamente diferentes propostas relativas aos mecanismos subjacentes ao MR.

No presente estudo serão pela primeira vez investigadas as relações entre o deslocamento M e o deslocamento O em tarefas de MR auditivo (unimodal) e audiovisual, para além do exame das correlações observadas em condições visuais unimodais (fornecendo uma possibilidade de replicação aos resultados reportados por Martins, 2015).

PARTE II

CONTRIBUTOS EMPÍRICOS

CAPÍTULO 4

PROCESSAMENTO AUDIOVISUAL NUMA TAREFA DE MOMENTO REPRESENTACIONAL: EFEITOS DA DURAÇÃO E DA ESPACIALIZAÇÃO DE UM SOM NA LOCALIZAÇÃO DE UM ALVO VISUAL

O conjunto de experiências que a seguir se apresenta procura responder às questões suscitadas em três estudos anteriores assentes em tarefas de momento representacional (MR) com alvos visuais (visão como modalidade relevante) e sons associados (a audição como modalidade não-relevante) – descritos na secção 3.2 do Capítulo 3, e aqui tomados como objeto de comparação, bem como às questões identificadas como em aberto na secção 3.3 do mesmo capítulo. Enquanto a maioria das investigações sobre o impacto do som no processamento visual do movimento se concentraram na deteção do movimento e/ou de propriedades como a sua direção, velocidade, coerência ou organização fenomenal (ver revisão de Teramoto et al., 2018), a investigação aqui apresentada partilha com os três estudos mencionados o entendimento de que o seguimento/localização de alvos constitui o teste mais direto à capacidade da audição para modular aspetos espaciais da perceção visual do movimento e de que as tarefas de MR intermodal fornecem um quadro operacional adequado para a produção desse teste. À semelhança dos estudos de comparação, por conseguinte, investigamos aqui os efeitos de disparidades audiovisuais na localização percebida do ponto de desaparecimento de um alvo visual em movimento. A escassez, até à data, de estudos de MR em contextos intermodais (Merz et al., 2020) poderá refletir a convicção de que a superioridade da visão sobre as restantes modalidades em tarefas de localização espacial se acentua particularmente com estímulos dinâmicos, envolvendo sinais (visualmente salientes) de movimento (Dick et al., 1987). Esta premissa não dispõe contudo

de suporte empírico claro. O presente trabalho pode, neste sentido, entender-se em parte como um teste empírico adicional ao pressuposto de uma preponderância completa da visão sobre a audição no processamento da localização espacial de estímulos em movimento.

4.1. HIPÓTESES E PREVISÕES

Um aspecto comum aos três estudos de comparação – para além da identificação de *enviesamentos de atração* em que a posição temporal (no caso de Schmiedchen et al., 2012, porventura também espacial) do som atraiu a localização percebida do termo do movimento visual – foi a ausência de efeitos benéficos de sons sincronizados com o termo do movimento visual. Este resultado é surpreendente, considerando o número de estudos que na literatura assinalam melhorias no processamento visual devidas a sons síncronos (e.g., Frassinetti et al., 2002; Seitz et al., 2006; Staufenbiel et al., 2011; Stein et al., 1996), incluindo em tarefas de localização espacial relativa de estímulos visuais (e.g., Vroomen & de Gelder, 2004).

Com base na evidência disponível sobre a capacidade do padrão de contingências AV numa tarefa promover ou cancelar efeitos benéficos de sons ocorrendo em simultâneo com eventos visuais (ver secção 3.3.3: estudos de Li et al., 2015; Lippert et al., 2007; Targher et al., 2017), colocámos a hipótese de que o número (afetando potencialmente a sua fiabilidade percebida) e magnitude (afetando potencialmente a sua saliência percebida) das disparidades AV fossem responsáveis por esse resultado. Em conformidade, uma manipulação central no nosso estudo diz respeito ao número e tamanho das disparidades AV utilizadas. Em contraste com os estudos de comparação, que incluíram seis assincronias AV, todas inferiores a 195 ms, no presente estudo foram utilizadas quatro assincronias com valores de ± 250 ms e ± 500 ms. Estes valores estão acima dos limiares de deteção de assincronia estabelecidos para a maioria das situações audiovisuais (ver Eg & Behne, 2015; Eg, Griwodz, Halvorsen, & Behne, 2015; van Eijk, Kohlrausch, Juola, & van de Par, 2008), sendo genericamente comparáveis aos valores utilizados por Li et al. (2015), Lippert et al. (2007) e Targher et al. (2017).

A hipótese subjacente a esta manipulação é a de que favoreceria a perceção da relevância para a tarefa das disparidades AV utilizadas e, desse modo, a ocorrência de benefícios induzidos por sons síncronos (e, eventualmente, por outras contingências temporais). Em termos de resultados esperados, corresponde às seguintes previsões: (1) em contraste com os estudos de comparação, as condições de sincronia AV deverão dar lugar no

nosso estudo a benefícios AV; (2) mesmo em condições de assincronia AV, é possível que a relevância percebida para a tarefa da informação auditiva permita suportar melhorias da execução (ao invés de enviesamento de atração). Do ponto de vista dos processos, é esperado que esta manipulação dirigida à percepção da relevância das associações AV para os objetivos da tarefa mobilize mecanismos multissensoriais distintos dos que suportam os de enviesamento de atração, operando à margem das janelas temporais estritas ($\pm \approx 100-150$ ms) evidenciadas nos estudos de comparação (Werner & Noppeney, 2010).

Uma segunda hipótese orientadora foi a de que a informação especificamente espacial do som (disparidades AV espaciais) teria um efeito modulador no momento representacional. A possibilidade de um impacto da informação espacial do som na localização de alvos estáticos encontra-se estabelecida (Alais & Burr, 2004; Battaglia et al., 2003). No caso do momento representacional, alguma evidência nessa direção foi reportada por Chien et al. (2013) e Schmiedchen et al. (2012), apoiando a ideia de um efeito direto, não mediado temporalmente, das propriedades espaciais do som. No entanto, como discutido acima (ver secção 3.3.1), em nenhum dos casos os resultados podem ser tomados como conclusivos, encontrando-se afetados por um problema metodológico central – a incapacidade de dissociar experimentalmente a informação espacial e temporal do som (Kawachi, 2016; Teramoto et al., 2018).

Com vista a resolver este problema, uma manipulação distintiva no nosso estudo consistiu na utilização de um desenho fatorial completo, combinando disparidades AV no início (*onset*) e no termo (*offset*) do movimento visual, em três tarefas contrabalanceadas envolvendo respetivamente: (1) assincronias AV veiculadas por sons contínuos não espacializados (tarefa temporal: T); (2) disparidades espaciais AV veiculadas por sons espacializados em movimento (com base em diferenças interaurais e uma função média de transferência relacionada com a cabeça, HRTF, apresentadas por meio de auscultadores), sincronizados e com a mesma duração do movimento visual (tarefa espacial: E); (3) disparidades temporais e espaciais covariadas, obtidas a partir do emparelhamento dos dois tipos de disparidades anteriores (tarefa espaço-temporal: ET).

O desenho fatorial desempenha aqui um papel metodológico central na dissociação das pistas temporais e espaciais do som, de acordo com o seguinte racional. O uso de estímulos AV sincronizados na tarefa E (disparidades espaciais) significa que movimentos auditivos e visuais com diferentes extensões (especialmente díspares), mas com a mesma duração (síncronos) foram obtidos pela utilização de diferentes velocidades da fonte sonora. Uma vez que a combinação de cada ponto de termo (*offset*) do movimento auditivo com

diferentes pontos de origem (*onsets*) corresponde na tarefa E a diferentes velocidades do estímulo sonoro, um efeito estatisticamente nulo do *onset* permitirá excluir um efeito da velocidade e atribuir uma eventual modulação do momento representacional visual às disparidades espaciais AV no termo do movimento. No caso de um efeito significativo das disparidades AV no *onset*, a estrutura fatorial do desenho deverá ainda assim permitir esclarecer o efeito das disparidades espaciais no *offset* através das suas possíveis interações com a velocidade do movimento auditivo. A hipótese de que correlacionar disparidades espaciais AV com a velocidade do som e não com a sua duração (implicando uma confusão com disparidades temporais) pode suportar uma dissociação operacional entre pistas acústicas temporais e espaciais encontra apoio em evidência sobre o carácter derivado da percepção da velocidade de fontes sonoras, relativamente à percepção da sua duração e distância percorrida (Carlile & Leung, 2016; T. C. A. Freeman et al., 2014; Roggerone, Vacher, Tarlao, & Guastavino, 2019). Adicionalmente, o desenho fatorial *onset* × *offset* permite esclarecer em geral o papel destes dois fatores, em particular do *onset*, relativamente ao qual se verificam divergências nos estudos de comparação (ver acima, secção 3.3.2, a discrepância de resultados entre Teramoto, Hidaka, et al. (2010) e Chien et al. (2013).

As duas hipóteses anteriores (e previsões associadas) dizem respeito em particular ao deslocamento M (momento representacional), não ao deslocamento O (gravidade representacional, no caso de movimentos horizontais). Na medida em que as manipulações utilizadas envolveram apenas a direção horizontal (azimute) do movimento (distâncias horizontais percorridas, origem do movimento, pontos de desaparecimento), as previsões relativas ao seu impacto na localização vertical percebida do alvo não são diretas. O típico efeito registado em experiências com alvos visuais animados de movimento horizontal é o de um deslocamento O para baixo, na direção da gravidade. Uma previsão genérica, baseada na hipótese de uma similaridade geral do momento representacional através de diferentes modalidades, é a de que condições visuais unimodais, auditivas unimodais e audiovisuais deveriam produzir identicamente deslocamentos O compatíveis com a gravidade representacional. O presente estudo, que inclui igualmente condições visuais unimodais intercaladas entre as apresentações audiovisuais e uma tarefa auxiliar com alvos auditivos unimodais, é, do nosso melhor conhecimento, o primeiro a proceder a uma avaliação empírica desta previsão.

Quanto aos efeitos mais específicos de variáveis como a distância percorrida, o ponto de desaparecimento ou a natureza das disparidades (temporais, espaciais ou espaço-temporais) no deslocamento O, não existe qualquer base anterior para previsões. Considerando, no

entanto, a existência documentada de divergências entre o MR visual e o MR auditivo quanto ao impacto modulador de algumas variáveis (designadamente, a velocidade do movimento e o papel dos movimentos oculares: ver secção 1.4 e, e.g., Getzmann, 2005a; Getzmann & Lewald, 2009; Schmiedchen et al., 2013) é possível esperar que, a par de um paralelismo global na direção do erro vertical, se verifiquem diferenças entre condições modais quanto ao impacto de variáveis moduladoras particulares.

As relações entre o deslocamento M e o deslocamento O constituem também aqui um objeto de interesse possível. Apenas um resultado, associado a uma dissertação de doutoramento não publicada (Martins, 2015) fornece um ponto de partida e termo de comparação. Esse resultado envolvendo apenas o MR visual, apontou para correlações positivas significativas entre os dois tipos de deslocamento (maior deslocamento M associado estatisticamente a maiores deslocamentos O) em amostras tanto de participantes afetados por paralisia cerebral como de participantes controlo. O exame da associação estatística dos dois tipos de erro (M e O) em condições visuais unimodais, auditivas unimodais e audiovisuais oferece, para além de um primeiro teste à generalidade do resultado obtido por Martins (2015) no domínio do MR visual, uma nova oportunidade de avaliação empírica de semelhanças e/ou diferenças entre condições modais no âmbito dos erros de localização espacial do movimento. Pode ainda contribuir com elementos úteis para o debate sobre os mecanismos subjacentes ao MR e à GR. Apesar de uma associação estatística, ou a ausência dela, não ter implicações diretas para a discussão de mecanismos causais, a ocorrência de correlações significativas consistentes entre as duas formas de deslocamento favoreceria explicações do MR que previssem essa correlação (e.g., uma extrapolação do movimento baseada na velocidade de transformação mental característica de cada participante: Munger et al, 1999), enquanto a sua ausência, pelo contrário, as desfavoreceria.

4.2. A TAREFA BASE: OPÇÕES OPERACIONAIS

As três tarefas utilizadas (temporal, espacial, espaço-temporal) constituem variantes de uma mesma tarefa de base, assente no desenho *onset* × *offset* e na instrução de localização da última posição percebida do alvo visual. Uma vantagem particular da utilização de um paradigma de MR no nosso estudo prende-se com o facto de o erro de localização para diante se ter revelado modulável tanto por mecanismos perceptivos de baixo nível (e.g., persistência

retiniana, enviesamento foveal), como por mecanismos de alto nível (e.g., conhecimento sobre a identidade e categoria do alvo) ou ainda formas variáveis de combinação entre processos *bottom-up* e *top-down* (e.g., antecipação perceptiva, incorporação de regularidades do ambiente) (ver secção 1.1 e 1.2 e, para revisões, Hubbard, 2005, 2014a, 2017). Esta característica surge como importante num estudo destinado a averiguar a possibilidade de, através da manipulação da relevância percebida para a tarefa de contingências AV, mobilizar processos intermodais distintos dos que usualmente têm sido destacados (de baixo nível, regidos por janelas temporais estritas de integração sensorial), conducentes a padrões de modulação distintos do processamento visual do movimento (e.g., benefícios AV *versus* enviesamentos de atração).

Algumas das escolhas operatórias inerentes à tarefa de base requerem justificação. A modalidade de resposta escolhida foi o posicionamento de um cursor controlado por um rato ótico (sem fios) sobre o ponto onde o alvo visual foi percebido pela última vez. Este método foi também utilizado por Chien et al. (2013), partilha com o método de apontar utilizado por Schmiedchen et al. (2012) a natureza motora (sensoriomotora) da resposta e distingue-se do método de julgamento perceptivo utilizado por Teramoto, Hidaka, et al. (2010). Relativamente a este último (os participantes avaliam se um estímulo sonda subsequente se encontra no mesmo local em que o alvo desapareceu ou num local diferente), a localização com rato ótico apresenta as seguintes vantagens: maior precisão da resposta (ao nível do pixel), medidas diretas da dispersão das respostas individuais, redução importante do número de ensaios necessários (ver Hubbard, 2005). Adicionalmente produz deslocamentos M de maior magnitude do que o método de julgamento perceptivo (Kerzel & Gegenfurtner, 2003), uma característica desejável para o objetivo de estudar o impacto da informação auditiva nos erros de localização. Como assinalado em Hubbard (2005), a localização através de um dispositivo apontador (rato) constitui a modalidade típica de resposta em estudos de MR baseados em alvos com movimento contínuo (*smooth motion*) – caso tanto do presente estudo como dos estudos de comparação.

Em contraste com todos os estudos de comparação, não foram impostos constrangimentos aos movimentos oculares dos participantes. A exigência de fixação ocular num ponto durante a tarefa (ou simplesmente de olhar em frente), foi adotada nos estudos anteriores por diferentes razões. No caso de Schmiedchen et al. (2012), tinha como objetivo possibilitar a comparação entre os efeitos de disparidades AV espaciais em visão central e na periferia, onde a fiabilidade da informação espacial auditiva aumenta relativamente à da visão (Hidaka et al., 2011; Perrott, Costantino, & Cisneros, 1993). Embora mais acentuados

na periferia, efeitos do som foram também observados em visão central. Contudo, dada a covariação sistemática das disparidades espaciais e temporais no estudo de Schmiedchen et al. (2012), é impossível estabelecer se estes efeitos, designadamente os documentados em visão central, se devem à informação espacial do som. Ao permitir movimentos oculares de seguimento do alvo (que o mantêm em visão central), em conjunto com uma dissociação experimental eficaz das pistas acústicas espaciais e temporais, procurámos dar resposta à questão irresolvida sobre se a informação auditiva espacial é capaz de modular o MR visual em condições de visão central/foveal.

No caso de Teramoto, Hidaka, et al. (2010), a exigência de fixação constante num ponto visou evitar potenciais “efeitos de confusão” devidos aos movimentos oculares. Na realidade, esta preocupação reflete sobretudo uma conceção particular do MR, entendido como um fenómeno exclusivamente sensorio-percetivo, contrária a outras conceções nas quais o planeamento motor e o processamento sensorio-motor constituem ingredientes essenciais do erro de localização para diante (J. S. Jordan et al., 2002). No atual estágio do debate, uma preocupação simétrica de que a restrição dos movimentos oculares possa encobrir dimensões centrais do MR (dos quais o próprio movimento ocular dependeria), é assim igualmente legítima (ver Hubbard, 2005, 2006, 2010, 2014). Adicionalmente, o MR com movimentos contínuos diminui significativamente, podendo mesmo desvanecer-se, quando o seguimento oculomotor é impedido (De Sá Teixeira et al., 2013; Kerzel, 2000; Stork & Müsseler, 2004). De um ponto de vista operacional, essa é uma característica indesejável no contexto de uma investigação sobre a modulação do MR visual pela informação do som.

4.3. MÉTODO

4.3.1. Participantes

Participaram como voluntários neste estudo 31 estudantes (25 do sexo feminino e 6 do sexo masculino) da Universidade de Coimbra e da Escola Superior de Tecnologia da Saúde de Coimbra do Instituto Politécnico de Coimbra, com idades compreendidas entre os 18 e os 22 anos ($M = 19.35$; $DP = 1.603$). Vinte e um (21) dos participantes receberam créditos do curso pela sua participação. Todos os participantes desconheciam o objetivo da experiência. Nenhum reportou a existência de problemas auditivos e todos apresentaram visão normal

ou corrigida para o normal. Foi obtido o seu consentimento informado, livre e esclarecido e todos os aspetos dos procedimentos obedeceram aos requisitos da Declaração de Helsínquia.

4.3.2. Estímulos

Os estímulos consistiram em animações *.avi* (a 60 frames por segundo: fps) incluindo apresentações visuais unimodais (sem som: V) e audiovisuais (associadas a sons: AV) de um disco (alvo) cinzento claro (modelo de cor RGB: 200, 200, 200) com 1 cm de diâmetro (≈ 35 pixéis [px] na resolução de ecrã utilizada; $\approx 0.7^\circ$ de ângulo visual [a.v.] a uma distância de visão de cerca de 84.5 cm), movendo-se horizontalmente da esquerda para a direita sobre um fundo preto a uma velocidade constante (8 cm/s ≈ 281 px/s $\approx 5.4^\circ$ a.v./s) e desaparecendo subitamente num ponto variável do ecrã. Esta velocidade situa-se na gama inferior das velocidades anteriormente utilizadas em estudos de MR visual (ver De Sá Teixeira et al., 2013; De Sá Teixeira & Oliveira, 2014; Hubbard, 1990; Hubbard & Bharucha, 1988). No contexto da perceção do movimento auditivo, velocidades mais baixas (desde que acima de 1° a.v./s: Perrott & Saberi, 1990) facilitam, em geral, a localização de fontes sonoras em movimento (Carlile & Leung, 2016; Chandler & Grantham, 1992; Perrott & Musicant, 1977), uma propriedade favorável no quadro deste estudo.

Os estímulos visuais unimodais (V), gerados no Interactive Physics 2000 v5.2.1.1 (MSC.Software Corporation, 2000), foram comuns a todas as tarefas. As componentes áudio dos estímulos audiovisuais (AV), por outro lado, incluíram três conjuntos distintos, um para cada tarefa (T, E e ET), correspondentes a diferentes condições de informação auditiva (temporal, espacial, espaço-temporal). Os sons foram gerados no Audacity v2.0.2 (Audacity Team, 2012), consistindo em ruído branco filtrado por um filtro passa-banda (largura de banda de 2 kHz, centrada em 2 kHz, frequência de amostragem de 44.1 kHz, precisão simples de *32-bit floating depth*; *fade-in* e *fade-out* linear de 15 ms). Esta banda de frequências é comparável com a amplitude de banda de 1-3 kHz utilizada por Getzmann e colaboradores no estudo do efeito Fröhlich (Getzmann, 2005b) e do momento representacional auditivo (Getzmann & Lewald, 2007, 2009; Getzmann et al., 2004). Diferentemente da localização de estímulos tonais e de banda estreita, que perde precisão nas frequências médias centradas em 2 kHz (Middlebrooks & Green, 1991), a precisão da localização de ruído filtrado de banda larga (duas oitavas ou mais) não depende da frequência central (Risoud et al., 2018; Yost, Loissele, Dorman, Burns, & Brown, 2013; Yost & Zhong, 2014).

A duração dos sons variou entre 0.5 s a 3 s, de acordo com as exigências do desenho experimental (ver **Figura 1**). Estes valores de duração foram obtidos, como explicado na secção seguinte, a partir das durações selecionadas para o movimento do alvo visual (1 s e 2 s), situadas na gama de durações comumente utilizada em estudos anteriores (e.g., Teramoto, Hidaka, et al., 2010; e Schmiedchen et al., 2012; Feinkohl et al., 2014). A espacialização do som, requerida nas tarefas E e ET, foi realizada com o slab3d v6.6.1 (NASA & Miller, 2012), utilizando diferenças interaurais de intensidade (*interaural level differences* – ILD) e de tempo (*interaural time differences* – ITD) e uma função média de transferência relacionada com a cabeça (*head related transfer function* – HRTF) para apresentação em auscultadores (Miller & Wenzel, 2002). Os sons espacializados na tarefa E deslocavam-se a velocidades constantes de 4, 6, 8, 10, ou 12 cm/s ($\approx 140, 211, 281, 351$ ou 421 px/s; $\approx 2.7^\circ, 4.1^\circ, 5.4^\circ, 6.8^\circ, 8.2^\circ$ a.v./s), de modo a percorrerem diferentes distâncias durante o tempo correspondente à duração do movimento visual. Todos os sons foram guardados como ficheiros .wav estéreo e combinados subsequentemente com os arquivos vídeo através do VirtualDub v1.9.11 (SourceForge, 2010). A duração de cada um dos vídeos resultantes (correspondentes aos estímulos audiovisuais) estendia-se até 500 ms após a desapareição do disco cinzento em movimento (isto é, incluía 30 *frames* posteriores aquele em que o disco era apresentado pela última vez).

4.3.3. Desenho experimental

Os diagramas da **Figura 1** (A e B) ilustram a lógica do desenho experimental, válida para todas as tarefas. O movimento visual é representado pelos retângulos pretos. O retângulo preto em (A) representa uma distância de 8 cm (≈ 281 px $\approx 5.4^\circ$ a.v.) percorrida pelo alvo visual, ou, de modo equivalente, uma duração de deslocamento de 1 s a uma velocidade constante de 8 cm/s ($d1: \approx 281$ px/s $\approx 5.4^\circ$ a.v./s). O retângulo preto em (B) representa uma distância percorrida de 16 cm (≈ 562 px $\approx 10.8^\circ$ a.v.) ou, de modo equivalente, uma duração de deslocamento de 2 s a uma velocidade constante de 8 cm/s ($d2: \approx 562$ px/s $\approx 10.8^\circ$ a.v./s). Distâncias/durações semelhantes têm sido utilizadas em estudos de MR visual (De Sá Teixeira et al., 2013; De Sá Teixeira & Oliveira, 2014; Hubbard & Lange, 2010; Hubbard & Ruppel, 1999). Os retângulos cinzentos em ambos os painéis representam durações do som (e.g., no caso dos sons não-lateralizados da tarefa T) ou distâncias percorridas pelo som (e.g., no caso dos sons espacializados da tarefa E), ilustrando as disparidades relativas entre o

início (*onset*) e termo (*offset*) dos sons e o início e termo do movimento visual. As disparidades representadas podem ser temporais (tarefa T) espaciais (tarefa E) ou espaciais e temporais (tarefa ET). Em todos os casos, correspondem a uma fração constante ($\pm 25\%$) da duração do movimento visual e/ou da sua distância percorrida. Por conseguinte, as assincronias AV no início e termo do movimento foram de -250, 0, +250 ms para movimentos visuais com duração de 1 s (d1), e de -500, 0, +500 ms para movimentos visuais de 2 s (d2). Da mesma forma, as disparidades espaciais AV foram de -2, 0, +2 cm (≈ 70 px $\approx 1.4^\circ$ a.v.) para movimentos visuais de 8 cm (d1) e de -4, 0, +4 cm (≈ 140 px $\approx 2.7^\circ$ a.v.) para movimentos de 16 cm. Na tarefa ET os dois tipos de disparidades foram apresentados de forma covariada; por outras palavras, disparidades temporais e espaciais com o mesmo sinal (+/-) e o mesmo tamanho proporcional relativamente à distância percorrida/duração do movimento visual foram emparelhadas: -250 ms e -2 cm, -500 ms e -4 cm, +250 ms e +2 cm, +500 ms e +4 cm.

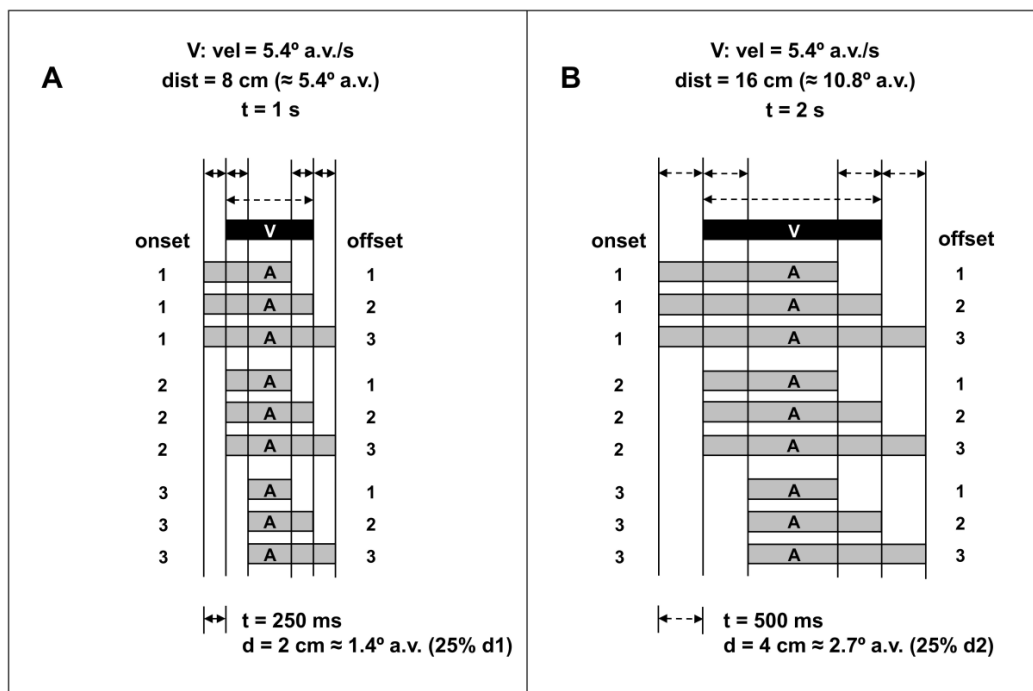


Figura 1. Representação esquemática do desenho experimental. Os dois retângulos pretos nos painéis A e B representam dois movimentos do alvo visual envolvendo distintas durações/distâncias percorridas (d1 e d2). Os dois conjuntos de retângulos cinzentos situados abaixo representam a duração e/ou distância percorrida pelos sons. Em cada um dos painéis, os desvios horizontais entre as extremidades dos retângulos pretos e cinzentos representam as disparidades audiovisuais no início (*onset*) e termo (*offset*) do movimento. Estas disparidades podem ser temporais, espaciais, ou espaço-temporais. Cada painel ilustra um desenho fatorial 3 (disparidades AV no *onset*) \times 3 (disparidades AV no *offset*).

Embora não indicado na **Figura 1**, cada uma das distâncias percorridas pelo alvo visual esteve associada a dois pontos distintos de desaparecimento, localizados 3 cm (≈ 105 px $\approx 2^\circ$ a.v.) à esquerda e 5 cm (≈ 176 px $\approx 3.4^\circ$ a.v.) à direita da linha média horizontal do ecrã. Isto permitiu que tanto alvos em movimento para o centro como afastando-se do centro do ecrã fossem incluídos na tarefa (ver Hubbard & Ruppel, 1999; Munger et al., 2006), evitando ao mesmo tempo uma simetria exata que poderia operar como uma informação posicional para os participantes. Estes valores são próximos dos $2^\circ - 3^\circ$ à direita ou esquerda do centro utilizados como pontos de desaparecimento por Teramoto, Hidaka, et al. (2010) e não muito distantes da oscilação horizontal (*jitter*) de $\pm 0.3^\circ$ a.v. em torno do ponto médio do ecrã empregue para o mesmo efeito por Chien et al. (2013). Considerando o tamanho das disparidades espaciais AV utilizadas, implicam que o termo (*offset*) dos movimentos auditivos (sons espacializados) cobriu uma gama de posições entre os $\approx -4.8^\circ$ a.v. (≈ -246 px) à esquerda do centro e os $\approx 6.1^\circ$ a.v. ($\approx +316$ px) à direita, nunca se aproximando do bordo direito do ecrã a $\approx 13.9^\circ$ a.v. (≈ 720 px) do centro.

O desenho em cada tarefa (T, E e ET) correspondeu assim a um desenho fatorial de medidas repetidas 3 (*onsets* auditivos: doravante, *onset*) $\times 3$ (*offsets* auditivos: doravante, *offset*) $\times 2$ (distâncias/durações: doravante, distância) $\times 2$ (pontos de desaparecimento: doravante, ponto). Cada condição definida pelo cruzamento dos níveis de *onset* e *offset* ocorreu deste modo quatro vezes. Para além das 36 apresentações AV correspondentes ao desenho principal, cada tarefa incluiu dois conjuntos extra de ensaios: (1) 4 apresentações de movimento visual sem som (duas distâncias percorridas com dois pontos de desaparecimento cada) permitindo comparar condições unimodais e bimodais; estas apresentações serão denotadas V_T, V_E e V_ET sempre que seja necessário distinguir entre os respetivos contextos audiovisuais; (2) 24 apresentações visuais e audiovisuais envolvendo várias distâncias percorridas e pontos de desaparecimento, todos distintos dos utilizados nos ensaios do desenho principal, de modo a contrariar o desenvolvimento de expectativas pelos participantes quanto aos pontos de desaparecimento. Estes ensaios incluíram 3 apresentações V e 21 AV (sete das quais sem disparidades), mobilizando oito pontos de desaparecimento e três distâncias percorridas, aleatoriamente distribuídas pelos 24 ensaios. As respostas a estes ensaios de controlo não foram registadas.

Os participantes realizaram as três tarefas em três sessões, separadas por alguns dias (M = 4.35, DP = 1.99, entre a primeira e segunda sessão; M = 4.45, DP = 1.65, entre a segunda e terceira sessão). Para garantir um controlo adicional dos efeitos de ordem, as tarefas foram

contrabalanceadas através dos participantes utilizando todas as ordenações possíveis ($3! = 6$). Tomando as tarefas T, E e ET como um fator intra-sujeitos adicional, o conjunto do estudo pode finalmente descrever-se como um desenho fatorial completo de medidas repetidas 3 (tarefa) $\times 3$ (*onset*) $\times 3$ (*offset*) $\times 2$ (distância) $\times 2$ (ponto).

Para além das tarefas principais, os participantes foram ainda convidados a realizar uma tarefa auxiliar que decorreu cerca de um mês após a última sessão (em dias: $M = 34.38$, $DP = 2.60$), consistindo em localizar o ponto de desaparecimento de sons em movimento (condições auditivas unimodais: A). Os estímulos foram um subconjunto dos arquivos áudio utilizados nas apresentações da tarefa E, limitados às condições em que o movimento auditivo se iniciava no mesmo local que o movimento visual (*onset* 2 nos painéis A e B da **Figura 1**) e terminava em cada um dos três *offsets* auditivos atribuídos (*offsets* 1, 2 e 3 na **Figura 1**). Esta tarefa destinou-se a testar o sucesso da espacialização dos sons, avaliando a exatidão da sua localização percebida. Todos os participantes (31) aceitaram realizá-la. Adicionalmente, um subconjunto de 10 participantes disponibilizou-se para um bloco de ensaios suplementar (contrabalanceado com o das apresentações auditivas), no qual identificaram a última posição percebida de alvos visuais em movimento (os mesmos utilizados nas tarefas principais). Diferentemente das condições visuais intercaladas em cada uma das tarefas (V_T, V_E e V_ET), estas apresentações unimodais ocorreram na ausência de contextos audiovisuais, sendo referidas abaixo como apresentações visuais isoladas (V_is).

4.3.4. Procedimento experimental

As animações foram apresentadas aleatoriamente num monitor LCD plano (Samsung, SyncMaster 940BW com 41×25.8 cm, resolução de ecrã de 1440×900 px e taxa de refrescamento de 60 Hz). Os estímulos auditivos foram apresentados em auscultadores (*Headphones* Sennheiser, HD 202; resposta de frequência: 18 Hz - 18 kHz; Sennheiser Electronic GmbH & Co. KG, Wedemark, Germany) com uma intensidade de 60 dB (A), medida à saída com um sonómetro (Brüel & Kjær, 2260; Brüel & Kjær, Nærum, Denmark). Todos os aspetos da apresentação dos estímulos e recolha das respostas foram geridos pelo programa SuperLab v4.0.7b (Cedrus Corporation, 2008).

As experiências decorreram na obscuridade, numa sala insonorizada (cabine insonorizada da Escola Superior de Tecnologia de Saúde de Coimbra). Os participantes encontravam-se sentados confortavelmente a uma distância de 84.5 cm do ecrã, com a

cabeça apoiada num suporte queixo-testa e os olhos alinhados com as linhas médias vertical e horizontal do ecrã. Uma gravação mono da palavra “atenção” (voz masculina, gravada no Praat v5.3.60; Boersma & Weenink, 2013) era apresentada dioticamente 1000 ms antes de cada estímulo, sem que os participantes fossem solicitados a fixar qualquer ponto ou região específica do ecrã (de modo a evitar “efeitos de marco”: Hubbard & Ruppel, 1999). Não foram impostas restrições aos movimentos oculares durante as tarefas. Os participantes foram instruídos a colocar um cursor “+”, controlado por um rato ótico sem fios, no local ocupado pelo centro do disco cinzento quando o perceberam pela última vez, clicando em seguida no botão esquerdo do rato. Comparado com o método do julgamento da localização de estímulos-sonda relativamente ao ponto de desaparecimento do alvo, apontar com um rato fornece uma medida direta de localização (com precisão ao nível do pixel) e requer um menor número de ensaios (Hubbard, 2005). As instruções não incluíam quaisquer referências aos sons apresentados para além da indicação de que a resposta pedida era a mesma quer o alvo visual fosse apresentado em silêncio ou acompanhado por um som. O cursor “+” tornava-se visível no centro do ecrã, 4 cm abaixo da linha média vertical, 500 ms após cada apresentação. Depois da resposta, a mensagem “clique para continuar” surgia no centro do ecrã até que o participante desse início a novo ensaio premindo um dos botões do rato. As respostas de localização foram registadas sob a forma de coordenadas XY do ecrã (em pixels). Previamente à realização de cada uma das tarefas, os participantes completaram um bloco de treino de 10 ensaios selecionados, aleatoriamente, entre os ensaios do bloco experimental, de modo a garantir a familiaridade com a tarefa e o bom cumprimento das instruções. O período de treino foi monitorizado e, sempre que necessário, foi dispensado *feedback* corretivo.

4.3.5. Análise dos dados

As análises gráficas e estatísticas assentaram sobretudo em ANOVAs de medidas repetidas conduzidas sobre os erros de localização. Os erros de localização horizontal foram, calculados como a diferença (em pixels) entre a coordenada x do ponto indicado pelo participante (resposta) e a coordenada x da localização final do alvo no ecrã. Erros positivos assinalam assim um deslocamento para diante da posição percebida do alvo e erros negativos um deslocamento para trás. Seguindo Hubbard (1995b, 1995c) referimo-nos ao erro horizontal como deslocamento M (isto é, na direção do movimento), uma expressão

teoricamente menos carregada do que a de “momento representacional”. Os erros de localização vertical foram, de modo semelhante, calculados como a diferença (em pixels) entre a coordenada y da resposta de localização do participante e a coordenada y da localização do alvo no ecrã (sempre a mesma: $y = 0$). Erros verticais negativos indicam assim localizações percebidas abaixo da linha em que decorreu o movimento horizontal, erros positivos localizações percebidas acima dessa linha. De acordo com Hubbard (1995b, 1995c), designamos o erro vertical como deslocamento O (ortogonal à direção do movimento), uma expressão teoricamente menos carregada do que a de “gravidade representacional”.

O valor $p = .05$ foi utilizado como critério de significância estatística nas análises. A correção épsilon de Greenhouse-Geisser foi usada em casos de violação de esfericidade (avaliada pelo teste de Mauchly). Todas as comparações aos pares subsequentes a ANOVAs foram ajustadas pela correção de Bonferroni para a inflação do erro de Tipo I. As comparações estatísticas de variâncias foram realizadas com o teste Pitman-Morgan para amostras correlacionadas (Morgan, 1939; Pitman, 1939; Snedecor & Cochran, 1989).

CAPÍTULO 5

DESLOCAMENTO M

Por razões de organização e simplicidade, os resultados para o deslocamento M (erro horizontal) e o deslocamento O (erro vertical) são apresentados e discutidos em capítulos separados. Dado que todas as manipulações realizadas envolveram o movimento horizontal e não o movimento vertical (fisicamente inexistente: invariância da coordenada y em todas as condições experimentais), a modulação do deslocamento M constitui o foco principal de interesse, tendo a análise do deslocamento O, neste contexto, sobretudo um valor complementar. A relação entre os resultados obtidos para os dois tipos de deslocamento (M e O) será objeto de análise no capítulo dedicado ao deslocamento O.

5.1. RESULTADOS

5.1.1. ANOVA global de medidas repetidas (5 vias)

Considerando o número elevado de variáveis independentes (5) que definem a estrutura de conjunto do estudo (envolvendo 5 efeitos principais, 10 interações de duas vias, 10 de três vias, 5 de quatro vias e 1 de cinco vias), optámos por, previamente à análise gráfica e estatística da operação dos fatores a diferentes níveis do desenho experimental (e.g., ao nível de cada tarefa) apresentar os resultados da ANOVA global de medidas repetidas 3 (tarefa) ×

3 (*onset*) × 3 (*offset*) × 2 (distância) × 2 (ponto) conduzida sobre os valores de deslocamento M.

Como pode ver-se na **Tabela 1**, todos os fatores exceto *onset* ($p = .942$) apresentaram efeitos principais significativos. Registaram-se três interações de duas vias estatisticamente significativas, todas envolvendo o fator distância: tarefa × distância ($p < .05$), *offset* × distância ($p < .001$), distância × ponto ($p < .01$). Os únicos resultados significativos para além destes dizem respeito a duas interações de três vias: tarefa × *offset* × distância ($p < .05$) e tarefa × *offset* × ponto ($p < .05$).

Tabela 1. Anova de medidas repetidas de 5 vias: 3 (tarefa) × 3 (*onset*) × 3 (*offset*) × 2 (distância) × 2 (ponto). A partir das interações de quatro vias, as designações dos fatores *Onset*, *Offset*, e Distância, foram abreviadas, respetivamente, para On, Off e Dist. Os valores de p reportados foram corrigidos pelo método de Greenhouse-Geisser sempre que a assunção de esfericidade foi violada. Por razões de simplicidade, a tabela apresenta nesses casos os valores não corrigidos de graus de liberdade (df) e ao valor de épsilon (ϵ) usado para a correção na coluna mais à direita da tabela. Os valores de p significativos ($p \leq .05$) estão assinalados por um asterisco.

Fatores	df_{effect}	df_{error}	F	p	η_p^2	ϵ
Tarefa	2	60	4.11	.032*	.12	.768
Onset	2	60	.06	.942	<.01	-
Offset	2	60	10.38	<.001*	.26	-
Distância	1	30	20.00	<.001*	.40	-
Ponto	1	30	12.13	.002*	.29	-
Tarefa × Onset	4	120	.70	.560	.02	.792
Tarefa × Offset	4	120	.60	.662	.02	-
Tarefa × Distância	2	60	3.96	.024*	.12	-
Tarefa × Ponto	2	60	.41	.538	.01	.532
Onset × Offset	4	120	.26	.904	<.01	-
Onset × Distância	2	60	1.34	.271	.04	-
Onset × Ponto	2	60	1.73	.192	.06	.802
Offset × Distância	2	60	11.59	<.001*	.28	-
Offset × Ponto	2	60	.95	.392	.03	-
Distância × Ponto	1	30	10.02	.004*	.25	-
Tarefa × Onset × Offset	8	240	1.08	.375	.04	-
Tarefa × Onset × Distância	4	120	.33	.799	.01	.739
Tarefa × Onset × Ponto	4	120	.83	.506	.03	-
Tarefa × Offset × Distância	4	120	2.60	.046*	.08	-
Tarefa × Offset × Ponto	4	120	2.97	.022*	.09	-
Tarefa × Distância × Ponto	2	60	1.37	.262	.04	-
Onset × Offset × Ponto	4	120	.817	.517	.03	-
Onset × Offset × Ponto	4	120	1.09	.366	.04	-
Onset × Distância × Ponto	2	60	.683	.477	.02	.788
Offset × Distância × Ponto	2	60	1.325	.273	.04	-
Tarefa × On × Off × Dist	8	240	1.182	.321	.038	.623

Tarefa × On × Off × Ponto	8	240	1.265	.263	.04	-
Tarefa × Off × Dist × Ponto	4	120	1.169	.321	.038	-
Tarefa × On × Dist × Ponto	4	120	.231	.921	.008	-
On × Off × Dist × Ponto	4	120	1.568	.187	.05	-
Tarefa × On × Off × Dist × Ponto	8	240	.643	.741	.021	-

A ausência de efeitos de *onset* (principal ou interações) indica que as disparidades AV apenas formam operatórias no *offset*. O efeito principal do fator tarefa ($p < .05$) indica que o tipo de pista auditiva (temporal, espacial ou espaço-temporal) modulou a magnitude do deslocamento M. A interação estatisticamente nula tarefa × *offset* ($p = .662$) sugere adicionalmente que isso não alterou, em todo o caso, o modo como as disparidades AV operaram através das tarefas.

A distância e o ponto de desaparecimento são variáveis teoricamente menos relevantes (tendo em conta as questões em apreço), cujo principal interesse reside no modo como potencialmente influenciam a ação dos diferentes fatores. O facto de ambas estarem associadas a efeitos principais significativos (respetivamente, $p < .001$ e $p = .01$) apoia a ideia de que se trata de variáveis intrinsecamente diferentes, contrariamente a indicações anteriores de que os efeitos atribuídos à distância poderiam resultar duma confusão metodológica com os pontos de desaparecimento em desenhos não-ortogonais (De Sá Teixeira & Oliveira, 2011). Dado que a magnitude das disparidades AV variou como uma percentagem fixa (25%) da distância percorrida pelo alvo visual (no caso das disparidades espaciais) ou, equivalentemente, da duração do percurso (no caso das disparidades temporais), as interações tarefa × distância e *offset* × distância podem assinalar um efeito modulador da magnitude das disparidades sobre a operação dos fatores tarefa e *offset*.

Estes resultados gerais são analisados mais detalhadamente nas secções seguintes, com recurso a análises gráficas e comparações envolvendo níveis mais específicos do desenho experimental.

5.1.2. Efeitos das disparidades audiovisuais relativas ao *offset*

A **Figura 2** apresenta o gráfico fatorial *onset* × *offset* (dados agregados através dos restantes fatores). O declive praticamente nulo das linhas indica a ausência de um efeito do *onset* (fator em abcissa), a separação vertical das linhas um efeito do *offset* (parâmetro das curvas) e o

paralelismo das linhas a ausência de interação entre os dois fatores. Estas indicações são estatisticamente suportadas pela ANOVA global (ver **Tabela 1**).

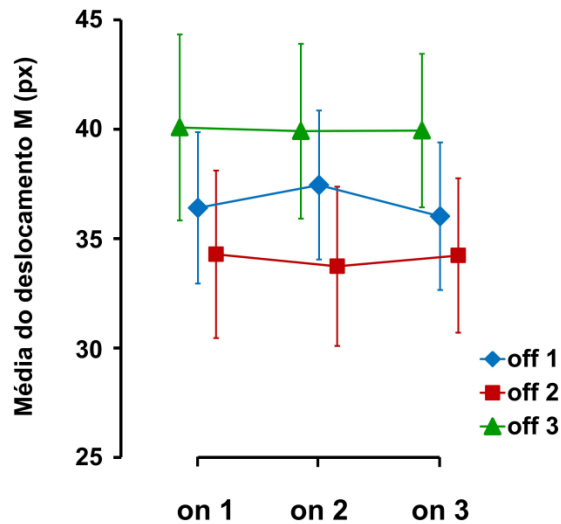


Figura 2. Gráfico fatorial *onset* × *offset* obtido com resultados agregados através dos restantes fatores. Na ordenada encontram-se as médias de deslocamento M. As barras de erro representam o erro padrão da média (EPM). Como os dados se encontram agregados através das tarefas (T, E, ET), *onset* e *offset* referem-se aqui indistintamente a disparidades temporais, espaciais ou espaço-temporais.

O resultado estatisticamente nulo da interação tarefa × *onset* × *offset* na **Tabela 1** ($p = .375$) sugere que este padrão fatorial se manteve estável através das tarefas. Os efeitos do *offset* organizam-se verticalmente, por ordem crescente de deslocamento M, do seguinte modo: off_2 (ausência de disparidades), off_1 (*offset* do som espacialmente e/ou temporalmente anterior ao *offset* visual) e off_3 (*offset* do som espacialmente e/ou temporalmente posterior ao *offset* visual). Comparações aos pares subsequentes revelaram uma diferença significativa entre off_2 e off_3 ($p < .001$), mas não entre off_1 e off_2 ($p = .09$) ou entre off_1 e off_3 ($p = .129$). O resultado não-significativo da interação tarefa × *offset* ($p = .662$; ver **Tabela 1**) indica adicionalmente que esta ordenação não é um artefacto da agregação por média e se mantém a mesma através das tarefas. Comparações aos pares subsequentes a ANOVAs conduzidas para cada tarefa (T, E, ET) revelaram o mesmo padrão estatístico obtido a partir da ANOVA global: diferenças significativas entre off_2 e off_3 (T: $p < .01$; E: $p = .05$, ET: $p = .01$) e resultados estatisticamente nulos entre off_1 e off_2 e entre off_1

e off_3 (menor p observado = .160, para off_1 versus off_2 na tarefa T; maior p observado = .933, para off_1 versus off_2 na tarefa ET).

5.1.3. Efeitos da magnitude das disparidades audiovisuais

A **Figura 3** ilustra a interação *offset* × distância assinalada como significativa na **Tabela 1** ($p < .001$). Esta interação envolve a localização relativa de off_1 face a off_2 e off_3 na distância percorrida mais curta (d1) e na distância mais longa (d2). As disparidades AV são proporcionalmente mais pequenas (metade do tamanho) em d1 (± 250 ms/ $\pm 1.4^\circ$ a.v.) do que em d2 (± 500 ms/ $\pm 2.7^\circ$ a.v.). Assim, com disparidades AV mais pequenas, off_1 produz o mesmo deslocamento M ($M=37.3$, ± 3.3 EPM) que off_2 ($M = 37.7$, ± 3.3 EPM), e ambos se distinguem de off_3 ($M = 44.9$, ± 3.8 EPM). Com disparidades AV maiores, por sua vez, off_1 ($M = 35.9$, ± 3.6 EPM) apresenta valores de deslocamento M semelhantes a off_3 ($M = 35.1$, ± 4.0 EPM), e ambos se distinguem de off_2 ($M = 30.5$, ± 3.9 EPM). Isto foi estatisticamente suportado por comparações aos pares que revelaram: em d1 – um resultado nulo entre off_1 e off_2 ($p = 1.0$) e diferenças significativas entre off_3 e off_1 ($p < .001$) e off_3 e off_2 ($p < .01$); em d2 – um resultado nulo entre off_1 e off_3 ($p = 1.0$) e diferenças significativas entre off_2 e off_1 ($p < .001$) e off_2 e off_3 ($p < .01$).

Os dois níveis de distância estão também relacionados com diferentes durações do som, maiores em d2. As diferenças de duração total, e não tanto o tamanho das disparidades AV, poderiam ser assim a causa do efeito da distância. Na realidade, essa possibilidade é excluída pela ausência de efeitos do fator *onset* (sendo que a variação do *onset* implica uma variação da duração do som). Esta interação pode assim interpretar-se como uma modulação do impacto das disparidades AV no *offset* pela sua magnitude.

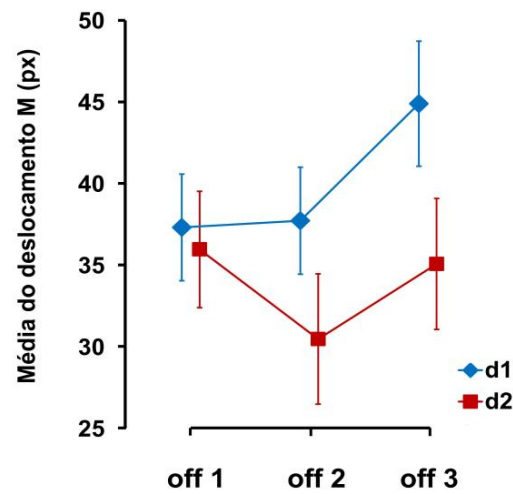


Figura 3. Gráfico fatorial *offset* × distância (d1, d2) (dados agregados através dos restantes fatores). As médias de deslocamento M encontram-se em ordenada. As barras de erro representam o erro padrão da média (EPM).

Um outro resultado significativo reportado na **Tabela 1**, relativo à interação de três vias tarefa × *offset* × distância ($p < .05$), assinala por sua vez que essa modulação foi dependente da tarefa. Na **Figura 4** o padrão fatorial *offset* × distância é ilustrado graficamente para cada uma das tarefas. É possível ver que a interação permanece nas tarefas T e ET, envolvendo disparidades temporais, mas desaparece (dando lugar a um padrão de paralelismo) na tarefa E, que envolve unicamente disparidades espaciais. Os resultados de ANOVAs de medidas repetidas conduzidas separadamente em cada tarefa foram concordantes com estas indicações da inspeção visual. A interação *offset* × distância foi significativa nas tarefas T, $F(2,60) = 8.280$, $p = .001$, $\eta_p^2 = 0.216$, e ET, $F(2,60) = 6.057$, $p < .01$, $\eta_p^2 = .168$, mas não na tarefa E, $F(2,60) = .809$, $p = .460$, $\eta_p^2 = .026$. Comparações aos pares subsequentes nas tarefas T e ET revelaram essencialmente o mesmo padrão estatístico encontrado na análise agregada: em d1 – resultados nulos entre off_1 e off_2 (T: $p = 1.0$; ET: $p = .912$) e diferenças significativas entre off_3 e off_1 (T: $p < .001$; ET: $p < .05$); em d2 – resultados nulos entre off_1 e off_3 (T: $p = .515$; ET: $p = 1.0$) e diferenças significativas entre off_1 e off_2 (T: $p < .01$; ET: $p < .05$).

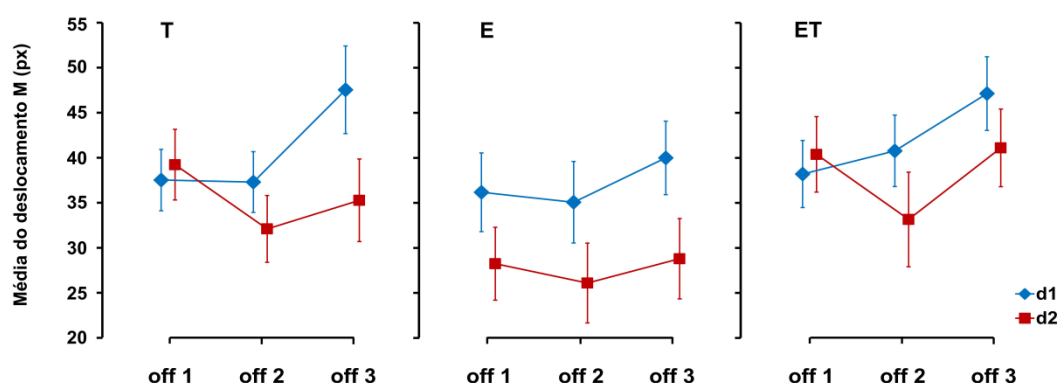


Figura 4. Gráficos fatoriais *offset* × distância em cada tarefa. T: tarefa temporal (apenas disparidades temporais). E: tarefa espacial (apenas disparidades espaciais). ET: tarefa espácio-temporal (disparidades temporais e espaciais covariadas). A média do deslocamento M encontra-se em ordenada. As barras de erro representam o erro padrão da média (EPM).

Tomados no seu conjunto, estes resultados indicam que é o tamanho das disparidades temporais no *offset* (assincronias), não o das disparidades espaciais (desalinhamentos espaciais), o atributo responsável pela interação *offset* × distância, ausente na tarefa E. Adicionalmente, sugerem a existência de uma janela temporal assimétrica no interior da qual assincronias AV de -250 ms (as mais pequenas, associadas a d1) são tratadas como equivalentes a *offsets* síncronos do som e do movimento visual.

Tal como a distância (d1, d2), o ponto de desapareção do alvo visual (p1, p2) apresentou um efeito significativo na ANOVA geral de 5 vias ($p = .002$; ver **Tabela 1**). No entanto, diferentemente da distância, não interagiu significativamente com o *offset* ($p = .392$; ver **Tabela 1**). O gráfico na **Figura 5** documenta esta ausência de interação, ilustrando um paralelismo quase perfeito das linhas relativas a p1 e a p2 (com maiores valores de deslocamento M em p2). Sendo o ponto de desapareção um fator ortogonal à distância e, por conseguinte, ao tamanho das disparidades (que variam com a distância), este resultado contribui para sublinhar, por contraste, o papel singular da magnitude das disparidades AV na modulação diferencial da localização percebida do alvo.

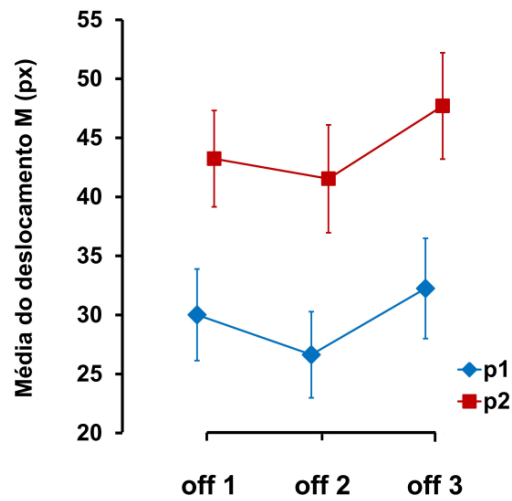


Figura 5. Gráfico fatorial *offset* × ponto de desaparecimento (p1, p2) (dados agregados através dos restantes fatores). As médias de deslocamento M encontram-se em ordenada. As barras de erro representam o erro padrão da média (EPM).

5.1.4. Efeitos da tarefa: Disparidades AV temporais, espaciais e espaço-temporais

A **Figura 6** apresenta graficamente o padrão fatorial *onset* × *offset* correspondente a cada tarefa (da esquerda para a direita: T, E e ET). Todos os gráficos exibem, embora com maior irregularidade, a tendência de paralelismo anteriormente observada na ANOVA geral (5 vias). ANOVAs de medidas repetidas conduzidas separadamente em cada tarefa revelaram em todos os casos um efeito nulo do *onset*, T: $F(1,644,49.307) = .890$, $p = .416$ ($\varepsilon = .822$), $\eta_p^2 = .029$; E: $F(2,60) = .320$, $p = .727$, $\eta_p^2 = .011$; ET: $F(2,60) = .360$, $p = .699$, $\eta_p^2 = .012$, um efeito principal significativo do *offset*, T: $F(2,60) = 5.413$, $p < .001$, $\eta_p^2 = .153$; E: $F(2,60) = 3.751$, $p = .029$, $\eta_p^2 = .111$; ET: $F(2,60) = 4.590$, $p = .014$, $\eta_p^2 = .133$. e uma interação *onset* × *offset* não significativa, T: $F(4,120) = .554$, $p = .696$, $\eta_p^2 = .018$; E: $F(4,120) = .949$, $p = .438$, $\eta_p^2 = .031$, ET: $F(4,120) = .940$, $p = .449$, $\eta_p^2 = .03$.

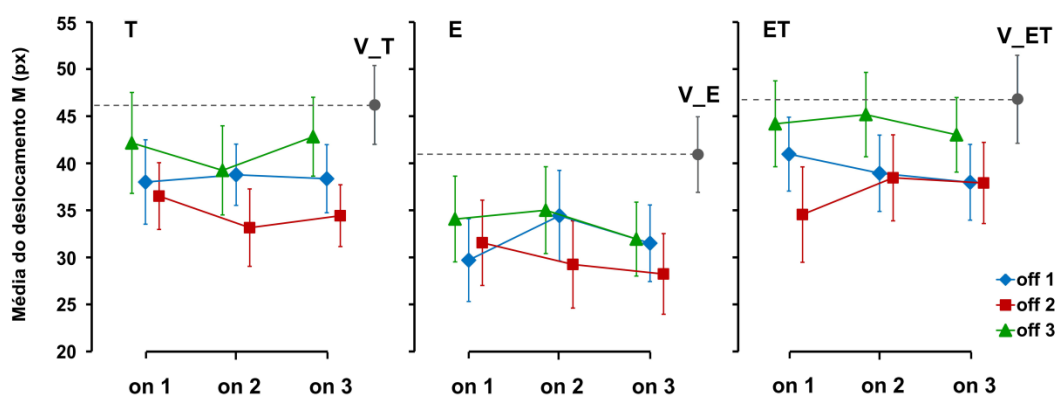


Figura 6. Gráficos fatoriais *onset* × *offset* em cada tarefa (T, E, ET). A linha a tracejado em cada gráfico corresponde ao deslocamento M obtido nas condições visuais intercaladas (V_T, V_E e V_ET). As médias de deslocamento M encontram-se em ordenada. As barras de erro representam o erro padrão da média (EPM).

Através das tarefas, os gráficos fatoriais localizam-se a diferentes alturas do eixo comum das ordenadas, com a tarefa E, ao centro, associada aos menores valores de deslocamento M, a tarefa ET, à direita, associada aos maiores valores e a tarefa T, à esquerda, ocupando uma posição intermédia. Estas diferenças entre tarefas correspondem na **Tabela 1** a um efeito principal significativo do fator tarefa ($p < .05$). Nenhuma das comparações aos pares subsequentes se mostrou significativa (menor p observado = .064, na comparação entre E e ET), mas o decréscimo relativo do deslocamento M na tarefa E (nível 2 do fator tarefa), foi assinalado por um contraste polinomial quadrático significativo, $F(1,30) = 4.984$, $p < .05$, $\eta_p^2 = .142$.

Considerando que todas as condições AV na tarefa E são síncronas, isto pode indicar que as diferenças entre tarefas dependem fundamentalmente da presença (nas tarefas T e ET) *versus* ausência (na tarefa E) de disparidades temporais AV no *offset*, com uma tendência para deslocamentos M mais reduzidos no último caso. Não é possível, contudo, com base no presente desenho, excluir que as disparidades espaciais na tarefa E tenham, em alternativa, sido responsáveis no todo ou em parte pela redução do erro. A tarefa ET, que inclui o mesmo padrão de disparidades espaciais (para além de disparidades temporais) apresenta os maiores valores de deslocamento M, sugerindo que, ao invés de operarem no sentido de reduzir o erro, as disparidades espaciais amplificam antes o efeito das disparidades temporais (aumento do erro relativamente às condições sem disparidade). Em qualquer caso, uma conclusão definitiva não é possível neste ponto.

A linha tracejada em cada um dos gráficos representa o deslocamento M observado nas condições visuais intercaladas. Estas condições mantiveram-se as mesmas através das tarefas, mas diferem na disposição em altura do mesmo modo que os contextos AV correspondentes (menores valores de deslocamento M em V_E , maiores valores em V_{ET} , V_T em posição intermédia). Isto sugere a possibilidade de um efeito contextual da tarefa nos ensaios visuais intercalados. No entanto, uma ANOVA de medidas repetidas de uma via, com as condições visuais intercaladas (V_T , V_E , V_{ET}) como níveis do único fator intra-sujeitos, produziu um resultado não-significativo, $F(1.417,41.093) = 1.804$ ($\epsilon = .0709$), $p = .185$, $\eta_p^2 = .059$.

A interação significativa tarefa \times distância na **Tabela 1** ($p < .05$) encontra-se graficamente ilustrada na **Figura 7**, revelando um decréscimo mais acentuado do deslocamento M na tarefa E relativamente às duas outras tarefas no nível de distância d2. ANOVAs de medidas repetidas de uma via realizadas separadamente para os níveis d1 e d2 (com a tarefa como fator intra-sujeitos), revelaram um resultado nulo em d1, $F(2,60) = 1.136$, $p = .264$, $\eta_p^2 = .043$, e um efeito significativo da tarefa em d2, $F(2,60) = 7.406$, $p = .001$, $\eta_p^2 = .198$. Comparações múltiplas aos pares documentaram uma diferença significativa entre as tarefas E e ET em d2 ($p < .01$), ficando a comparação entre as tarefas E e T, por sua vez, associada a um $p = .06$.

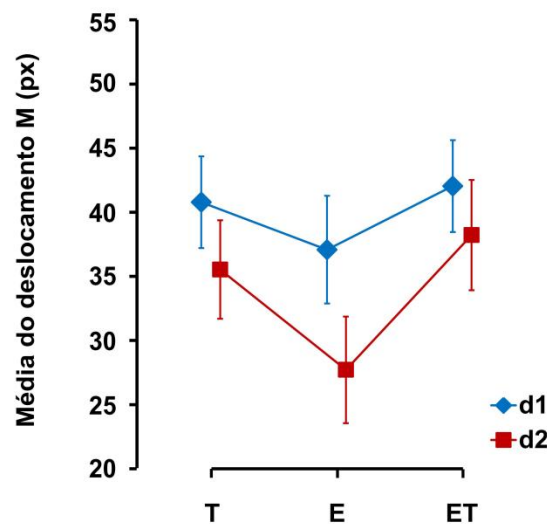


Figura 7. Gráfico fatorial tarefa (T, E, ET) \times distância (d1, d2) (dados agregados através dos restantes fatores). A média do deslocamento M encontra-se em ordenada. As barras de erro representam o erro padrão da média (EPM).

Dado que o nível d2 de distância está associado a disparidades AV de maior magnitude, estes resultados admitem duas interpretações alternativas. Ou as maiores disparidades temporais (d2) em T e ET produziram maiores erros de localização comparativamente às apresentações síncronas da tarefa E do que as pequenas disparidades temporais (d1); ou as maiores disparidades espaciais (d2) na tarefa E produziram uma maior redução do erro relativamente às tarefas com assincronias (T e ET) do que as pequenas disparidades espaciais (d1). A primeira interpretação surge como bastante mais plausível do que a segunda, apoiando a hipótese de um papel primário da informação temporal nas diferenças entre tarefas. Contudo, como assinalado antes, o presente desenho não permite conclusões definitivas quanto a este ponto.

5.1.5. Apresentações audiovisuais e visuais unimodais: Efeitos comparativos

Um traço comum a todos os gráficos na **Figura 6** é o de as linhas a cheio que representam condições AV se situarem abaixo das linhas a tracejado que representam as condições visuais (V) intercaladas. Esta característica não conhece exceções, sendo válida para todos e em cada um dos níveis de *onset* (fator em abcissa). Por conseguinte, independentemente da presença de disparidades, as condições AV conduziram invariavelmente a melhorias na execução da tarefa.

Esta redução do erro mostrou ser significativa quando avaliada por testes *t* emparelhados entre ensaios V e AV (agregados) em cada uma das tarefas – T: $t(30) = 4.509, p < .001$; E: $t(29) = 3.433, p = .002$; ET: $t(30) = 3.385, p = .002$. Um participante foi removido da comparação na tarefa E por apresentar um valor extremo (mais de três vezes superior ao intervalo interquartil) de deslocamento M nas apresentações visuais isoladas. Com vista à comparação das condições AV em cada um dos níveis de *offset* com as respetivas condições V intercaladas, foram ainda conduzidas ANOVAs de medidas repetidas de uma via em cada tarefa, com os níveis V, off_1, off_2 e off_3 como níveis do único fator intra-sujeitos, seguidas de comparações múltiplas aos pares. Em todas as tarefas, o deslocamento M em V foi significativamente maior do que em off_2 e off_1. T: V (M=46.19, DP=23.33) > off_2 (M=34.69, DP=18.57), $p < .001$, e V > off_1 (M=38.37, DP=19.57), $p = .014$. E: V (M=40.94, DP=22.08) > off_2 (M=29.679, DP=23.86), $p = .002$, e V > off_1 (M=31.88, DP=22.48), $p = .015$. ET: V (M=46.82, DP=26.19) > off_2 (M=36.97, DP=24.29), $p = .001$, e V > off_1 (M=39.29, DP=20.96), $p = .018$. Apesar de consistente em todos os gráficos, a redução do erro nas

condições AV off_3 em relação a V não teve expressão estatística em nenhuma das tarefas (menor p observado = .145).

A **Figura 8** fornece uma visão geral comparativa dos efeitos das apresentações AV e V intercaladas através das tarefas (identificadas pelas letras T, E, ET após *underscore*). Maiores deslocamentos M no bloco das condições V (colunas brancas) relativamente ao bloco das condições AV (colunas cinzentas) são aparentes. A diferença estatisticamente significativa entre os dois blocos (V e AV) assinalada no gráfico por um asterisco foi obtida numa MANOVA de medidas repetidas que utilizou o tipo de condição como uma variável de agrupamento com dois níveis (V e AV) e as tarefas (T, E, ET) como três medidas distintas do efeito (múltiplas variáveis dependentes): Traço de Pillai = .579, $F(3,27) = 12.355$, $p < .001$.

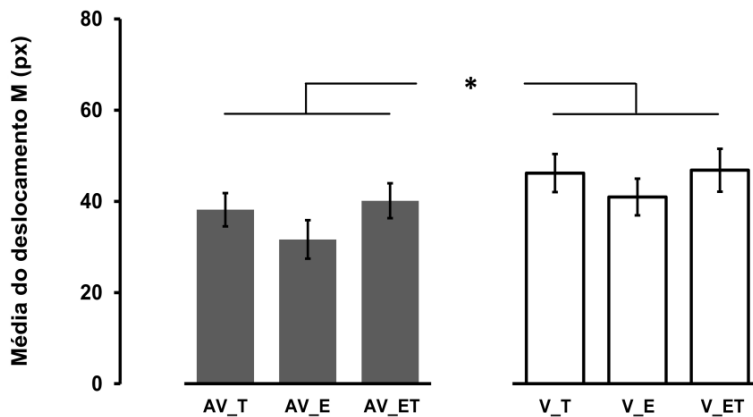


Figura 8. Efeitos comparativos das apresentações AV e V intercaladas. As colunas a cinzento representam o deslocamento M agregado através de todas as apresentações AV em cada uma das tarefas (AV_T, AV_E, AV_ET). As colunas a branco correspondem à média do deslocamento M nas condições V intercaladas em cada tarefa (V_T, V_E, V_ET). As barras de erro representam o erro padrão da média (EPM).

Para além das comparações entre médias, as condições V e AV foram igualmente comparadas quanto às variâncias através do teste de Pitman-Morgan para amostra correlacionadas. Exceto quando indicado diferentemente, os valores de p fornecidos são unilaterais. Essas comparações identificaram uma redução significativa da variância nas condições AV (relativamente às condições V intercaladas) nas tarefas T e ET, mas não na tarefa E. $V_T (s^2 = 544.14) > AV_T (s^2 = 413.77)$, $t(29) = 1.748$, $p = .045$; $V_ET (s^2 = 685.74) > AV_ET (s^2 = 451.78)$, $t(29) = 2.774$, $p = .005$; $V_E (s^2 = 487.49) \approx AV_E (s^2 = 532.04)$, $t(28) =$

.374, $p = .356$. Não se observaram diferenças significativas quanto à variância entre condições V (menor p bilateral observado = .264) ou entre condições AV (menor p bilateral observado = .382). Considerados conjuntamente, os resultados desta subsecção mostram que as condições AV bimodais não apenas resultaram em maior exatidão da localização (melhoria do desempenho), como (exceto na tarefa E) numa menor dispersão da localização (redução da variância) ao nível do desempenho de grupo.

A amostra de 10 participantes que, na experiência auxiliar, realizaram também uma tarefa de localização de alvos visuais unimodais fornece, apesar das limitações associadas a um N reduzido, a possibilidade de comparações adicionais. Diferentemente das condições visuais intercaladas, que podem ter sido afetadas pelo contexto AV específico (T, E, ET) em que foram inseridas, as condições visuais unimodais na tarefa auxiliar (V_is) encontravam-se desacompanhadas de qualquer contexto AV. Torna-se assim possível obter indicações sobre o efeito da bimodalidade (condições AV) relativamente a condições visuais independentes (V_is), bem como sobre eventuais efeitos do contexto AV nas apresentações visuais intercaladas, tomando como termo de comparação a localização percebida do alvo nas condições V_is. A **Figura 9** apresenta a média do deslocamento M (ordenada) para as diversas condições AV, V intercaladas, e V_is neste subgrupo de dez participantes.

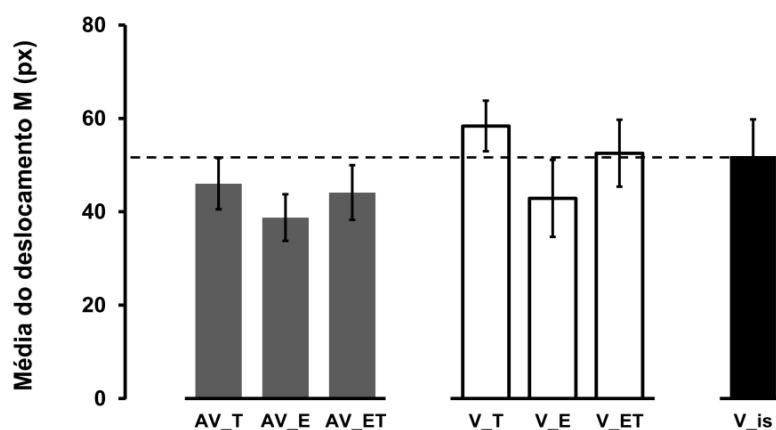


Figura 9. Médias de deslocamento M nas apresentações AV (colunas a cinzento), V intercaladas (colunas a branco) e V isoladas (coluna preta) fornecidas pelo subgrupo de 10 participantes que realizaram também a tarefa auxiliar com condições visuais unimodais independentes de qualquer contexto audiovisual (V_is). A linha tracejada de referência corresponde aos valores de V_is. As barras de erro representam o erro padrão da média (EPM).

O padrão de resultados relativos às condições AV (colunas a cinzento) e V intercaladas (colunas a branco) é essencialmente análogo ao ilustrado na **Figura 8** para o conjunto dos participantes, com maiores valores de deslocamento M no bloco das condições V intercaladas do que no das condições AV e, em cada um destes blocos, maiores reduções do deslocamento M na tarefa E (onde todas as condições AV são síncronas). Relativamente às condições V_{is} (coluna preta e linha tracejada), as apresentações AV continuam a mostrar valores de erro de localização inferiores. Esta diferença revelou-se significativa numa ANOVA de medidas repetidas de uma via com V_{is}, AV_T, AV_E e AV_{ET} como níveis do único fator intra-sujeitos, $F(3,24) = 3.113$, $p = .045$, $\eta_p^2 = .280$. Nenhuma das comparações aos pares subsequentes atingiu o critério de significância estatística (o que, considerando o reduzido efetivo amostral, pode simplesmente refletir uma inadequação do poder estatístico, acentuada pelo carácter conservador da correção de Bonferroni). A variância das condições AV foi inferior à das condições V_{is} ($s^2 = 557.64$) em todas as tarefas, sendo essa diferença estatisticamente significativa (teste Pitman–Morgan para amostra correlacionadas, p unilateral) nas tarefas T [$s^2 = 240.878$; $t(7) = 1.857$, $p = .05$] e ET [$s^2 = 216.248$; $t(7) = 2.182$, $p = .032$]. Estes resultados reforçam e estendem as conclusões anteriores relativas aos efeitos da bimodalidade na localização percebida do alvo visual. Com efeito, tal como nas comparações com as condições V intercaladas (analisadas atrás para o conjunto dos participantes), a comparação com as condições V_{is} neste subgrupo de participantes aponta para uma redução do erro de localização e da dispersão das respostas nas condições AV.

Por sua vez, a comparação entre V_{is} e as apresentações V intercaladas mostram valores de deslocamento M semelhantes em V_{is} e V_{ET}, inferiores em V_E e superiores em V_T. Estas diferenças não tiveram expressão estatística, $F(3,24) = 1.693$, $p = .195$, $\eta_p^2 = .175$ (ANOVA de medidas repetidas de uma via, com V_{is}, V_T, V_E e V_{ET} como níveis do único fator intra-sujeitos). O mesmo sucedeu com as comparações de variâncias entre V_{is} e as apresentações V intercaladas em cada tarefa (menor p observado = .402, entre V_{is} e V_T). Estes resultados limitam de alguma forma a hipótese de uma modulação contextual das condições V intercaladas pelos respetivos contextos AV. No entanto, o facto de se basearem num subgrupo reduzido de participantes (cerca de 1/3 do total), com implicações importantes na redução do poder estatístico, permite apenas atribuir-lhes um valor indicativo.

5.1.6. Um teste à localização percebida do termo (*offset*) do movimento de estímulos auditivos

A manipulação experimental dos pontos terminais do movimento de estímulos auditivos (através da espacialização do som nas tarefas E e ET) baseou-se na premissa de que a localização percebida desses pontos refletiria no essencial quer a sua disposição no espaço quer a sua posição relativamente ao *offset* do movimento visual. Caso contrário, a observação de que todas as condições AV nas tarefas E e ET produzem menos deslocamento M do que as condições V poderia resultar de um erro de localização auditiva, consistindo em perceber o *offset* espacial dos sons como ocorrendo sempre atrás do *offset* do movimento visual – mesmo quando ocorria adiante. Do mesmo modo, a ordem invariável dos efeitos de *offset* através das tarefas (erro crescente na ordem off_2, off_1, off_3), poderia hipoteticamente resultar, nas tarefas E e ET, da localização percebida de off_2 ser erradamente percebida como espacialmente anterior à de off_1.

O valor explicativo deste tipo de conjeturas está limitado à partida pelo facto de a tarefa T, onde os sons não foram espacializados, produzir o mesmo padrão de resultados que as restantes tarefas. Por outro lado, não se vislumbra um mecanismo óbvio capaz de explicar de que modo off_2, por exemplo, poderia ser erroneamente percebido, de forma sistemática, como situado espacialmente atrás de off_1. Em qualquer caso, a tarefa ancilar, na qual os participantes localizaram o ponto de desaparecimento de sons em movimento, facultou um teste direto a este tipo de suposições.

A **Figura 10** apresenta os deslocamentos M obtidos com movimentos de alvos auditivos unimodais (A) na tarefa auxiliar (barras brancas) e, para efeitos de comparação, os deslocamentos M em apresentações AV (barras cinzentas) e V intercaladas (barras pretas) na tarefa E. Os estímulos auditivos unimodais (A) consistiram nas componentes áudio das apresentações AV na tarefa E em que os movimentos visual e auditivo se iniciavam no mesmo local (nível 2 de *onset*). Para que as comparações fossem legítimas, por conseguinte, apenas esse subconjunto particular de apresentações AV foi considerado na **Figura 10**. O valor 0 (zero) na abcissa representa a posição verídica do *offset* do movimento visual (V_{off}) e a posição pretendida do *offset* 2 do movimento auditivo (A_{off_2}), que deveriam ser coincidentes. Os valores -2° e $+2^\circ$ correspondem às posições médias pretendidas do *offset* 1 (A_{off_1}) e do *offset* 3 (A_{off_3}) do movimento auditivo, respetivamente, em coordenadas angulares normalizadas ao *offset* do movimento visual. A utilização de valores angulares em

vez de pixels ($2^\circ \approx 105$ px) destina-se a facilitar a comparação com a magnitude do momento representacional auditivo reportada em trabalhos anteriores na literatura. Os valores de deslocamento M são representados pelas barras horizontais em termos de desvios relativamente à localização verídica/pretendida.

Dois participantes foram removidos por apresentarem deslocamentos M superiores a uma vez e meia a amplitude do intervalo interquartil em pelo menos um dos níveis de *offset* auditivo. O gráfico mostra que o movimento auditivo produziu sempre um erro de localização para diante, significativamente diferente de zero (0) em A_off_1 ($M = 0.82^\circ$, $\pm 0.39^\circ$ EPM), $t(28) = 2.075$, $p < .05$, e A_off_3 ($M = 1.36^\circ$, $\pm 0.32^\circ$ EPM), $t(28) = 4.042$, $p < .001$, mas não em A_off_2 ($M = 0.6^\circ$, $\pm 0.32^\circ$ EPM), $t(28) = 1.904$, $p < .067$. Os tamanhos médios do erro foram um pouco inferiores aos reportados por Getzmann et al. (2004), que variaram entre 2° ($\pm 1.1^\circ$ EPM) a 4° ($\pm 1.2^\circ$ EPM) nos movimentos da esquerda para a direita, mas de magnitude comparável aos observados em Getzmann & Lewald (2009), em torno de 1.5° ($\pm 0.5^\circ$ EPM), e em Feinkohl et al. (2014), de 0.88° e 1.26° (respetivamente, para velocidades de $25^\circ/s$ e $50^\circ/s$). Entre estes diferentes estudos existem diferenças importantes, relativas, por exemplo, ao método de resposta, modo de apresentação do som, velocidades do alvo, pontos de desaparecimento, duração dos trajetos. Apesar disso, o facto de termos obtido resultados não muito distintos dos observados em estudos em campo livre (como os de Getzmann), ou com recurso a HRTF individualizadas para apresentação em auscultadores (como o de Feinkohl et al.), constitui uma indicação favorável à qualidade da espacialização do som.

Não se registaram diferenças significativas de deslocamento M entre os *offsets* auditivos (A_off_1 , A_off_2 , A_off_3), $F(2,56) = 2.803$, $p = .069$, $\eta_p^2 = .091$, nem entre qualquer das condições A e as apresentações V intercaladas (V_E: $M = 0.9^\circ$; $\pm 0.12^\circ$ EPM), $F(3,84) = 1.905$, $p = .135$, $\eta_p^2 = .064$. Este resultado converge com a observação de Getzmann & Lewald (2009) de valores de MR semelhantes nas modalidades visual e auditiva (no entanto, para um resultado divergente, ver Schmiedchen et al., 2013). Os valores de EPM, representados pelas barras de erro, são claramente maiores nas condições A do que nas condições V_E, e atingem o seu mínimo nas condições AV. Comparações com o teste de Pitamn-Morgan revelaram diferenças significativas de variância entre todas as condições A_off e as apresentações V_E ($p < .001$), bem como entre as primeiras e as apresentações AV correspondentes na tarefa E ($p < .001$). Apesar de todas as condições AV terem variâncias inferiores à de V_E, não foram neste caso documentadas diferenças significativas (menor p observado = .197, associado à comparação entre V_E e AV_off_1).

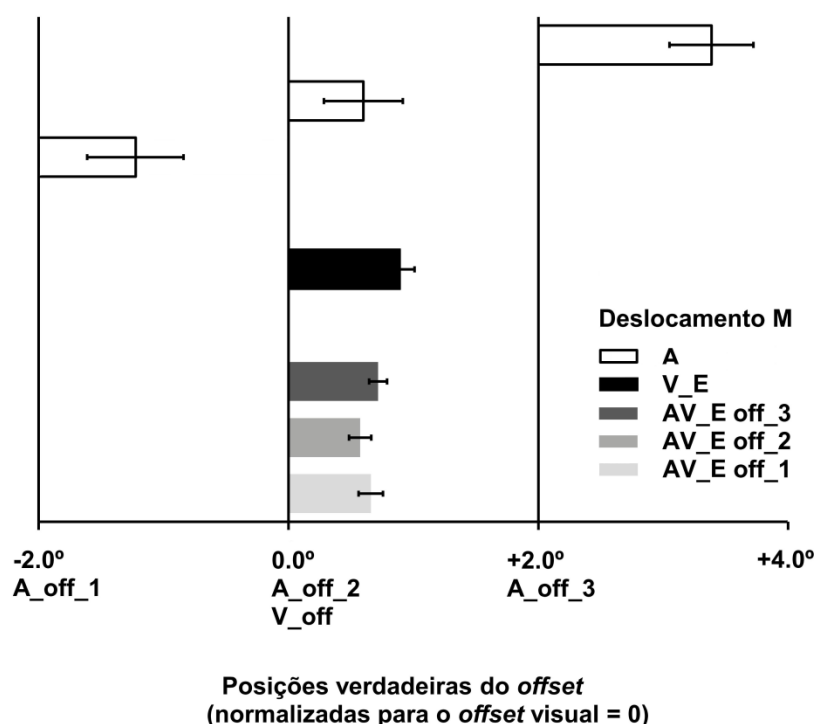


Figura 10. Deslocamento M obtido com movimentos auditivos (A) na tarefa auxiliar, no subconjunto de ensaios AV da tarefa E que utilizou componentes áudio correspondentes a esses movimentos (barras cinzentas), e com as condições visuais intercaladas da tarefa E (barras pretas). O valor 0.0° na abcissa representa a posição verdadeira do *offset* visual (V_off) e a posição pretendida do *offset* 2 (A_off_2), que deveriam ser coincidentes. Os valores - 2° e + 2° representam as posições médias pretendidas do *offset* 1 (A_off_1) e do *offset* 3 (A_off_3) do movimento auditivo, respetivamente, em coordenadas angulares normalizadas ao *offset* do movimento visual. As médias de deslocamento M são dadas pelas barras sob a forma de desvios relativamente à posição verdadeira/preendida. As barras de erro representam o erro padrão da média (EPM).

De modo fundamental para os nossos propósitos, A_off_1 foi percebido como estando localizado atrás (especialmente antes, ou mais à esquerda) tanto de A_off_2 como de V_off. Simetricamente, A_off_3 foi localizado pelos participantes adiante (especialmente à frente, ou mais à direita) tanto de A_off_2 como de V_off. Em concordância com a coincidência espacial pretendida entre ambos, A_off_2 foi efetivamente localizado na vizinhança de V_off (apenas 0.3° à esquerda). Uma ANOVA de medidas repetidas de uma via realizada sobre as coordenadas físicas indicadas pelos participantes, com AV_off_1, AV_off_2, AV_off_3 e V_off como níveis do único fator intra-sujeitos, produziu um resultado significativo, $F(3,84) = 66.327$, $p < .001$, $\eta_p^2 = .793$. Comparações aos pares subsequentes revelaram diferenças significativas entre todos os níveis de *offset* de A ($p < .001$), bem como entre V_E versus

A_off_1 ($p < .001$) e A_off_3 ($p < .001$). A comparação entre V_E e A_off_2, por sua vez, foi estatisticamente nula ($p = 1.00$). A localização relativa dos *offsets* espaciais auditivos, tanto uns em relação aos outros como em relação ao *offset* do movimento visual, foi assim corretamente percebida.

Adicionalmente, não parece existir qualquer forma consistente de derivar a localização percebida dos *offsets* AV (barras cinzentas) como uma combinação direta da localização percebida dos *offsets* auditivos (barras brancas) e do *offset* visual. Isto exclui a atração para a localização percebida do som (viés de atração) como uma explicação viável dos resultados obtidos nas tarefas E e ET. Invalida ainda o modelo Bayesiano da máxima verosimilhança (MMV: Alais & Burr, 2004; Ernst & Banks, 2002) como um quadro de interpretação plausível. De acordo com este modelo, a localização nas condições AV deveria poder derivar-se como uma média ponderada das informações percebidas auditiva e visual. A impossibilidade de o fazer aponta para a inaplicabilidade do MMV à modulação AV do deslocamento M no nosso estudo.

O recurso à amostra de 10 participantes que realizaram adicionalmente uma experiência de localização de alvos visuais (V_is) no quadro da tarefa auxiliar permite ainda comparar o deslocamento M entre alvos auditivos e alvos visuais independentes de qualquer contexto audiovisual (não intercalados entre apresentações AV). A **Figura 11** (A e B) representa o deslocamento M associado a alvos auditivos (losangos) e visuais (triângulos) como um desvio (lido no eixo das ordenadas) à diagonal traçada em cada um dos gráficos. A linha diagonal representa uma relação linear perfeita entre a posição física dos *offsets* (correspondente a uma coordenada x no ecrã, em abcissa) e a sua localização percebida, correspondente à coordenada x da resposta dos participantes (em ordenada). Desvios para cima relativamente à diagonal indicam erros de localização para diante; desvios para baixo indicam erros de localização para trás. O sistema de coordenadas tem origem (convencional) no centro do ecrã, pelo que as coordenadas à esquerda têm um sinal negativo e as coordenadas à direita um sinal positivo. O gráfico A (em cima) ilustra os resultados obtidos com a distância percorrida d_1 (mais curta) e o gráfico B (em baixo) os resultados na distância percorrida d_2 (maior). Em ambos os gráficos, os dois círculos a preto identificam as respostas ideais (refletindo uma perceção exata) de localização do *offset* visual em cada um dos dois pontos de desaparecimento (p_1 e p_2). No conjunto, os gráficos A e B permitem examinar a posição relativa percebida dos *offsets* auditivos (A_off1, A_off_2, A_off_3) entre si e relativamente ao *offset* visual (V_off) percebido em cada uma das distâncias (d_1 , d_2) e pontos de desaparecimento (p_1 , p_2).

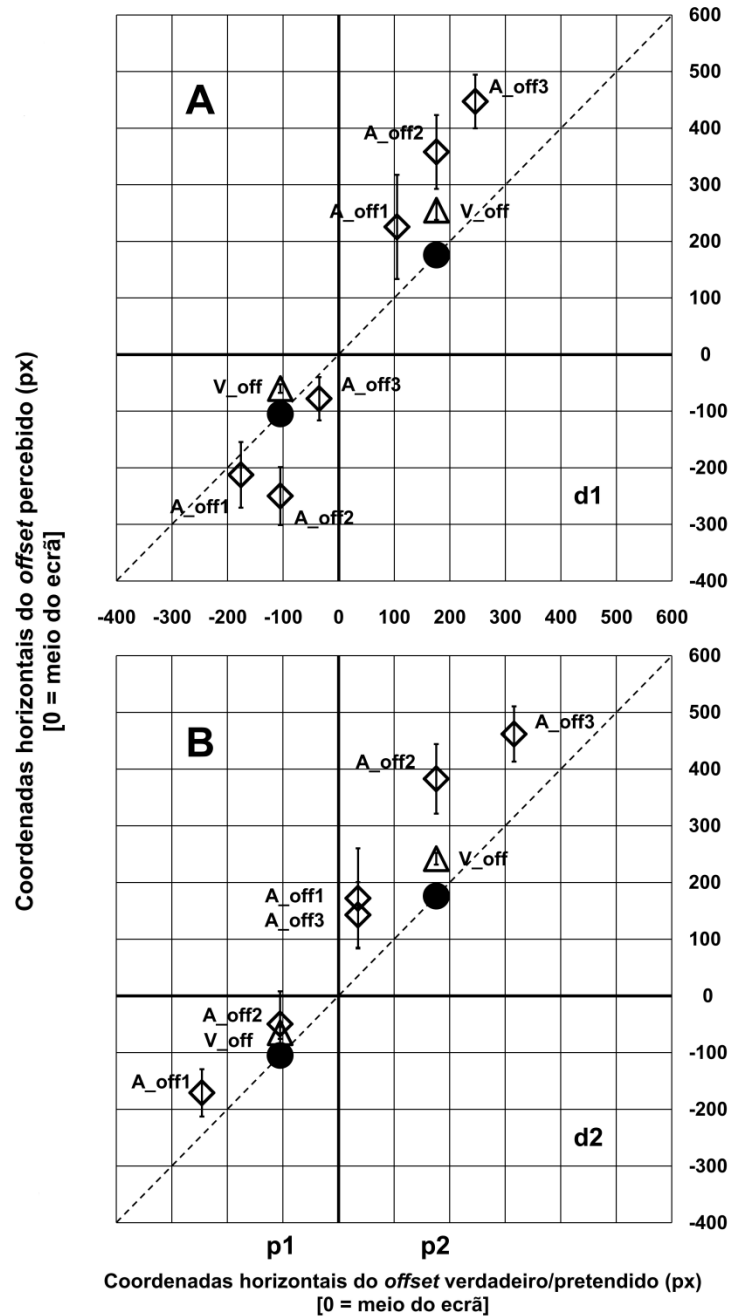


Figura 11. Respostas de localização (posições percebidas) do *offset* de alvos dinâmicos auditivos (A_off_1, A_off_2, A_off_3), representadas por losangos, e de alvos dinâmicos visuais apresentados à margem de qualquer contexto audiovisual (V_is, no gráfico V_off), representados por triângulos, em função da sua posição verdadeira (no caso do alvo visual) e pretendida (no caso dos sons espacializados). A posição verdadeira/preterida, em abcissa, é dada pela coordenada x do ecrã correspondente ao *offset*

do alvo. A posição percebida (na ordenada) é dada pela coordenada x da resposta de localização dos participantes. O sistema de coordenadas tem origem convencional no centro do ecrã, com valores negativos à esquerda e positivos à direita. Os dois círculos a preto em cada gráfico representam as respostas ideais (refletindo uma percepção exata) de localização do *offset* visual em cada um dos dois pontos de desaparecimento do alvo (p_1 e p_2). Os gráficos A (em cima) e B (em baixo) ilustram, respetivamente, os resultados obtidos com trajetos correspondentes aos níveis de distância d_1 (trajetos mais curtos) e d_2 (trajetos mais longos). A diagonal tracejada em cada gráfico representa uma correspondência linear perfeita entre as coordenadas reais/pretendidas do *offset* e as coordenadas percebidas. O deslocamento M é dado pelos desvios (lidos no eixo das ordenadas) em relação à diagonal.

Com uma única exceção no ponto 1 em d_1 (gráfico A, quadrante inferior esquerdo), onde se observa uma inversão da posição relativa de A_{off_1} e A_{off_2} , os resultados mostram, como antes, uma correspondência geral entre a disposição espacial pretendida dos *offsets* do som e a sua disposição relativa percebida. A_{off_1} foi percebido em todos os quadrantes como atrás do *offset* visual (V_{off}) e, com exceção, de novo, do quadrante inferior esquerdo do gráfico A (d_1, p_1), A_{off_3} foi percebido como à frente de V_{off} . Por conseguinte, tal como relativamente às condições visuais intercaladas, a posição dos *offsets* auditivos relativamente ao *offset* das apresentações visuais isoladas (V_{is}) foi corretamente percebida. Quanto a A_{off_2} , a sua localização pelos participantes relativamente a V_{off} variou com a distância e o ponto, apresentando em todo o caso uma tendência dominante (em três dos quatro quadrantes) para ser percebido à frente de V_{off} . Este resultado é indicativo, uma vez mais, da incapacidade do modelo da máxima verosimilhança (MMV) para explicar a redução de deslocamento M nas condições AV da tarefa E (em particular no nível 2 de *offset*) como uma média ponderada da informação espacial auditiva e visual. Considerando o reduzido número de participantes na amostra ($N=10$), estas observações devem tomar-se acima de tudo como sugestivas e complementares. Ainda assim, reforçam a coerência geral dos resultados analisados na subsecção e, desse modo, o sentido das conclusões apontadas.

5.2. DISCUSSÃO

5.2.1. Conclusões gerais

Os resultados relativos ao deslocamento M suportam de forma robusta duas conclusões gerais. Em primeiro lugar, de modo consistente em todas as tarefas, as disparidades AV no *onset* não tiveram efeitos na localização percebida do ponto de desaparecimento do alvo, enquanto as disparidades AV no *offset* tiveram efeitos significativos. Isto contrasta com a conclusão de Teramoto, Hidaka, et al. (2010) de que a sincronia AV no *onset* é uma condição necessária para que o som module o MR visual (não havendo modulação no caso de assincronia). É possível que os resultados de Teramoto, Hidaka, et al. (2010) permaneçam válidos no domínio das pequenas assincronias AV (< 160 ms) e dos efeitos de “atração para o som” que estes autores reportaram. No entanto, efeitos de atração semelhantes foram também obtidos por Chien et al. (2013), com valores de assincronia comparáveis, utilizando sons transientes na vizinhança temporal do *offset* do movimento. Estas discrepâncias entre estudos poderão dever-se a diferenças de procedimento, designadamente o uso no presente estudo, como no de Chien et al. (2013), de uma resposta motora de localização, em alternativa ao método de sonda perceptiva empregue por Teramoto, Hidaka, et al. (2010). Pode, por exemplo, admitir-se que a mobilização pela resposta de processos no ramo sensoriomotor do sistema visual (Milner & Goodale, 1995), potenciada ainda, no nosso estudo, pela liberdade de movimentos oculares de perseguição do alvo (Cui, Razavi, O’Neill, & Paige, 2010; Jay & Sparks, 1984; Kubovy & Van Valkenburg, 2001; Zmigrod & Hommel, 2009), tenha fornecido formas alternativas de “unificação” audiovisual, dispensando a sincronia no *onset*. De forma consistente com os nossos próprios resultados, Teramoto, Hidaka, et al. (2010) reportaram um efeito estatisticamente nulo da duração dos sons (que se estendeu até aos 1307 ms). Isso torna implausível que diferenças na duração dos sons entre o nosso estudo e o de Teramoto, Hidaka, et al. (2010) possam explicar a divergência relativamente ao papel da sincronia no *onset*. Em qualquer caso, o nosso resultado de um efeito nulo do *onset* impõe limites adicionais à ideia de que assincronias AV no *onset* do movimento sejam um impedimento, em qualquer circunstância, a um impacto do som no MR visual.

A ausência de efeitos do fator *onset* na tarefa E é metodologicamente importante. As disparidades espaciais encontravam-se confundidas nesta tarefa com as distâncias

percorridas pelos sons espacializados e com a sua velocidade, que variavam ambas com o *onset* espacial do som. Deste modo, o resultado nulo do *onset* exclui uma influência de qualquer dessas variáveis (distância percorrida e velocidade), permitindo atribuir os efeitos observados, especificamente, às disparidades espaciais no *offset*. Este resultado ilustra pela primeira vez um efeito direto (não mediado temporalmente) da informação espacial do som no MR visual. Mostra ainda que correlacionar as disparidades AV espaciais com a velocidade do movimento auditivo e não com a duração do movimento (como em Schmiedchen et al., 2012) fornece um meio efetivo de dissociar pistas auditivas temporais e espaciais por meio do desenho experimental.

A segunda conclusão refere-se ao tipo de efeito modulador dos sons. Embora em graus variáveis, em todas as condições AV se observaram efeitos de melhoria audiovisual (redução do erro de localização, comparativamente às condições visuais unimodais). De modo significativo, a maior redução do erro para diante em todas as tarefas ocorreu nas condições sem disparidades, não quando o *offset* do som foi temporalmente ou espacialmente anterior ao *offset* visual. Este último deveria, no entanto, ter sido o caso se o *offset* percebido do movimento visual fosse atraído pelo *offset* do som. Por fim, exceto na tarefa E, as condições AV apresentaram reduções significativas na variância comparativamente às condições V unimodais. O desempenho da localização ao nível do grupo foi assim não apenas mais exato (melhor localização), como mais concentrado em torno da resposta média de localização. Este padrão de efeitos pode legitimamente qualificar-se como melhoramento bimodal, contrastando com os enviesamentos de atração (para o som) reportados em todos os estudos de comparação. Deve assim notar-se que, sendo o objetivo da tarefa atribuída aos participantes em todos os estudos localizar o mais exatamente possível o *offset* do alvo visual, apenas no nosso estudo a informação auditiva se revelou instrumental para o propósito (i.e., foi funcionalmente relevante-para-o-objetivo). Este resultado é concordante com a expectativa inicial de que utilizar contingências AV perceptivamente mais salientes e em menor número poderia promover a relevância da informação auditiva para a tarefa e alterar qualitativamente o impacto do som na localização do *offset* visual, induzindo efeitos de benefício AV (ver secção 3.3.3 e 4.1).

As diferenças de deslocamento M entre condições visuais intercaladas (V_T, V_E, V_ET) nas diferentes tarefas apresentaram um perfil ordinal idêntico ao das diferenças entre condições AV (ver **Figura 8**), sugerindo alguma forma de modulação das condições V unimodais pelo contexto audiovisual. Isso poderia ter lugar através de alterações do critério de resposta, efeitos de plasticidade perceptiva, ou uma combinação de ambos os fatores.

Melhorias no processamento visual unimodal após exposição repetida a contingências AV foram antes documentadas (Shams et al., 2011) e sugeridas como de natureza perceptiva, mais do que cognitiva (R. Kim, Seitz, & Shams, 2008). No entanto, as diferenças entre condições V intercaladas não atingiram a significância estatística e a existência duma modulação AV contextual deve apenas tomar-se como uma sugestão provisória, dependente de futuras replicações.

5.2.2. Evidência de mecanismos guiados pelos dados (*bottom-up*)

Subjacente à expectativa anterior encontrava-se também a hipótese de que a manipulação do tamanho e número de disparidades AV incluídas no desenho poderia mobilizar processos multissensoriais distintos dos que suportam os enviesamentos de atração. Teramoto, Hidaka, et al. (2010) e Chien et al. (2013) interpretaram ambos os seus resultados em termos de integração multissensorial – unificação dos fluxos auditivos e visual num percepto AV unificado. Este percepto AV contendo o contributo do *offset* temporal do som seria o responsável pelos “enviesamentos de atração” verificados. Schmiedchen et al. (2012) recusam a integração sensorial em favor de um mecanismo pré-perceptivo (interações intermodais dispensando a “unificação” perceptiva), mas que, tal como o mecanismo de integração convocado por Teramoto, Hidaka, et al. (2010) e Chien et al. (2013), obedecem a janelas estreitas de tolerância à assincronia, próximas dos 100 ms. A predominância de janelas temporais estritas (“regra temporal”) é tipicamente encarada como indicativa de processos multissensoriais guiados pelos dados (*bottom-up*). Chien et al. (2013) relacionam neste sentido os seus resultados e a integração sensorial estudada em neurónios do córtex superior (Meredith & Stein, 1983). Esta leitura ganha um sentido substantivo adicional no quadro destes estudos, dado que os enviesamentos de localização operaram sem consideração pelo propósito da tarefa (localizar o alvo o mais exatamente possível), mostrando-se assim impermeáveis aos objetivos.

Na realidade, o presente estudo forneceu também evidência consistente com o envolvimento parcial de uma janela de tolerância à assincronia AV, especificamente a equivalência dos efeitos de facilitação AV obtidos com apresentações síncronas e com assincronias AV de -250 ms (*offset* do som temporalmente anterior ao *offset* visual na distância d1). Esta equivalência foi replicada nas tarefas T e ET, onde estas condições forneceram os maiores benefícios AV (em d1), de magnitude equivalente, e não teve

contrapartida nas disparidades unicamente espaciais (tarefa E). A assimetria da equivalência relativamente aos 0 ms (deixa de fora assincronias de + 250 ms) distingue-se da assimetria típica das janelas de integração audiovisual, geralmente mais estreitas quando o som é o elemento precedente na assincronia (Conrey & Pisoni, 2006; Dixon & Spitz, 1980; Schmiedchen et al., 2012; van Wassenhove et al., 2007), mas mantém-se compatível com indicações recentes de que diferentes mecanismos podem estar envolvidos na tolerância às assincronias, dependendo de qual a modalidade que precede (Cecere et al., 2016). Targher et al. (2017) reportaram uma equivalência similar envolvendo estes mesmos dois valores de assincronia (0 ms e -250 ms), interpretando-a como indicativa duma janela temporal de integração audiovisual. De modo semelhante, tomamos aqui a equivalência observada como um indicador plausível do envolvimento de processos multissensoriais guiados pelos dados (*bottom-up*) e regidos por uma regra temporal nas condições AV síncronas e com assincronias de -250 ms.

5.2.3. Mecanismos explicativos: Proposta de um quadro interpretativo do conjunto dos resultados

Face ao conjunto dos resultados, são dois os pontos essenciais a explicar. Por que é que as condições AV síncronas produziram os maiores benefícios de desempenho no presente estudo e nenhuma melhoria de localização nos estudos de comparação? E de que modo assincronias AV excedendo largamente a janela temporal de ± 100 ms estabelecida como condição de efetividade nos estudos de comparação lograram ainda assim produzir efeitos de facilitação no presente estudo? Sugerimos como uma explicação plausível, válida para ambas as questões, a intervenção no nosso estudo de processos atencionais endógenos, i.e., guiados por objetivos (Donohue et al., 2015; Koelewijn et al., 2010; Talsma, Senkowski, & Woldorff, 2009; Tang et al., 2016; van Ee et al., 2009). Nesta proposta explicativa, detalhada em seguida, a informação temporal relevante veiculada pelo som deve ser endogenamente selecionada e processada de modo a repercutir-se favoravelmente (e.g., através do afinamento do *timing* de desaparecimento do alvo) no processamento visual da localização espacial (Driver & Spence, 2000; Talsma et al., 2010; Tang et al., 2016). A relevância percebida para a tarefa das disparidades AV, promovida no nosso estudo pela sua saliência perceptiva (maior magnitude) e fiabilidade (menor número), forneceria a base indispensável para a operação da atenção endógena, não disponível nos estudos de comparação.

O impacto de uma propriedade (e.g., informação sobre o *timing*) numa outra propriedade (e.g., informação sobre o local) através de modalidades sensoriais distintas (e.g., audição e visão) (*cross-sensory cascading*; Driver & Spence, 2000) tem sido concebido como um análogo multissensorial da difusão da atenção baseada-no-objeto (Donohue, Roberts, Grent-'t-Jong, & Woldorff, 2011; Driver & Spence, 2000; Fiebelkorn et al., 2010; Molholm et al., 2007; Tang et al., 2016; Zimmer, Itthipanyanan, Grent-'t-Jong, & Woldorff, 2010; Zimmer, Roberts, Harshbarger, & Woldorff, 2010). Atribuímos, nesse sentido, um papel relevante na explicação proposta à distinção entre difusão de atenção guiada-pelos-dados e guiada-por-representações (Fiebelkorn et al., 2010; Molholm et al., 2007). A primeira assenta em objetos perceptivos multimodais resultantes da integração multissensorial *bottom-up*, que admitimos terem ocorrido nas condições síncronas e com assíncronas AV de -250 ms. Isto poderia resultar não apenas numa “objetualidade” (*objecthood*) mais acentuada da percepção intermodal nessas condições (Shomstein & Behrmann, 2008), como numa captura exógena da atenção pelos perceptos AV emergentes (Koelewijn et al., 2010; Talsma et al., 2010; Tang et al., 2016), fornecendo rapidamente uma base para a difusão subsequente de atenção endógena. Ambos os fatores poderiam, deste modo, ter contribuído para a obtenção dos maiores benefícios AV nestas condições. Para que a atenção endógena pudesse operar a partir daqui, no entanto, teria que ser capaz de aceder seletivamente às informações auditivas relevantes para a tarefa no interior do percepto integrado (AV). Isso não parece ser uma condição especialmente problemática, dada a larga prevalência empírica da chamada “fusão intermodal fraca” (*intermodal weak fusion*) (Clark & Yuille, 1990; Ernst & Bühlhoff, 2004; Hillis et al., 2002; Landy, Maloney, Johnston, & Young, 1995), em que a informação unimodal se mantém acessível (não é perdida ou dissolvida) nos objetos multimodais. Em contraste, nos estudos de comparação a atenção exógena não seria capaz, por si só, de contrariar o enviesamento de atração induzido pelos perceptos AV, podendo mesmo acentuá-los.

A difusão da atenção (endógena) guiada-por-representações operaria, por sua vez, nas restantes condições de assincronia AV (+ 250 ms, ± 500 ms). Embora concebida por Fiebelkorn et al. (2010) como baseada, sobretudo, em associações multissensoriais de longo prazo entre propriedades de objetos conhecidos, diferentes linhas de evidência indicam que a aprendizagem de correspondências intermodais durante uma sessão experimental é também capaz de originar representações coerentes (Heron, Roach, Hanson, McGraw, & Whitaker, 2012; Hidaka et al., 2015; R. Kim, Seitz, Feenstra, & Shams, 2009; Lewkowicz, 2003; Parise & Ernst, 2016; Parise et al., 2012; Seitz, Kim, van Wassenhove, & Shams, 2007;

Shams et al., 2011). O traço distintivo é que as representações não são já guiadas primariamente pela proximidade temporal de coocorrência (implicando janelas temporais específicas), mas pela detecção-aprendizagem de padrões estruturados de contingências multissensoriais no ambiente. Isto explicaria o facto de que, mesmo com assincronias mais longas (e.g., ± 500 ms), os sons lograram ainda produzir efeitos de melhoria AV no nosso estudo, e que esses efeitos tenham podido ocorrer (embora com diferenças de grau) tanto quando o som precedia (pré-ativamente), como quando sucedia (retroativamente) ao *offset* visual.

As condições síncronas e com assincronias de -250 ms foram interpretadas como envolvendo uma contribuição de processos integração sensorial guiados-pelo-estímulo (*bottom-up*) por razões empíricas discutidas atrás. No entanto, uma assincronia de -250 ms situa-se apesar de tudo fora dos limites das janelas temporais de integração observadas nos estudos de comparação. Isto pode compreender-se como uma modulação contextual das margens da janela de integração. Esse tipo de modulação temporal encontra-se bem documentado, sabendo-se igualmente que depende de vários fatores (Colonius & Diederich, 2010; Spence & Frings, 2020; Vroomen & Keetels, 2010), entre os quais a atenção e as características da tarefa (Donohue et al., 2015; Mégevand et al., 2013). A atenção a contingências AV repetidas na tarefa poderia assim ter resultado numa reconsideração dos limites da janela de integração adequada no contexto da tarefa (Targher et al., 2017). Isso estaria de acordo, no geral, com a noção de “janela temporal ótima variável de integração audiovisual” proposta por Colonius & Diederich (2010).

Ainda relativamente a esta questão, os efeitos distintos das assincronias AV de -250 ms e -500 ms (com maior facilitação na primeira) parecem mais adequadamente interpretáveis pela amplitude de uma janela de integração (tratando os 0 ms e os -250 ms como equivalentes) do que por mecanismos alternativos conjeturáveis, como efeitos bifásicos de pista (e.g., inibição de retorno e facilitação reflexiva: Blurton, Greenlee, & Gondan, 2015) ou efeitos de orientação temporal (Rohenkohl, Coull, & Nobre, 2011). As situações associadas à ocorrência de inibição de retorno (respostas mais lentas a alvos apresentados em locais na periferia onde há mais de 200-300 ms esteve uma pista: Klein, 2000; Spence & Driver, 1998) são demasiado distintas da presente situação experimental para permitirem qualquer forma de transposição útil. Soma-se a isso que a equivalência observada entre as condições síncrona e -250 ms AV não é favorável a uma facilitação automática (reflexiva) associada ao intervalo pista-alvo de 250 ms. Como indicado por vários autores (e.g., L. Chen, Shi, & Müller, 2010; Spence, 2007), sons especialmente não-

informativos não parecem influenciar as respostas de localização através duma orientação automática da atenção para a localização do alvo visual, e sim melhorando o agrupamento perceptivo entre acontecimentos multissensoriais próximos no tempo. O quadro da orientação temporal (*temporal orienting*: Coull & Nobre, 1998), diferentemente da orientação espacial, parece assim mais adequado à presente situação, mas levaria neste caso (pistas temporais de -500 ms e de -250 ms com igual validade) a esperar resultados semelhantes ou superiores nas condições de -500 ms (Weinbach & Henik, 2012), diferentemente do observado.

5.2.4. Objeções e alternativas possíveis: Discussão argumentada

Várias objeções gerais ao quadro explicativo proposto podem basear-se em mecanismos não específicos. A melhoria AV no nosso estudo poderia ter resultado de as assincronias em que *offset* do som precede o do movimento visual agirem como sinais preparatórios inespecíficos (ver, e.g., Los & Van der Burg, 2013). Contudo, num paradigma de período preparatório variável, como é o caso no presente estudo, as vantagens medidas através do abaixamento do tempo de reação (TR) aumentam com a extensão do período de preparação até aos -600 ms (Los & Van der Burg, 2013) e mesmo -3 s (Niemi & Näätänen, 1981). Maiores benefícios seriam assim esperáveis com assincronias de -500 ms (considerado o intervalo de “alerta” ótimo por Posner & Boies, 1971) do que com assincronias de -250 ms ou com as condições AV síncronas, contrariamente ao observado. Junta-se a circunstância de que os *offsets* do som posteriores ao *offset* visual, nos quais não existe, por definição, período preparatório, induziram também, ainda que em menor grau, efeitos de melhoria AV. Tentar explicar este resultado como um efeito de calibração das respostas assente nos ensaios em que o *offset* auditivo precede o visual choca com o facto de que as condições visuais intercaladas não lograram o mesmo benefício. Por fim, os efeitos não-específicos de alerta reportados consistem frequentemente num aumento de rapidez à custa da exatidão (Fernandez-Duque & Posner, 1997; Weinbach & Henik, 2012), enquanto os ganhos de exatidão são aqui o resultado característico.

Poderia também argumentar-se que os efeitos de melhoria AV se poderiam explicar simplesmente por um enviesamento geral da resposta, consistindo em responder mais à esquerda sempre que um som é apresentado. Isto deixaria por explicar os efeitos específicos das disparidades AV no *offset*, ou a razão pela qual um enviesamento semelhante não surgiu

nos estudos de comparação. Adicionalmente, não é claro por que deveria um enviesamento para a esquerda ocorrer, em primeiro lugar. Se se tratasse de um mecanismo geral de correção (não ir tanto para a direita na direção do movimento), deveria igualmente aplicar-se às condições visuais intercaladas. Se se tratasse duma interrupção dos processos de extrapolação do movimento (por exemplo, a ultrapassagem oculomotora do alvo: Kerzel, 2000, 2005) pelo *offset* abrupto do som, o padrão de efeitos específicos observados permaneceria ainda assim por explicar (e.g., por que produz o *offset* 2 do som uma resposta mais à esquerda do que o *offset* 1). Pode notar-se, entretanto, que uma versão modificada desta última conjetura, em que a extrapolação (e.g., Cooper & Munger, 1993; Freyd & Finke, 1984; Kerzel, 2003b; Maus & Nijhawan, 2009) seria modelada por um afinamento do *timing* dos sinais visuais de posição (induzido de forma intermodal pelo som), seria compatível com o quadro explicativo que sugerimos e com o conjunto dos resultados.

Os efeitos das disparidades AV espaciais requerem uma qualificação adicional, que aponta, neste caso, para um limite do desenho experimental. A tarefa E produziu as respostas de localização mais exatas no conjunto das tarefas, o que poderá ter resultado simplesmente do facto de todas as condições AV serem aí síncronas. Com o desenho experimental utilizado, não é possível estabelecer se as pistas auditivas espaciais contribuíram também de algum modo para esse efeito de facilitação. Isso requeria pelo menos duas tarefas adicionais, uma incluindo apenas sons sincronizados e não-lateralizados (não espacializados), a outra apenas condições em que, para além de síncronos, os sons realizariam o mesmo movimento que o alvo visual, mantendo-se espaço-temporalmente alinhados com ele. Maior exatidão da localização nesta última tarefa indicaria uma contribuição da informação auditiva espacial para a melhoria da localização.

Um ponto importante, neste particular, é assim distinguir entre a verificação de um efeito direto (não temporalmente mediado) das disparidades espaciais na modulação do MR visual e a caracterização desse efeito como envolvendo ou não um contributo para a facilitação AV da resposta. Como assinalado, esta última questão não pode esclarecer-se no quadro do presente estudo. Em todo o caso, o desalinhamento espacial dos *offsets* auditivo e visual resultou, comparativamente às condições sem disparidade (*offset* 2 na tarefa E), numa menor redução do erro de localização (ver **Figura 6**, gráfico central), sendo esta diferença significativa. Isto é também consistente com a observação de que a tarefa ET, onde as disparidades espaciais se acrescentaram às disparidades temporais (presentes na tarefa T), apresentou os menores valores de redução do erro (relativamente às condições V) entre todas as tarefas (ver **Figura 6**, gráfico da direita). O impacto direto documentável das

disparidades espaciais na tarefa E consistiu assim em reduzir os efeitos facilitadores da sincronia AV e, eventualmente (não é possível estabelecer), do alinhamento espacial dos *offsets* auditivo e visual. Como nas outras tarefas, em todo o caso, estes efeitos foram inconsistentes com “enviesamentos de atração” para a localização do som. Os resultados da experiência auxiliar (ver **Figura 10**), mostrando que a posição relativa dos sons foi corretamente percebida, excluem adicionalmente explicações invocando eventuais confusões perceptivas na localização auditiva.

5.2.5. Sumário provisório

Em síntese, os resultados relativos ao deslocamento M suportaram a capacidade da informação auditiva para modular o MR visual em condições de visão central, tanto através de pistas auditivas temporais como de pistas auditivas espaciais.

Em contraste com os estudos de comparação anteriores, que reportaram enviesamentos de atração impermeáveis aos objetivos da tarefa, o presente estudo registou efeitos generalizados de melhoria AV (orientados para o objetivo). A hipótese de que incrementar a saliência perceptiva (recorrendo a disparidades maiores) e a fiabilidade (reduzindo-lhes o número/heterogeneidade) das contingências AV poderia alterar qualitativamente o impacto do som no MR visual resultou assim apoiada.

Este resultado pode ser plausivelmente explicado por uma interação, mediada pelas características da tarefa, entre processamento multissensorial e atenção intermodal (e.g., Busse et al., 2005; Fiebelkorn et al., 2010; Koelewijn et al., 2010; Macaluso et al., 2016; Staufenbiel et al., 2011; Talsma et al., 2010; Tang et al., 2016). O papel atribuído, na interpretação proposta, a contingências AV não primariamente dependentes da estrita proximidade temporal (regra temporal) converge com o reconhecimento crescente da importância de associações aprendidas na adaptação perceptiva intermodal (Heron et al., 2010; Li et al., 2015; Seitz et al., 2006; Shams et al., 2011) e com uma conceção do processamento multissensorial como um mecanismo flexível de exploração dos padrões causais e estatísticos presentes no ambiente (Ernst, 2007; Fiebelkorn et al., 2010; Hidaka et al., 2015; R. Kim et al., 2008; Parise et al., 2012; Seitz et al., 2007).

CAPÍTULO 6

DESLOCAMENTO O

6.1. RESULTADOS

As manipulações dos *onsets* e *offsets* auditivos foram realizadas nas coordenadas horizontais, convertendo assim as medidas de deslocamento M na variável dependente prioritária. Contudo, as investigações do chamado momento representacional com movimento horizontal revelam igualmente a tendência para uma localização do alvo abaixo da posição onde de facto desapareceu (Hubbard, 1995c, 1995a, 1995b, 2010). Este erro vertical, ortogonal ao movimento (no caso do movimento horizontal) foi interpretado como um análogo representativo da atração exercida pela gravidade e, desse modo, designado como “gravidade representacional” (Hubbard, 1995c, 1995b, 1997, 2020). A relação entre os dois tipos de erro, deslocamento M e deslocamento O (para movimentos horizontais), mantém-se em debate, sendo a evidência disponível escassa (Hubbard, 2020). Em particular, não são conhecidos estudos sobre a gravidade representacional com alvos sonoros especializados nem sobre o efeito de estímulos auditivos espacializados e/ou com diferentes durações no deslocamento O de alvos visuais. Hubbard & Courtney (2010) e Hubbard & Ruppel (2013) abordaram a gravidade representacional num contexto acústico e audiovisual, mas variando os estímulos auditivos no domínio das frequências, não nos domínios espacial ou temporal. A

análise dos erros de localização vertical nas três tarefas do presente estudo, bem como na tarefa ancilar, realizada com estímulos auditivos unimodais em movimento, reveste-se por isso neste quadro de um interesse adicional.

À semelhança do deslocamento M , o deslocamento O foi medido como a diferença em píxeis entre a coordenada y da resposta de localização do participante (localização percebida) e a da localização do alvo no momento do seu desaparecimento (localização real). Valores negativos indicam localizações percebidas abaixo da linha em que decorreu o movimento horizontal, valores positivos localizações percebidas acima dessa linha.

6.1.1. Efeitos das tarefas (T, E, ET) e das condições modais (V, A e AV)

Na **Figura 12** é possível verificar que os valores médios do erro de localização vertical obtido em qualquer das condições – apresentações visuais intercaladas (V_T, V_E, V_ET), audiovisuais (AV) e auditivas unimodais (AE_is) – foram negativos. A linha tracejada (e a coluna a preto) corresponde ao valor médio de deslocamento O obtido por 10 dos 31 participantes em condições visuais unimodais apresentadas numa tarefa autónoma, sem contexto audiovisual. Embora não autorize comparações diretas com os valores representados nas colunas (obtidos a partir do conjunto dos participantes), sugere uma tendência quanto à magnitude do deslocamento O em condições inteiramente visuais. Por apresentar valores de deslocamento O excedendo em mais de três vezes o intervalo interquartil foi removido da análise um participante na condição V_E. Em todas as condições, a última posição percebida do alvo foi significativamente inferior à posição real (valores de significância unicaudais): V_T: $t(30) = 3.358, p = .002$; V_E: $t(29) = 4.261, p < .001$; V_ET: $t(30) = 5.009, p < .001$; AV_T: $t(30) = 5.380, p < .001$; AV_E: $t(30) = 5.592, p < .001$; AV_ET: $t(30) = 5.857, p < .001$; AE_is: $t(30) = 1.949, p = .031$.

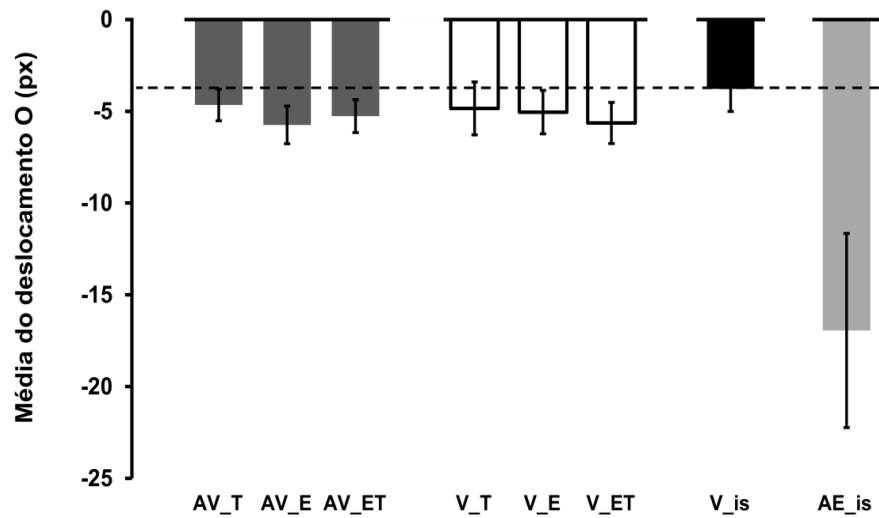


Figura 12. Deslocamentos O (valores médios em px) obtidos, nas condições audiovisuais em cada tarefa (colunas a cinzento escuro), nas condições visuais intercaladas (colunas a branco) e na tarefa ancilar com estímulos auditivos unimodais espacializados (coluna a cinzento claro). As barras de erro correspondem ao erro padrão da média. A linha tracejada (e coluna a preto) corresponde à média de deslocamento O fornecida por 10 dos 31 participantes em condições visuais unimodais apresentadas numa tarefa autónoma, sem contexto audiovisual.

As barras de erro (representando o erro padrão da média) indicam uma variância claramente superior da localização percebida na condição auditiva unimodal (AE_is) relativamente às restantes condições e menores variâncias nas condições AV do que nas condições V intercaladas. A comparação de variâncias entre condições V e AV foi conduzida em cada uma das tarefas através do teste de Pitman-Morgan para amostras emparelhadas. Com uma única exceção (V_E *versus* AV_E), revelou diferenças significativas em todos os restantes casos: V_T ($s^2 = 64.60$) > AV_T ($s^2 = 23.20$), $t(29) = 3.74$, $p = .004$; V_E ($s^2 = 42.074$) > AV_E ($s^2 = 32.67$), $t(28) = .802$, $p = .206$; V_ET ($s^2 = 39.26$) > AV_ET ($s^2 = 25.00$), $t(29) = 1.97$, $p = .0285$ (valores de p unicaudais). À semelhança da dispersão das respostas na coordenada x (deslocamento M), a dispersão das respostas na coordenada y (deslocamento O) foi por conseguinte significativamente reduzida nas condições AV relativamente às apresentações V unimodais, indicando um efeito audiovisual de aumento da precisão da localização. As comparações entre a condição auditiva unimodal AE_is ($s^2 = 1340.67$) e cada uma das condições V intercaladas (V_T, V_E, V_ET) produziram também resultados estatisticamente significativos, respetivamente $t(29) = 12.12$, $p < .001$, $t(28) = 14.47$, $p < .001$ e $t(29) = 15.82$,

$p < .001$. Tal como na coordenada horizontal, a dispersão das respostas nas apresentações auditivas unimodais foi assim significativamente superior à das apresentações visuais também na coordenada y , exprimindo, sem surpresa, uma maior precisão da localização espacial em visão central do que no domínio da audição.

A diferença de deslocamento O entre condições AV (cinzento escuro) e V intercaladas (branco) não foi estatisticamente significativa. Uma ANOVA bifatorial de medidas repetidas com T, E e ET como níveis de um fator tarefa e V e AV como níveis de um fator condição modal, revelou um resultado estatisticamente nulo da condição modal (AV versus V), $F(1,29) = .015, p = .903, \eta_p^2 = .001$. Diferentemente do deslocamento M, a adição de um som não teve assim um efeito de redução do erro de localização vertical. O fator tarefa (T, E, ET), envolvendo o tipo de pista auditiva, apresentou também resultados não significativos, $F(2,58) = .199, p = .820, \eta_p^2 = .007$. O mesmo sucedeu em duas ANOVAs de medidas repetidas conduzidas separadamente para as condições V e AV, com tarefa como único fator intra-sujeitos, ($p = .940$ e $p = .384$, respetivamente). De modo a comparar as condições AV e V com a condição AE_is, as duas ANOVAs unimodais anteriores foram repetidas com inclusão de AE_is como um quarto nível do fator tarefa. Em ambos os casos, os resultados revelaram-se não significativos, $F(3,87) = 1.188, p = .319, \eta_p^2 = .039$, na ANOVA com condições V intercaladas, $F(3,90) = 1.411, p = .245, \eta_p^2 = .045$, na ANOVA envolvendo condições AV. Apesar de graficamente óbvia, a diferença entre médias de deslocamento O na condição AE_is e nas condições AV e V intercaladas não atingiu por conseguinte o critério de significância estatística.

6.1.2. Efeitos da distância percorrida/duração do trajeto do alvo: Comparação entre condições modais (AV, V, A)

O gráfico da **Figura 13** apresenta os deslocamentos O em cada tarefa separadamente para cada uma das distâncias percorridas pelo alvo ($d1, d2$) nas condições AV. O erro de localização vertical diminui com o aumento da distância percorrida em todas as tarefas. Este efeito da distância percorrida revelou-se estatisticamente significativo numa ANOVA bifatorial de medidas repetidas com tarefa (T, E, ET) e distância ($d1, d2$) como fatores intra-sujeitos, $F(1,30) = 5.117, p = .031, \eta_p^2 = .146$. Quando avaliado separadamente em cada tarefa (testes t para amostras emparelhadas), apenas atingiu o critério da significância estatística

da tarefa ET, $t(30) = 2.478$, $p = .019$. O perfil de diferenças entre tarefas foi o mesmo em d1 e d2, com os maiores deslocamentos O na tarefa E e os menores valores na tarefa T. Estas diferenças não tiveram expressão estatística nas ANOVAs de medidas repetidas associadas a cada uma das distâncias, com tarefa como único fator intra-sujeito: d1 - $F(2,60) = .305$, $p = .738$, $\eta_p^2 = .010$; d2 - $F(2,60) = 1.726$, $p = .187$, $\eta_p^2 = .054$.

A **Figura 14** apresenta graficamente os resultados correspondentes obtidos nas condições V intercaladas. A tendência de diminuição do deslocamento O com o aumento da distância percorrida pelo alvo mantém-se nas tarefas T e E, mas não na tarefa ET. Não foi no entanto documentado qualquer efeito agregado significativo da distância, $F(1,29) = .002$, $p = .966$ (ANOVA de medidas repetidas com tarefa e distância como fatores intra-sujeitos) e as comparações t emparelhadas conduzidas separadamente por tarefa produziram resultados não significativos (menor $p = .338$, observado na tarefa ET).

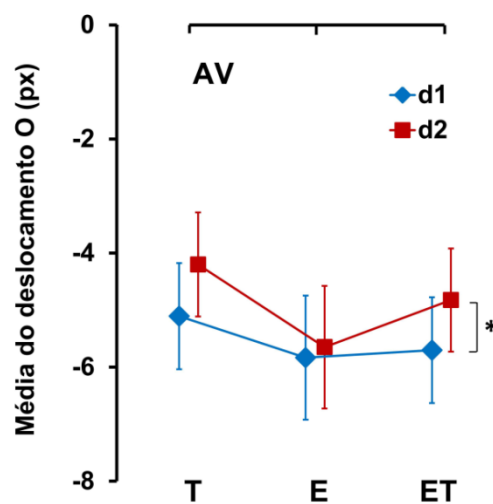


Figura 13. Deslocamento O (valores médios em px) obtido nas condições AV em cada tarefa (AV_T, AV_E, AV_ET) para cada uma das duas distâncias percorridas/durações de trajeto do alvo (d1 e d2). As barras de erro representam o erro padrão da média.

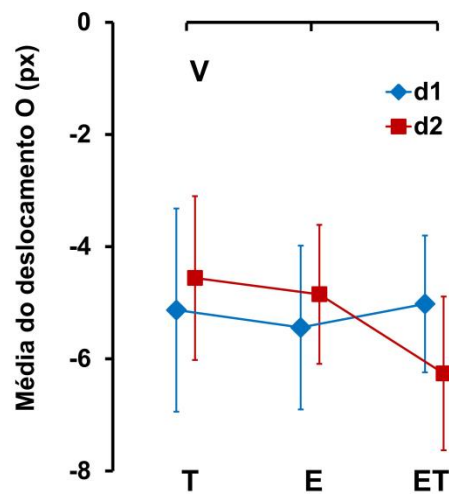


Figura 14. Deslocamentos O (valores médios em px) obtidos nas condições V intercaladas em cada tarefa (V_T, V_E, V_ET) para cada uma das duas distâncias percorridas/durações de trajeto do alvo (d1 e d2). As barras de erro representam o erro padrão da média.

Considerando que em todos os casos anteriores a resposta envolve a localização de um alvo visual (acompanhado de um som nas condições AV), justifica-se a comparação com os resultados conhecidos do efeito da distância sobre o deslocamento O de alvos visuais unimodais em movimento horizontal. Apesar da escassez dos estudos disponíveis, o resultado padrão consiste num aumento do erro de localização vertical com o aumento da extensão do trajeto percorrido pelo alvo (De Sá Teixeira et al., 2013; De Sá Teixeira & Oliveira, 2011; ver Hubbard, 2020, para revisão), contrário, por conseguinte, ao padrão aqui observado nas condições AV.

Atribuir esta discrepância a um efeito do estímulo auditivo apresentado nas condições AV implica assumir que não decorre de outros fatores distintivos do presente estudo (e.g., velocidade do alvo, natureza dos trajetos, atraso imposto à resposta de localização). Apesar de constituírem uma amostra reduzida (N=10), os participantes que procederam na tarefa auxiliar à localização dos alvos visuais isolados oferecem a possibilidade de examinar a resposta de localização à margem de qualquer efeito audiovisual direto ou contextual, contribuindo assim para esclarecer esta questão. O gráfico da esquerda na **Figura 15** apresenta o deslocamento O médio obtido nas apresentações visuais isoladas (V_is) da tarefa auxiliar para as distâncias d1 e d2. É possível verificar que a maior distância percorrida pelo alvo se encontra associada a maiores deslocamentos O, de acordo com o reportado na

literatura e contrariamente ao observado nas condições AV e V intercaladas (em contexto audiovisual) do nosso estudo. Um dos dez participantes foi removido por apresentar um deslocamento O em d2 superior a três vezes o intervalo interquartil (de notar, contudo, que a manutenção desse participante conduziria a uma acentuação, não a uma redução, da diferença no sentido verificado). Esta diferença não teve expressão estatística, $t(8) = .732$, $p = .485$ (amostras emparelhadas). O reduzido poder do teste (calculado para uma ANOVA de medidas repetidas unifatorial, o poder observado foi de .099), associado ao pequeno efetivo amostral, constitui em si mesmo uma razão provável para este resultado. No entanto, enquanto em d1 o deslocamento O não diferiu significativamente de zero (0), $t(8) = .971$, $p = .360$, em d2 revelou-se pelo contrário significativo, $t(8) = 2.806$, $p = .023$.

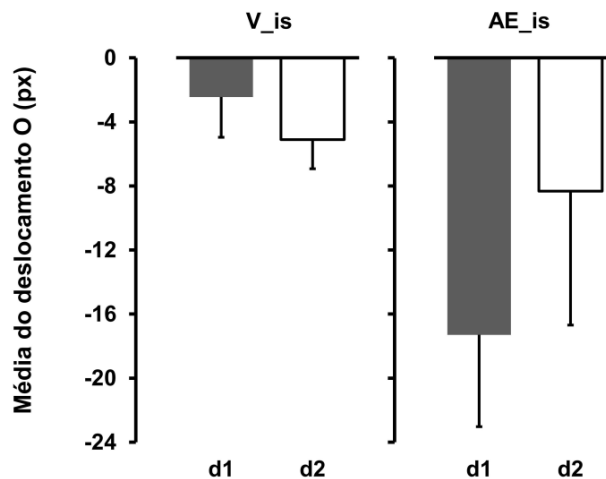


Figura 15. Deslocamentos O (valores médios em px) obtidos na tarefa auxiliar com apresentações unimodais V e A em cada um dos níveis de distância percorrida (d1 e d2). *Painel esquerdo:* condições V_is. *Painel direito:* condições AE_is. As barras de erro representam o erro padrão da média.

O sentido contrário do efeito da extensão/duração do trajeto do alvo nas condições V_is relativamente às condições V intercaladas e AV aponta assim para um impacto da informação auditiva presente nas tarefas audiovisuais, suscitando a questão de como a extensão/duração do trajeto afeta a localização de alvos auditivos unimodais em movimento (AE_is). No painel direito da **Figura 15** estão representadas as médias de deslocamento O obtidas na tarefa auxiliar para as condições AE_is, mostrando um decréscimo claro do erro de localização vertical com o aumento da distância percorrida. Esta diferença não foi significativa, $t(30) = .498$, $p = .622$. No entanto, o deslocamento O diferiu significativamente

de zero (0) em d1, $t(30) = 2.572, p = .019$, e apresentou-se estatisticamente nulo em d2, $t(30) = 1.355, p = .186$, contrariamente ao sucedido nas condições V_{is}. Estes resultados indicam assim que o efeito da distância percorrida no erro de localização vertical é na realidade oposto (respetivamente, crescente e decrescente) nas modalidades visual e auditiva e são consistentes com uma atribuição da inversão do efeito da distância/duração na localização de alvos visuais nas condições AV e V intercaladas ao impacto da informação auditiva presente nas três tarefas.

6.1.3. Efeitos do ponto de desapareção: Comparação entre condições modais (AV, V, A)

A **Figura 16** e a **Figura 17** apresentam, respetivamente para as condições AV e V intercaladas, as médias de deslocamento O nas três tarefas (T, E, ET) em cada um dos pontos de desapareção do alvo. Ambos os gráficos documentam um aumento do deslocamento O, transversal às três tarefas, no ponto de desapareção mais extremo no sentido do movimento, à direita do centro do ecrã (p2). Este efeito não atingiu significância estatística em nenhuma das condições modais: $F(1,30) = 1.601, p = .216, \eta_p^2 = .051$ e $F(1,29) = 1.562, p = .221, \eta_p^2 = .051$, respetivamente, nas condições AV e V (ANOVAs de medidas repetidas com tarefa e ponto de desapareção como fatores). Testes *t* emparelhados conduzidos separadamente em cada uma das tarefas produziram igualmente resultados estatisticamente nulos do ponto de desapareção (menor *p* observado = .177, associado à tarefa T, nas condições AV; menor *p* observado = .288, associado à tarefa E, nas condições V intercaladas).

Nas condições AV (**Figura 16**), o perfil da diferença entre tarefas é idêntico nos dois pontos (p1 e p2), com os maiores erros de localização vertical na tarefa E. Este mesmo padrão foi observado nas condições AV no caso das duas distâncias percorridas pelo alvo (ver **Figura 13**), surgindo assim como graficamente consistente. No entanto, tal como no caso das distâncias, não deu lugar a diferenças estatisticamente significativas entre tarefas em qualquer dos pontos: p1 - $F(2,60) = 1.153, p = .323, \eta_p^2 = .037$; p2 - $F(2,60) = .578, p = .564, \eta_p^2 = .019$. Nas condições V intercaladas as diferenças entre tarefas apresentam um padrão distinto e variam com o ponto de desapareção. O fator tarefa não apresentou também diferenças significativas em nenhum dos pontos: p1 - $F(2,58) = .065, p = .937, \eta_p^2 = .002$; p2 - $F(2,58) = .095, p = .910, \eta_p^2 = .003$.

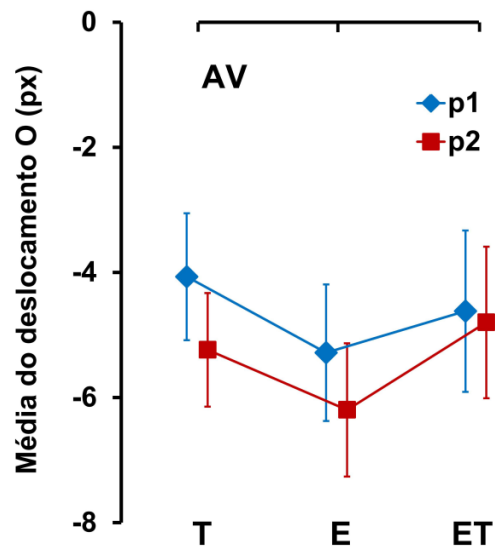


Figura 16. Deslocamento O (valores médios em px) obtido nas condições AV em cada tarefa (AV_T, AV_E, AV_ET) para cada um dos dois pontos de desaparecimento do alvo (p1 e p2). As barras de erro representam o erro padrão da média.

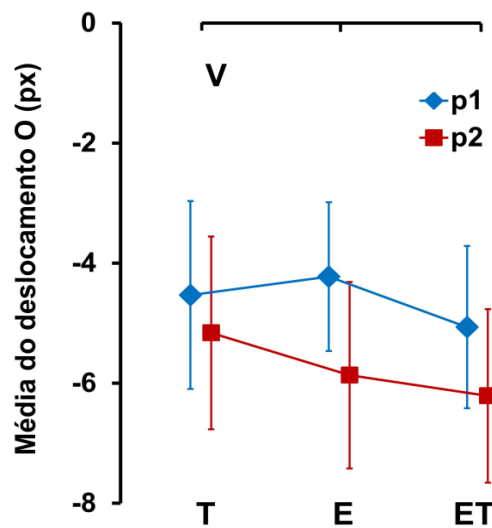


Figura 17. Deslocamento O (valores médios em px) obtido nas condições V intercaladas (V_T, V_E, V_ET) para cada um dos dois pontos de desaparecimento do alvo (p1 e p2). As barras de erro representam o erro padrão da média.

A amostra de 10 participantes que realizaram a tarefa auxiliar com condições visuais independentes de qualquer contexto audiovisual possibilita de novo comparar os efeitos do ponto de desaparecimento devidos a informação unicamente visual (V_is) e exclusivamente auditiva (AE_is). Regra geral, os estudos de MR visual que variaram o ponto de desaparecimento do alvo mantiveram essa variação confundida com a da distância percorrida pelo alvo, não permitindo dissociar os efeitos das duas variáveis. Do nosso melhor conhecimento, apenas um estudo foi conduzido até à data onde o ponto de desaparecimento e a distância percorrida foram variados ortogonalmente (De Sá Teixeira & Oliveira, 2011). Esse estudo analisou o deslocamento M numa tarefa de MR visual, não oferecendo por conseguinte informação sobre a modulação do deslocamento O pelos pontos de desaparecimento. A manipulação ortogonal dos dois pontos de desaparecimento (p1 e p2) e das distâncias (d1 e d2) no presente estudo possibilita examinar, pela primeira vez, os efeitos do ponto de desaparecimento no deslocamento O, quer de alvos visuais, quer auditivos.

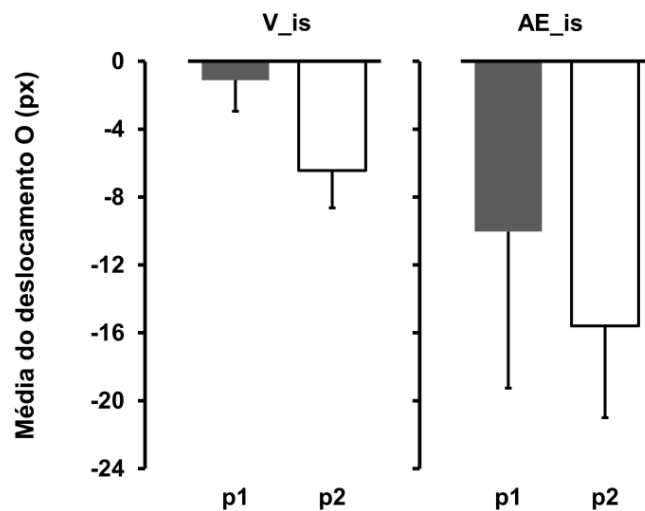


Figura 18. Deslocamentos O (valores médios em px) obtidos na tarefa auxiliar com apresentações unimodais V e A em cada um dos níveis do ponto de desaparecimento. *Painel esquerdo:* condições V_is. *Painel direito:* condições AE_is. As barras de erro representam o erro padrão da média.

Os painéis esquerdo e direito da **Figura 18** apresentam, respetivamente para as condições V_is e AE_is da tarefa auxiliar, os valores médios do erro vertical em cada um dos pontos de desaparecimento. Tal como nas condições AV e V intercaladas, o deslocamento O aumenta em ambos os casos no ponto de desaparecimento mais extremo no sentido do

movimento. Em nenhuma das condições a diferença teve expressão estatística, V_{is} : $t(8) = 1.661$, $p = .135$; AE_{is} : $t(30) = .743$, $p = .463$. No entanto, só em p2 o erro de localização vertical na condição V_{is} diferiu significativamente de zero (0), $t(8) = 2.914$, $p = .019$, o mesmo sucedendo na condição AE_{is} , $t(30) = 2.881$, $p = .007$.

Tomados no seu conjunto, os resultados indicam assim que, diferentemente do observado para a distância percorrida, que opera inversamente nas modalidades A (maior erro vertical na distância d1) e V (maior erro vertical na distância d2), o ponto de desaparecimento produz efeitos de deslocamento O de mesmo sentido nas duas modalidades (maior deslocamento O em p2). Sem surpresa, esse mesmo sentido emergiu no contexto das tarefas audiovisuais, tanto nas condições AV como nas condições V intercaladas.

6.1.4. Efeitos do começo (*onset*) e termo (*offset*) do estímulo auditivo nas condições audiovisuais (AV)

Uma ANOVA de medidas repetidas de 5 vias, 3 (tarefa) \times 2 (distância/duração) \times 2 (ponto de desaparecimento) \times 3 (*onset*) \times 3 (*offset*) não apresentou efeitos principais ou interações (de qualquer ordem) significativas. Apenas o fator distância/duração (d1, d2) se aproximou do critério de significância estatística, $F(1,30) = 3.838$, $p = .059$, $\eta_p^2 = .113$.

Os 3 painéis da **Figura 19** representam graficamente os valores médios de deslocamento O em função do começo (*onset*) do estímulo auditivo nas apresentações audiovisuais e com o termo (*offset*) do som como parâmetro das curvas em cada uma das três tarefas (da direita para a esquerda: T, E, ET). As linhas tracejadas correspondem à média de deslocamento O obtida nas condições V intercaladas em cada tarefa. Nas análises seguintes deve ter-se em conta que valores crescentes de deslocamento O correspondem na realidade a valores mais negativos (numericamente decrescentes) na ordenada e valores decrescentes ou reduções no deslocamento O a valores menos negativos (numericamente crescentes) na ordenada.

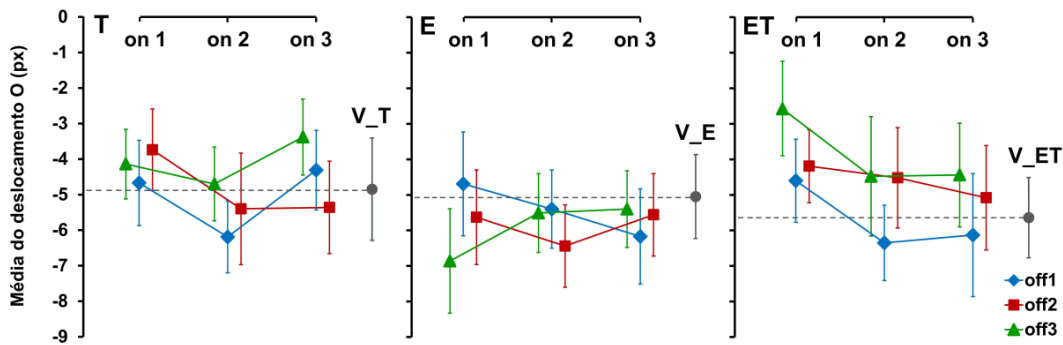


Figura 19. Gráficos fatoriais *onset* \times *offset* em cada uma das tarefas AV (T, E, ET, da esquerda para a direita). O deslocamento O encontra-se em ordenada (valores médios em px), os níveis de *onset* em abcissa e os níveis de *offset* parametrizam as curvas. As linhas tracejadas representam o erro vertical médio nas condições V_intercaladas. As barras de erro representam o erro padrão da média.

O gráfico correspondente à tarefa ET, no painel da direita, apresenta o padrão fatorial mais claro, com um efeito crescente do *onset* no deslocamento O (menores erros verticais no *onset* 1, quando o som tem início mais cedo e atrás da origem movimento visual) e um efeito crescente do *offset* (menores erros verticais no *offset* 3, quando o som termina mais tarde e à frente do movimento visual). Estas indicações foram parcialmente suportados por efeitos principais significativos do *offset* numa ANOVA de medidas repetidas conduzida na tarefa ET, respetivamente $F(2,60) = 5.360$, $p = .007$, $\eta_p^2 = .152$. Comparações aos pares subsequentes, com ajuste de Bonferroni, mostraram uma diferença significativa entre os níveis 1 e 3 de *offset* ($p = .006$).

De forma mais irregular, o gráfico correspondente à tarefa T, no painel da esquerda, ilustra igualmente uma tendência para menores erros verticais no nível 3 de *offset* (som termina mais tarde do que o movimento visual). Contudo, nenhum dos fatores apresentou efeitos significativos na ANOVA associada: *onset* - $F(2,60) = 1.817$, $p = .171$, $\eta_p^2 = .057$; *offset* - $F(2,60) = 1.231$, $p = .299$, $\eta_p^2 = .039$. O gráfico correspondente à tarefa E, no painel central, documenta uma sobreposição desordenada das linhas, sugerindo a ausência de qualquer efeito sistemático dos fatores. Esta sugestão foi suportada por uma ANOVA de medidas repetidas, que produziu resultados estatisticamente nulos: *onset* - $F(2,60) = .02$, $p = .978$; *offset* - $F(2,60) = .440$, $p = .646$, $\eta_p^2 = .014$. Apesar de apenas o fator *offset*, na tarefa ET, ter produzido um efeito significativo, a inspeção visual dos gráficos obtidos separadamente para os níveis de *onset* (médias agregadas através dos níveis de *offset*) e de *offset* (médias

agregadas através dos níveis de *onset*) revelam tendências partilhadas entre as tarefas T e ET e um perfil distintivo da tarefa E. Os valores de deslocamento O associados aos três níveis de *onset* em cada uma das tarefas (T, E, ET) são apresentados na **Figura 20**, e os valores associadas aos três níveis de *offset*, também para cada uma das tarefas, na **Figura 21**.

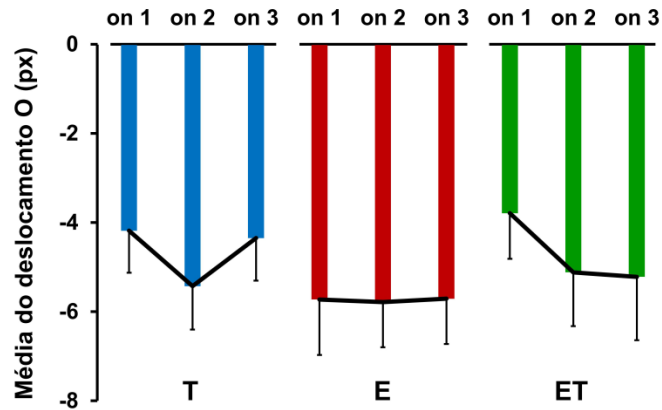


Figura 20. Deslocamento O (valores médios em px) obtidos nas condições AV para cada um dos níveis de *onset* (on 1, on 2, on 3) em cada uma das tarefas (T, E, ET). As barras de erro representam o erro padrão da média.

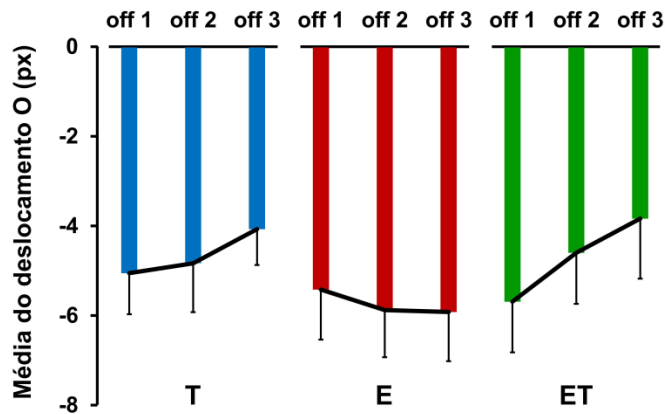


Figura 21. Deslocamento O (valores médios em px) obtidos nas condições AV para cada um dos níveis de *offset* (off 1, off 2, off 3) em cada uma das tarefas (T, E, ET). As barras de erro representam o erro padrão da média.

Os perfis relativos ao *onset* documentam uma tendência semelhante nas tarefas T e ET para valores de deslocamento O mais reduzidos no nível 1 de *onset* e deslocamentos O superiores (mais negativos) no nível 2 de *onset*. As duas tarefas divergem quanto à posição relativa do nível 3 de *onset*, que em T se associa a valores de erro vertical próximos dos de *onset* 1 e em ET valores de erro próximos dos de *onset* 2. Entretanto, a tarefa E distingue-se das restantes por duas características: valores de deslocamento O invariavelmente superiores (mais negativos) e virtualmente idênticos através dos três níveis de *onset* (ver linha horizontal a preto na **Figura 20**). Quanto aos valores de *offset*, pode notar-se a convergência das tarefas T e ET num mesmo perfil de decréscimo do valor de deslocamento O com a passagem dos níveis 1 a 3 (menor valor em *offset* 3: ver linhas a preto com declive positivo na **Figura 21**) e, de novo, um padrão de resultados distinto na tarefa E, que apresenta erros verticais muito semelhantes em todos os níveis de *offset*, com apenas uma leve tendência para a acentuação do deslocamento O com a passagem dos níveis 1 a 3.

Considerando que na tarefa E, diferentemente das tarefas T e ET, as discrepâncias audiovisuais são unicamente espaciais, envolvendo sons com a mesma duração do estímulo visual (as variações do *onset* e do *offset* determinam diferentes extensões do movimento do som, percorridas a diferentes velocidades, mas não diferentes durações), as tendências assinaladas nas tarefas T e ET e ausentes na tarefa E poderão refletir um impacto das propriedades de duração do som no erro vertical. Com efeito, o nível 1 de *onset* (que nas tarefas T e ET apresenta os menores valores de deslocamento O) e o nível 3 de *offset* (que em ambas as tarefas apresenta também o menor erro vertical) encontram-se em média associados às maiores durações do estímulo auditivo. Os resultados observados poderiam assim sugerir que o aumento da duração do som associado induz uma maior resistência ao deslocamento O do alvo visual ou, mais exatamente, uma maior resistência aos efeitos da gravidade representacional (deslocação para baixo do ponto de desapareção percebido relativamente ao ponto de desapareção real).

Como assinalado atrás, estas tendências tiveram expressão estatística, no caso do *offset*, na tarefa ET, onde o fator *onset* se aproximou igualmente do critério de significância estatística ($p = .075$). São além disso consistentes com a verificação anterior de uma diminuição do deslocamento O com a distância percorrida pelo alvo (o que, a uma velocidade constante, como é o caso, implica uma maior duração do trajeto) quer no caso de alvos auditivos unimodais (AE_is) (ver **Figura 15**, painel da direita) quer nas condições AV das tarefas T e ET (ver **Figura 13**). A interpretação destes resultados como um efeito das propriedades de duração do som pode encontrar algum apoio na literatura sobre a

prioridade da informação de duração (e não de velocidade) na percepção do movimento auditivo e suas características (Carlile & Leung, 2016; T. C. A. Freeman et al., 2014; Roggerone et al., 2019). Embora utilizada conjuntamente com a informação de distância, de acordo com alguns modelos (e.g., Perrott, Costantino, & Ball, 1993), para o cálculo da velocidade do movimento de uma fonte sonora, a duração poderá, enquanto dimensão processada de forma independente, ter efeitos distintos da distância no que se refere à localização vertical percebida do movimento, em particular no caso de movimentos horizontais. A circunstância de estes efeitos serem particularmente visíveis na tarefa ET, na qual a variação das durações envolve sons em movimento (e não estáticos, como na tarefa T), é compatível com o aumento documentado da duração percebida de estímulos em movimento (dinâmicos) relativamente aos seus equivalentes estáticos (Brown, 1995; Šimko, Aalto, Lippus, Włodarczak, & Vainio, 2015) – com a implicação de que o aumento da duração percebida (e não da duração física) do som constituiria neste cenário o fator responsável pelos efeitos discutidos.

Contrastando com os efeitos de melhoria audiovisual observados no deslocamento M (redução do erro de localização horizontal nas condições AV relativamente às condições V intercaladas), a **Figura 19** assinala que, no caso do deslocamento O, a componente auditiva das apresentações AV não contribui, em geral, para uma redução do erro vertical. Na tarefa E (gráfico central) o deslocamento O nas condições AV é maior (mais negativo) do que o registado nas condições V_E intercaladas (representadas pela linha tracejada). Na tarefa T situa-se em valores médios praticamente idênticos ao das condições V_T. Apenas na tarefa ET o deslocamento O na maioria das condições audiovisuais se apresenta tendencialmente menor (menos negativo) do que nas condições V_ET. Nenhuma das comparações estatísticas realizadas (ANOVAs de medidas repetidas *onset* × *offset*, com inclusão das apresentações visuais intercaladas como um dos níveis de *offset*, seguidas de testes *t* emparelhados com ajuste de Bonferroni) evidenciou diferenças significativas entre condições AV e V_intercaladas em qualquer das tarefas ($p < .05$), apoiando desse modo a ausência de uma redução do erro de localização nas condições audiovisuais.

6.1.5. Correlações entre o deslocamento M e o deslocamento O

A questão das relações entre o deslocamento M e o deslocamento O constitui um problema irresolvido, não sendo claro se envolvem o mesmo mecanismo, mecanismos diferentes ou

uma sobreposição parcial de mecanismos (De Sá Teixeira, 2016; De Sá Teixeira & Oliveira, 2011; Hubbard, 1995c, 1995a, 1995b, 2005, 2020). Uma questão associada, apesar de relativamente independente, é a de saber se, enquanto variáveis dependentes, apresentam covariações ou correlações na sua magnitude ou se, em alternativa, variam ortogonalmente (de forma independente) de um ponto de vista estatístico. Apesar de não oferecer respostas diretas à questão da existência de mecanismos comuns aos dois tipos de erro (um problema de causalidade, não de associação), o exame descritivo do perfil de associação das duas variáveis poderá contudo fornecer elementos úteis para a discussão.

Do nosso melhor conhecimento, a correlação entre o deslocamento M e o deslocamento O não tem sido objeto de análise, excetuando-se os dados de uma tese não publicada (Martins, 2015) que encontrou correlações significativas entre a magnitude dos dois deslocamentos numa tarefa de momento representacional com estímulos visuais, tanto num grupo de participantes com paralisia cerebral como num grupo de participantes controlo (participantes com maiores deslocamentos M tendiam também a apresentar maiores deslocamentos O). A generalidade destes resultados, obtidos na modalidade visual com movimentos horizontais, não foi nunca avaliada em estudos posteriores. Quanto à eventual associação entre os dois erros em contextos auditivos (A) ou audiovisuais (AV), não existem até à data, dados disponíveis.

Na **Tabela 2** encontram-se resumidas as correlações (r de Pearson) registadas entre os erros de localização no sentido do movimento (M) e ortogonais ao movimento (O) nas condições AV, V intercaladas, auditivas espacializadas (AE) e, para os dez participantes que as forneceram na tarefa ancilar e visuais isoladas (V_{is}). Nenhuma das correlações se revelou significativa ($p < .05$). De notar que associações crescentes dos dois tipos de erro deveriam neste caso manifestar-se por correlações negativas significativas (dada a associação do erro vertical com valores negativos) e associações decrescentes por correlações positivas significativas. Em nenhuma das condições modais, por conseguinte, os dois tipos de erro mostraram estar associados. Este resultado limita a generalidade dos dados reportados em Martins (2015), sugerindo eventualmente um papel importante de características específicas da tarefa na ocorrência de covariações de magnitude entre as duas formas de deslocamento. Por outro lado, aponta pela primeira vez para uma independência estatística do deslocamento M e do deslocamento O tanto na modalidade auditiva como na modalidade visual no contexto de apresentações AV.

Tabela 2. Correlações (r de Pearson) entre os deslocamentos M e O nas diferentes condições modais.

	Deslocamento O			
	AV	V int.	AE_is	V_is
Deslocamento M	.099	-.119	.114	-.179
<i>sig.</i> (bicaudal)	.595	.523	.548	.645
N	31	31	30	9

Tabela 3. Correlações (r de Pearson) entre os deslocamentos M e O nas condições AV e V intercaladas em cada uma das tarefas (T, E, ET).

	Deslocamento O					
	AV_T	AV_E	AV_ET	V_T	V_E	V_ET
Deslocamento M	.107	.096	.156	.237	-.220	.019
<i>sig.</i> (bicaudal)	.566	.609	.403	.200	.242	.998
N	31	31	31	31	30	31

As correlações apresentadas na **Tabela 3** entre o deslocamento M e o deslocamento O, agora no contexto de cada uma das tarefas (T, E, ET), reforçam as conclusões extraídas da **Tabela 2**, documentando de novo a ausência de correlações significativas tanto nas condições AV como V intercaladas.

6.2. DISCUSSÃO

6.2.1. Semelhanças e diferenças entre a modalidade visual e auditiva no deslocamento O

Uma primeira indicação da análise do deslocamento O é a de que, tal como no caso de alvos visuais, o erro vertical associado à chamada “gravidade representacional”, ocorre igualmente com alvos auditivos unimodais (AE_is), alvos visuais unimodais intercalados em contextos audiovisuais (V_int), e alvos visuais acompanhados por um som (apresentações AV). Os

valores de deslocamento 0 foram mais pronunciados nas apresentações auditivas isoladas (AE_is) (ver **Figura 12**), mas não se verificaram diferenças significativas entre as condições modais (V_int, A, AV). Trata-se, ao que sabemos, do primeiro resultado indicando a existência de erros de localização vertical com uma direção análoga (no sentido da ação da gravidade) nas modalidades auditiva e visual quando se utilizam alvos em movimento horizontal. Acrescentando-se à evidência já acumulada sobre a existência de deslocamento M para diante na localização de alvos auditivos espacializados (Getzmann, 2005a; Getzmann & Lewald, 2007, 2009; Getzmann et al., 2004; Schmiedchen et al., 2013), estes dados contribuem para suportar a hipótese de mecanismos subjacentes comuns à localização de alvos dinâmicos nas duas modalidades (ver, e.g., Hubbard, 2014a; Schmiedchen et al., 2013) e, mais geralmente, duma possível natureza amodal do momento representacional (Hubbard, 2017).

Um possível limite destes resultados prende-se com a sua eventual dependência relativamente a características do som utilizado, em particular a banda de frequências selecionada (centrada nos 2 kHz). A existência de correspondências intermodais consistentes entre a altura do som, associada às propriedades de frequência, e a sugestão de localizações em altura no espaço encontra-se bem documentada (sons mais agudos são associados a objetos com localizações mais elevadas, sons mais graves a objetos com localizações mais baixas: ver Spence, 2011, para uma revisão). Estas correspondências abrem a possibilidade de que a ocorrência de um erro “para baixo” com alvos auditivos em movimento horizontal, bem como a magnitude desse erro possa depender de forma importante da gama de frequências utilizada. Uma vez que a frequência do som se manteve constante em todas as condições, essa hipótese não pode ser descartada nem validada pelo presente estudo, permanecendo uma questão em aberto.

O paralelismo dos resultados relativos ao deslocamento 0 nas modalidades visual e auditiva não foi, por outro lado, completo. Na verdade, enquanto algumas das variáveis que modularam a magnitude do erro vertical apresentaram efeitos convergentes nas duas modalidades (como o ponto de desaparecimento do alvo), outras mostraram ter efeitos divergentes (como a distância percorrida pelo alvo). Tal como na modalidade visual (V_is e V_int), o ponto de desaparecimento mais extremo no sentido do movimento (p2) apresentou deslocamentos 0 mais acentuados nas condições auditivas unimodais (AE_is) (ver **Figura 18**). Contudo, a distância percorrida pelo alvo/duração do trajeto apresentou efeitos distintos e mesmo opostos nas duas modalidades, com a maior distância (d2) associada a

uma redução do deslocamento O nas condições AE_is e a maiores deslocamentos O no caso das condições V_is (ver **Figura 15**).

Semelhanças globais e divergências locais entre as duas modalidades têm sido igualmente assinaladas na literatura relativa ao deslocamento M. Por exemplo, enquanto com alvos visuais o deslocamento M aumenta com a velocidade do alvo, efeitos crescentes (Feinkohl et al., 2014), decrescentes (Getzmann et al., 2004) e condicionais à direção do movimento (Schmiedchen et al., 2013) foram reportados com alvos auditivos. Uma outra divergência reportada envolve o efeito dos movimentos oculares: o deslocamento M de alvos visuais diminui substancialmente, chegando mesmo a desaparecer, na ausência de movimentos oculares (Kerzel, 2000, 2005), mas mantém-se idêntico com e sem movimentos oculares no caso de alvos auditivos (Getzmann, 2005a; Getzmann & Lewald, 2009; Schmiedchen et al., 2012). A verificação de uma divergência quanto aos efeitos da distância no deslocamento O replica assim um padrão de discrepância pontual entre as duas modalidades, já aparente no contexto dos estudos de momento representacional e sugestivo de uma sobreposição parcial, não completa, dos mecanismos envolvidos na produção deste tipo de erros de localização nos domínios visual e auditivo.

6.2.2. Mecanismos de integração da informação visual e auditiva no deslocamento O: Avaliação indicativa do modelo de máxima verosimilhança da integração multissensorial

Dado o sentido contrário dos efeitos da distância no deslocamento O em cada uma das modalidades, a forma como esta variável modula o erro vertical nas condições AV adquire um interesse particular. Como indicado acima (ver **Figura 13**), o deslocamento O foi mais reduzido em d2 do que em d1 nas condições AV em todas as tarefas (embora apenas de forma residual na tarefa E), em conformidade com o sentido do efeito da distância nas apresentações auditivas unimodais (AE_is) e em contraste com o aumento do deslocamento O em d2 nas condições visuais isoladas (V_is). Isto sugere um impacto importante da informação auditiva na localização vertical de um alvo visual, capaz mesmo de inverter o sentido usual do efeito da distância em situações visuais unimodais. No entanto, em si mesmo, não esclarece o mecanismo de combinação da informação auditiva com a informação visual, nem tão-pouco se um mesmo mecanismo pode igualmente acomodar o efeito nas condições AV do ponto de desaparecimento do alvo, que, contrariamente à distância operou de

forma convergente nas condições visuais e auditivas isoladas (aumento do erro vertical com o ponto mais extremo no sentido do movimento: ver **Figura 18**).

Uma das propostas mais influentes sobre o mecanismo de combinação de informações intermodais é o Modelo de Máxima Verosimilhança da integração sensorial (MMV), de inspiração Bayesiana (Alais & Burr, 2004; Ernst & Banks, 2002). O modelo prevê uma integração por média ponderada, sendo a ponderação das informações efetuada pela fiabilidade (operacionalizada como o inverso da variância) respetiva de cada modalidade. No caso do deslocamento M este modelo revelou-se inaplicável (ver discussão anterior), mas poderia porventura oferecer um quadro de explicação unitário para os efeitos das variáveis distância percorrida e ponto de desaparecimento do alvo no contexto do deslocamento O. As **Figuras 22 e 23** contribuem elementos úteis, ainda que limitados, para avaliar essa possibilidade. Uma das condições requeridas para um teste sumário à aplicabilidade do MMV é dispor da variância das respostas de localização a estímulos visuais à margem de qualquer contexto audiovisual e da variância das respostas a estímulos auditivos unimodais. Apenas os dez (10) participantes que realizaram também a experiência com condições V_is, para além da experiência com condições AE_is, na tarefa auxiliar puderam ser assim utilizados para o efeito. Adicionalmente, porque as condições AE_is se limitaram a trajetos iniciados no nível 2 de *onset* (correspondentes aos pontos de origem do movimento visual), apenas o subconjunto das condições AV das tarefas E e ET (sons espacializados) em que o trajeto do som teve origem em *onset* 2 permite comparações com as previsões deriváveis do modelo. Estas circunstâncias condicionam o alcance das análises seguintes, que devem por conseguinte ser olhadas como tendo sobretudo um valor indicativo.

A **Figura 22** apresenta as médias de deslocamento O nas condições visuais isoladas (V_is: círculos verdes), auditivas unimodais (AE_is: círculos azuis), audiovisuais (AV: barras cinzenta) e visuais intercaladas (V_int: barras vermelhas), com os respetivos erros padrão da média, em cada uma das distâncias. Como indicado acima, apenas os dados obtidos por 10 dos 31 participantes, limitados a um subconjunto de condições audiovisuais nas tarefas E e ET e às condições visuais intercaladas nessas duas tarefas contribuíram para o cálculo das médias (os valores das condições AV e V_int representam médias agregadas através das tarefas E e ET). Tanto em d1 como em d2, a barra cinzenta (condições AV) localiza-se entre os círculos verdes (V_is) e azuis (AE_is). Esta localização é compatível, qualitativamente, com uma integração AV por média ponderada. No entanto, a maior proximidade do deslocamento O nas condições AV à média das condições auditivas unimodais (AE_is) em d1 implicaria uma maior ponderação da informação auditiva do que da informação visual, contrária ao que a

variabilidade superior dos resultados em AE_is implicaria, de acordo com o MMV (ver erros padrão da média na **Figura 22**). Em d2, o deslocamento nas condições AV encontra-se ligeiramente mais próximo da média das condições V_is do que das condições AE_is. Contudo, a dispersão muito superior dos resultados em AE_is requereria, de acordo com o MMV, que esta proximidade fosse bastante mais acentuada.

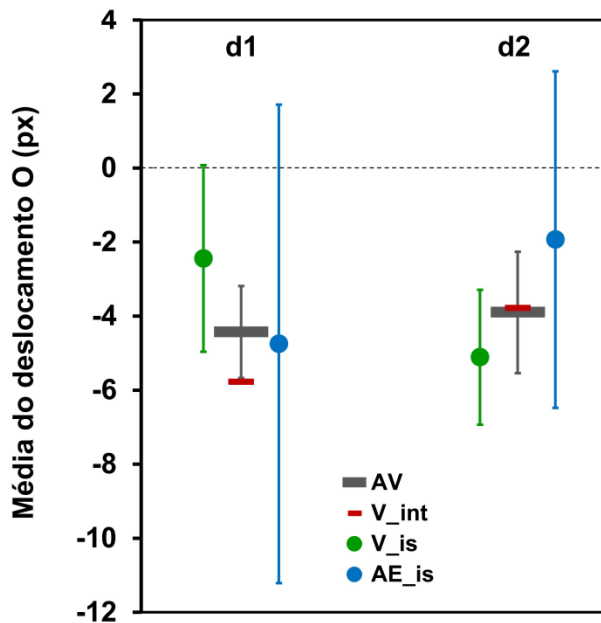


Figura 22. Média do deslocamento O obtida nas condições V sem contexto audiovisual (V_is: círculo verde), auditivas unimodais (AE_is: círculo azul), AV (barra cinza) e V intercaladas (V_int: barra vermelha) em cada uma das condições de distância percorrida pelo alvo (d1 e d2). Os valores destas duas últimas condições representam médias agregadas através das tarefas E e ET. De modo a assegurar comparabilidade entre condições, apenas os dados fornecidos pelos 10 participantes que realizaram as tarefas visual unimodal e auditiva unimodal na experiência ancilar foram utilizados. As barras de erro representam o erro padrão da média.

Na **Tabela 4** apresentam-se os pesos relativos (W_V e W_A) das informações V_is e AE_is calculados com base nas fórmulas

$$W_V = \frac{1/s_V^2}{1/s_V^2 + 1/s_A^2} \quad \text{e} \quad W_A = \frac{1/s_A^2}{1/s_V^2 + 1/s_A^2},$$

e os pesos relativos estimados com a ferramenta Solver do Excel por minimização da discrepância entre os valores de deslocamento O previstos a partir do MMV e os valores observados, submetida à restrição de que os dois pesos relativos deveriam somar 1. É possível observar a inversão do sentido das ponderações entre os valores calculados e estimados em d1 (peso estimado de AE_{is} = 0.863, substancialmente superior ao valor estimado de V_{is} = 0.137) e, em d2, a sobreponderação relativa dos valores estimados de AE_{is} (0.381) relativamente aos valores calculados pelas fórmulas de ponderação normativas (0.138).

Tabela 4. Pesos calculados (com base no MMV) e estimados da informação visual unimodal (V_{is}) e auditiva unimodal (AE_{is}) quanto à localização vertical percebida do alvo, em d1 e d2, num cenário de integração AV por média ponderada.

	Peso V_{is}		Peso AE_{is}	
	Calculado	Estimado	Calculado	Estimado
d1	0.880	0.137	0.120	0.863
d2	0.862	0.619	0.138	0.381

A **Figura 23** mostra, por sua vez que, tal como no caso da distância, a barra cinzenta, correspondente às condições AV, se situa entre os círculos verdes (V_{is}) e azuis (AE_{is}) tanto nas condições p1 como p2, um resultado compatível com uma integração das informações visual e auditiva por média ponderada. Contudo, em p1, a maior proximidade aos valores médios de AE_{is} é de novo contrária à maior ponderação da informação V_{is} que deveria esperar-se (de acordo com o MMV) face há muito menor variabilidade da localização visual. Em p2, a média dos valores AV encontra-se mais próxima dos valores V_{is} , de acordo com o esperado. Contudo, face à diferença de variabilidade entre as duas fontes de informação, essa proximidade fica ainda aquém da que seria normativamente prevista pelo modelo MMV. Os valores expressos na **Tabela 5** documentam isso mesmo, ilustrando, em p1, um peso estimado da informação auditiva (0.701) superior ao da informação visual (0.299) e em p2 um peso empiricamente estimado da informação auditiva (0.316) superior à sua ponderação normativa (0.120).

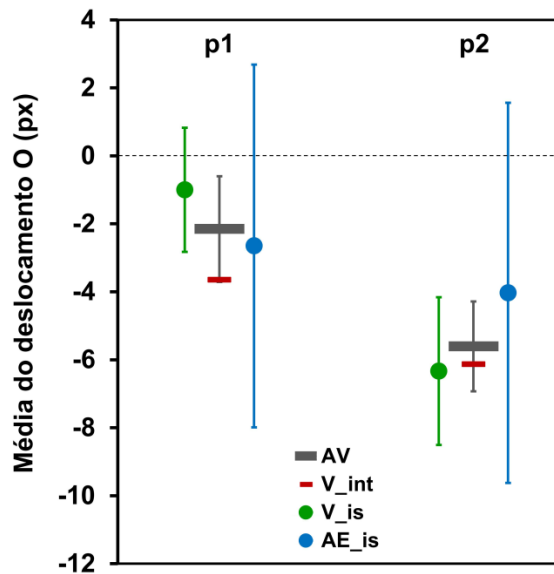


Figura 23. Média do deslocamento O obtida nas condições V sem contexto audiovisual (V_is: círculo verde), auditivas unimodais (AE_is: círculo azul), AV (barra cinza) e V intercaladas (V_int: barra vermelha) em cada um dos pontos de desaparecimento do alvo (p1 e p2). Os valores destas duas últimas condições representam médias agregadas através das tarefas (E, ET). Apenas os dados fornecidos pelos 10 participantes que realizaram as tarefas visual unimodal e auditiva unimodal na experiência auxiliar foram utilizados. As barras de erro representam o erro padrão da média.

Tabela 5. Pesos calculados (com base no MMV) e estimados da informação visual unimodal (V_is) e auditiva unimodal (AE_is) quanto à localização vertical percebida do alvo, em p1 e p2, num cenário de integração AV por média ponderada.

	Peso V_is		Peso AE_is	
	Calculado	Estimado	Calculado	Estimado
p1	0.905	0.299	0.095	0.701
p2	0.880	0.684	0.120	0.316

A indicação resultante destas análises é assim a de que, enquanto um modelo de média permanece em princípio aplicável à combinação das informações visuais e auditivas no caso do deslocamento O (abarcando tanto os efeitos da distância percorrida pelo alvo como os do seu ponto de desaparecimento), o esquema de ponderação ótima pelo inverso da

variância previsto pelo MMV não é empiricamente suportado. Estes resultados alinham-se facilmente com outros disponíveis na literatura questionando a generalidade de um mecanismo de combinação ótima da informação procedente de diferentes sentidos e sugerindo, em alternativa, formas de ponderação subótimas e não-ótimas (Arnold et al., 2019; Meijer et al., 2019).

A evidência relativamente às condições V_{int} (barras vermelhas) sugere, diferentemente, que um eventual efeito do contexto audiovisual nas condições visuais intercaladas é incompatível com um modelo de média. Isto resulta da localização do deslocamento O das condições V_{int} , tanto em $d1$ como em $p1$, fora do intervalo definido pelos círculos verde (V_{is}) e azul (AE_{is}). Esta indicação é consistente com um cenário no qual os efeitos do contexto AV passariam pela alteração de critérios de decisão quanto à localização do alvo (um mecanismo primariamente cognitivo) e não por alterações especificamente perceptivas (Shams, 2012).

6.2.3. Semelhanças e diferenças entre o deslocamento O e o deslocamento M

Contrariamente aos efeitos de melhoria da localização (redução do erro, maior exatidão) induzidos pela informação auditiva nas condições AV no caso do deslocamento M com alvos em movimento horizontal, a informação auditiva não contribuiu para a melhoria da localização vertical do alvo nas condições AV. Esta diferença pode relacionar-se com as discussões anteriores sobre a aplicabilidade de um modelo de integração por média no caso do deslocamento O , e da sua inadequação de princípio no caso do deslocamento M . A integração de informações sensoriais por média ponderada tem sido proposta como um mecanismo de natureza fundamentalmente sensorial-perceptiva (Alais & Burr, 2004; Ernst & Banks, 2002). Diferentemente, os mecanismos que propusemos como explicação para o melhoramento audiovisual registado no deslocamento M com movimentos horizontais, envolvem de modo crítico a seleção pela atenção endógena (*top-down*, ordenada a fins) das propriedades temporais do som e a sua utilização para o refinamento da localização espacial horizontal do alvo visual. Os mecanismos de integração sensorial invocados em Teramoto, Hidaka, et al. (2010) e Chien et al. (2013) revelaram-se incapazes, em si mesmos, de produzir melhoramento audiovisual, mesmo em condições de sincronia entre o *offset* dos movimentos visual e auditivo. No conjunto, é assim possível conjecturar que a diferença entre o deslocamento M e o deslocamento O quanto aos efeitos das condições audiovisuais

(benéficos, no primeiro caso, mas não no segundo) traduz uma diferença nos mecanismos subjacentes, de natureza sobretudo sensorial-percetiva no deslocamento ortogonal e percetivo-atencional (com componentes de atenção endógena) no deslocamento M horizontal.

As correlações entre o deslocamento O e o deslocamento M revelaram-se em todos os casos nulas (ver **Tabela 2** e **Tabela 3**), indicando uma independência estatística entre os dois tipos de erro. Este resultado é fornecido pela primeira vez para condições auditivas unimodais e condições audiovisuais e contrário ao único resultado por nós conhecido envolvendo condições visuais unimodais (Martins, 2015). Importa notar que, em si mesma, a ausência de associação estatística não implica a inexistência de mecanismos comuns, do mesmo modo que a emergência de correlações significativas não constituiria prova de mecanismos partilhados. Em todo o caso, alguns dos mecanismos propostos na literatura como explicações para os erros de localização de alvos em movimento implicariam relações de covariação entre os dois tipos de erro. Por exemplo, a hipótese de que a extrapolação da posição futura de um alvo móvel envolve a imaginária mental, assumindo adicionalmente que a velocidade de transformação mental é uma característica estável de cada participante (ver Munger, Solberg, Horrocks, & Preston, 1999), leva a prever uma correlação entre deslocamentos M e deslocamentos O. A ausência de correlação significativa entre as duas variáveis tem neste sentido algum valor seletivo relativamente a interpretações concorrentes do momento representacional, sendo mais desfavorável a algumas do que a outras. Em geral, combinada com indicações como as discutidas no parágrafo anterior, surge como mais consentânea com a admissão de mecanismos distintos para os deslocamentos M e O, do que o contrário.

Por fim, alguns dos resultados relativos ao deslocamento O suportam, de forma limitada, uma atribuição do efeito da variável distância nas condições AE_{is} (redução do erro vertical em d2) à maior duração do som e não tanto à extensão do trajeto percorrido. Um desses resultados é a ocorrência dos maiores erros verticais na tarefa E (ver **Figura 13** e **Figura 16**), que utiliza apenas sons síncronos, com duração idêntica à do movimento visual (embora percorram num terço das vezes extensões maiores do que o alvo visual), enquanto nas tarefas T e ET, associadas a valores de deslocamento O inferiores, se incluem sons (um terço de todos os apresentados) cuja duração excede a do movimento visual. Um segundo resultado, relacionado com o anterior, é a ocorrência, nas tarefas T e ET (mas não na tarefa E) dos menores erros médios de localização vertical nas condições com início no nível 1 de *onset* (som começa antes do movimento visual) e nas condições com termo no nível 3 do

fator *offset* (som termina mais tarde do que o movimento visual). São estes os níveis associados a uma maior duração do som nas condições audiovisuais (sons iniciados nos níveis 2 e 3 de *onset* têm em média menor duração do que os sons iniciados no nível 1; sons com termo nos níveis 1 e 2 de *offset* têm em média menor duração do que os concluídos em off 3: ver **Figura 1**). Estes resultados podem apenas tomar-se como sugestivos, mantendo-se a sua generalidade condicional a replicações posteriores do estudo. No entanto, a confirmarem-se em novos estudos, podem indicar que diferentes propriedades do som são tidas em conta no melhoramento audiovisual do deslocamento M (a sua redução baseada na localização temporal do *offset* do som) e na integração da informação visual e auditiva no contexto do deslocamento O (a duração total do som).

CAPÍTULO 7

DISCUSSÃO GERAL E CONCLUSÃO

7.1. PRINCIPAIS RESULTADOS

Nesta rubrica da discussão são enunciados e qualificados de forma resumida os principais resultados da dissertação, considerando em particular a resposta às questões identificadas como em aberto no ponto 3.3. do enquadramento.

(1) *O presente estudo documentou a possibilidade de efeitos consistentes de melhoramento audiovisual na localização do ponto de desaparecimento de um alvo visual em movimento.* Este resultado contrasta com os “enviesamentos de atração” (em direção à localização temporal ou espacial do som) reportados em estudos anteriores que utilizaram assincronias audiovisuais com magnitudes inferiores a 200 ms e em maior número do que as do presente estudo. Confirma assim a importância de aspetos contextuais da tarefa (e.g., saliência perceptiva e fiabilidade das assincronias AV) na determinação dos efeitos da informação auditiva na localização de alvos visuais dinâmicos. Um aspeto particular deste resultado envolve as condições de sincronia audiovisual (ausência de disparidades temporais entre os estímulos visual e auditivo) que, fornecendo um termo de comparação equivalente através dos diferentes estudos, apenas no nosso estudo produziram efeitos de melhoramento audiovisual (na realidade, os maiores valores observados de redução do

deslocamento M). Este ponto assinala o grau em que a relevância percebida da informação auditiva para a tarefa se sobrepõe à estrita coincidência ou proximidade temporal (regra temporal imposta por uma janela de integração AV) na produção de aperfeiçoamentos da execução (orientados para os objetivos da tarefa). Do ponto de vista da interpretação, estes resultados suportam uma conceção multi-mecanismo da combinação da informação sensorial, assente numa interação flexível, modulada pela tarefa, entre processamento multissensorial e atenção intermodal, com destaque, no caso dos efeitos de melhoramento AV, para um regime endógeno (orientado por objetivos) do envolvimento atencional.

(2) *Tanto as propriedades temporais como espaciais do som modularam significativamente a magnitude do deslocamento M (momento representacional)*. Trata-se, tanto quanto nos é dado conhecer, do primeiro resultado não ambíguo de um efeito da informação espacial veiculada pelo som na localização percebida de alvos visuais em movimento. Efeitos na localização espacial de alvos dinâmicos visuais mediados por pistas auditivas temporais foram anteriormente reportados (Chien et al., 2013; Teramoto, Hidaka, et al., 2010; Teramoto et al., 2018) mas nenhum resultado conclusivo sobre efeitos diretos de pistas auditivas espaciais (Kawachi, 2016; Teramoto et al., 2018). Por outro lado, efeitos diretos (não temporalmente mediados) de pistas espaciais acústicas na localização de estímulos visuais estáticos desfocados (submetidos a um tratamento de *blur*) foram anteriormente estabelecidos (Alais & Burr, 2004; Battaglia et al., 2003), mantendo-se em aberto se poderiam estender-se a estímulos em movimento.

Este resultado articula na realidade três aspetos. (i) *Uma indicação metodológica essencial*: o seu estabelecimento dependeu da capacidade de dissociar experimentalmente os efeitos da informação temporal e espacial do som, baseada numa combinação do desenho fatorial *onset* × *offset* com a covariação, nas condições espaciais (tarefa E), da magnitude das disparidades AV com a velocidade (e não a duração) do movimento auditivo. Esta estratégia possibilitou, pela primeira vez, uma dissociação controlada dos dois tipos de pistas, ultrapassando a confusão entre disparidades temporais e espaciais que se mantinha em estudos anteriores (e.g., Schmiedchen et al., 2012), comprometendo a interpretação dos resultados. (ii) *A ausência de constrangimentos impostos aos movimentos oculares permitiu estabelecer este resultado no contexto da visão central/foveal*, afastando dúvidas, inspiradas tanto pela hipótese da “adequação da modalidade” (Welch & Warren, 1980) como pela sua revisão à luz do Modelo da Máxima Verosimilhança (MMV) (Alais & Burr, 2004; Ernst & Banks, 2002), de que um eventual efeito da informação auditiva espacial pudesse ocorrer

fora da visão periférica, onde a capacidade de localização visual se encontra significativamente degradada. (iii) *A sugestão de que os efeitos específicos de melhoramento AV se devem à informação temporal e não à informação espacial veiculada pelo som.* Embora, por limitações do desenho, não seja possível excluir em definitivo efeitos de melhoria devidos à coincidência espacial da informação auditiva e visual, vários dados (e.g., o aumento relativo do erro de localização na tarefa ET e o perfil da interação entre os fator tarefa e o fator distância, associado a disparidades com diferentes magnitudes: ver discussão dos resultados da **Figura 6**) sugerem que o principal efeito das disparidades espaciais consistiu numa limitação do benefício das contingências temporais entre a visão e audição, em particular da sincronia audiovisual. Esta conclusão é compatível com evidência recente de que a sobreposição espacial (“regra espacial”) possa não constituir uma condição geral de combinação multissensorial e sim um fenômeno essencialmente dependente da tarefa (Spence, 2013). Em particular, poderia suportar uma variante da hipótese da “adequação da modalidade” na qual as propriedades espaciais do som, não sendo perceptivamente ignoradas em tarefas de localização visual, não seriam contudo consideradas (pelo menos em visão central) para o melhoramento AV dessa localização.

(3) *Papel respetivo das disparidades audiovisuais no começo e termo do movimento.* Um resultado robusto do estudo foi a ausência de efeitos das disparidades AV no início do movimento (efeitos estatisticamente nulos do fator *onset*). Este resultado é distinto do reportado por Teramoto, Hidaka, et al. (2010) para sons contínuos (como os do presente estudo), que indicou, ao contrário, a necessidade de sincronia AV no começo do movimento para que as disparidades AV no termo do movimento modulassem o momento representacional. Teramoto, Hidaka, et al. (2010) interpretaram a sua observação em termos da hipótese “suposição de unidade” (Vatakis & Spence, 2007; Warren et al., 1981; Welch & Warren, 1980, 1986; para uma revisão ver Y.-C. Chen & Spence, 2017). A sincronia das duas modalidades no *onset* constituiria uma fonte necessária da assunção de que os dois tipos de informação (auditiva e visual) têm origem no mesmo objeto, tornando possível a integração das informações assíncronas no *offset* e, desse modo, a modulação do deslocamento M pelo som. No entanto, os efeitos documentados por Teramoto, Hidaka, et al. (2010) correspondem a enviesamentos de localização, não a melhoramentos AV. Uma explicação possível para a diferença entre o nosso resultado e o encontrado por Teramoto, Hidaka, et al. (2010) poderia ser assim a de que a “suposição de unidade” não é já necessária para efeitos de melhoria AV. Em alternativa, podia suceder que a “suposição de unidade” se

mantivesse na realidade atuante no nosso estudo, mas baseada noutras fontes que não a associação das duas modalidades no *onset*: por exemplo, o facto de a resposta utilizada no nosso estudo ser de localização motora (diferentemente dos julgamentos perceptivos utilizados por Teramoto, Hidaka, et al., 2010) poderia suportar uma associação estável do som com fases do planeamento motor da localização do alvo visual, favorecendo por essa via uma “suposição de unidade”. Mais decisivo parece, contudo, ser o facto de os efeitos registados por Teramoto, Hidaka, et al. (2010) terem como condição necessária que as assincronias AV no termo do movimento se mantenham abaixo dos ± 160 ms, enquanto os efeitos verificados no nosso estudo se baseiam em assincronias ≥ 250 ms. A diferença de resultados quanto ao papel do *onset* constitui plausivelmente, neste contexto, uma expressão adicional da autonomia das contingências AV no nosso estudo (cujo efeito de melhoramento é primariamente “orientado para a tarefa”) relativamente aos constrangimentos impostos por janelas temporais estritas de integração multissensorial.

(4) *Semelhanças e diferenças entre as modalidades visuais e auditivas.* A comparação do momento representacional nas modalidades visual e auditiva (com sons espacializados) foi abordada anteriormente apenas num número reduzido de estudos e unicamente para o deslocamento M. Os resultados conhecidos apontam para um padrão geral de semelhança, com um deslocamento positivo no sentido do movimento também na modalidade auditiva (Getzmann, 2005a; Getzmann & Lewald, 2009; Getzmann et al., 2004; Mateeff & Hohnsbein, 1988; Schmiedchen et al., 2013), e para diferenças no efeito modulador de algumas variáveis, como a velocidade do alvo (efeitos crescentes no deslocamento M com alvos visuais, mas variáveis no caso de alvos auditivos: ver Feinkohl et al., 2014; Getzmann et al., 2004; Schmiedchen et al., 2013) ou as restrições impostas aos movimentos oculares (redução do erro para diante com alvos visuais, ausência de efeitos com alvos auditivos: ver Getzmann, 2005a; Getzmann & Lewald, 2007, 2009; Schmiedchen et al., 2012).

Os resultados obtidos no nosso estudo apontam na mesma direção, juntando novos elementos relativos ao deslocamento O (gravidade representacional) e a outras variáveis moduladoras não consideradas anteriormente no quadro da comparação entre modalidades, como a distância percorrida pelo alvo. Tanto no caso do deslocamento M como do deslocamento O, as duas modalidades apresentaram um comportamento paralelo, com deslocamentos horizontais positivos e verticais negativos. Por outro lado, os efeitos da distância percorrida mostraram ter sentidos contrários nas duas modalidades. No deslocamento M, o erro diminuiu com a distância para alvos visuais (mesmo em contextos

audiovisuais) e aumentou com a distância para alvos auditivos unimodais. No deslocamento O, o erro vertical aumentou com a distância para alvos visuais, mas diminuiu com a distância para alvos auditivos. A ideia de que tanto mecanismos comuns (eventualmente preditivos) como mecanismos específicos a cada modalidade poderão contribuir para os erros de localização de alvos dinâmicos na visão e na audição (Hubbard, 2014a; Schmiedchen et al., 2013) sai assim reforçada da presente dissertação.

(5) *Relações entre deslocamento M e deslocamento O em contextos unimodais visuais e auditivos e em contextos audiovisuais.* Nenhuma correlação significativa entre os deslocamentos M e O foi encontrada em qualquer das condições modais (visual, auditiva, AV). Embora se trate de uma simples medida de associação, que não permite excluir a existência de mecanismos comuns aos dois tipos de deslocamento, este resultado é, em qualquer caso, mais favorável ao envolvimento de diferentes mecanismos (ou diferentes componentes de um mecanismo diferenciado) no momento representacional horizontal e na gravidade representacional do que à hipótese de um só mecanismo comum. No caso particular das condições AV, esta indicação foi ainda suportada pela sugestão de que diferentes propriedades poderão ter determinado os efeitos do som nas localizações horizontal, por um lado, e vertical, por outro, do alvo visual: as diferenças temporais entre o termo do som e do movimento visual, no primeiro caso (deslocamento M), e a duração total do som no segundo (deslocamento O).

(6) *Limites dos modelos Bayesianos de integração ótima.* A redução do deslocamento M (melhoramento AV) observada nas tarefas E e ET não é compatível com uma integração por média ponderada da informação espacial auditiva e visual e sim com a utilização da informação temporal do som como forma de aperfeiçoar a localização temporal do alvo visual (resultando numa melhor localização espacial). Já no caso do deslocamento O, a combinação das informações espaciais veiculadas pela visão e pela audição através de uma média ponderada permanece em geral possível. Contudo, o esquema de ponderação afasta-se marcadamente da ponderação ótima pela fiabilidade (inverso da variabilidade) da informação sensorial, prevista pelo Modelo Bayesiano da Máxima Verosimilhança. Na medida em foram obtidos a partir dos 10 participantes que realizaram, para além da tarefa auditiva unimodal, a tarefa visual unimodal na ausência de contexto audiovisual, estes últimos resultados relativos ao deslocamento O devem no entanto considerar-se como sobretudo indicativos.

7.2. IMPLICAÇÕES DE NATUREZA APLICADA

Os resultados encontrados contêm algumas implicações de natureza aplicada. No que se refere ao deslocamento M, fornecem indicações essencialmente paralelas, obtidas com sujeitos normovisuais e no domínio da localização de estímulos visuais dinâmicos, às obtidas por Targher et al. (2017) em participantes com visão subnormal (redução da acuidade ou do campo visual sem possibilidade de correção por cirurgia, lentes ou fármacos) no domínio da detecção visual. Partindo da evidência acumulada sobre melhorias na detecção de estímulos visuais (e.g., luminância, contraste) induzidas por um som temporal e espacialmente alinhado em participantes normovisuais (Frassinetti et al., 2002; Hairston, Wallace, et al., 2003), com défices visuais associados a lesão cerebral (Frassinetti, Bolognini, Bottari, Bonora, & Làdavas, 2005) e com visão subnormal, Targher et al. (2017) investigaram as propriedades temporais do melhoramento audiovisual (i.e., o efeito de diferentes regimes de assincronia AV) numa amostra de participantes com visão deficitária (ver também Targher, Occelli, & Zampini, 2012). A conclusão foi a de que o melhoramento era largamente modulado por mecanismos atencionais “de-cima-para-baixo”. Na discussão final, os autores sublinham que essa conclusão abre caminho a estratégias AV de reabilitação da visão menos constrangidas pela exigência de sobreposição (coincidência espacial) e simultaneidade (coincidência temporal) dos estímulos A e V do que a literatura inicial levaria a supor.

De modo semelhante, a interpretação do conjunto dos resultados relativos ao deslocamento M no nosso estudo aponta para um papel fundamental de mecanismos atencionais, dependentes da estrutura da tarefa, nos efeitos de melhoramento audiovisual. O papel destes mecanismos vai mesmo mais longe do que o proposto em Targher et al. (2017), para quem o efeito atencional consiste em flexibilizar os limites da janela temporal de integração audiovisual (ampliando o intervalo no qual é possível integrar os dois tipos de informação). No caso do nosso estudo, a detecção e aprendizagem de associações intermodais (AV) de alto nível, independentes em princípio de uma janela de integração (Fiebelkorn et al., 2010), parece desempenhar um papel autónomo na mobilização de uma atenção intermodal endógena, capaz de suportar efeitos de melhoramento mesmo com intervalos temporais de ± 500 ms. Assincronias AV desta magnitude (-500 ms) foram já utilizados em programas de reabilitação da visão de pacientes com escotomas hemianópsicos que resultaram em benefícios suportados por melhores estratégias de exploração atencional do campo visual e

melhor planeamento oculomotor (Bolognini, Rasi, Coccia, & Làdavas, 2005; Tinelli, Purpura, & Cioni, 2015; ver também Passamonti, Bertini, & Làdavas, 2009).

O paralelismo, documentado na literatura sobre o melhoramento AV da deteção visual, entre os resultados obtidos em amostras de sujeitos com défices de visão e normovisuais torna assim plausível que a redução do erro de localização no sentido do movimento (deslocamento M) observado no nosso estudo possa igualmente verificar-se em amostras de participantes com limitações da visão. Condicionalmente a essa verificação e à sua generalidade, a utilização de estratégias de reabilitação AV para défices específicos da perceção visual do movimento e, em particular, da localização motora/interseção de alvos móveis, constitui uma linha de investigação aplicada concebível. Uma indicação específica relevante neste quadro, procedente dos nossos resultados, é a de que é sobretudo a informação temporal veiculada pelo som, não a sua informação espacial, que suporta os efeitos de redução do erro no sentido do movimento. Isto é consistente com a ideia de que a chamada “regra espacial” (Stein, 1998), segundo a qual a combinação da informação de diferentes modalidades é facilitada pela sua apresentação no mesmo local (informação espacial coincidente), poderá não constituir uma condição geral de combinação intermodal e sim um fenómeno dependente da natureza da tarefa (Spence, 2013). Adicionalmente, a verificação de que na tarefa E as condições com disparidades espaciais entre o termo dos movimentos visual e auditivo aumentaram o valor do erro comparativamente às condições sem disparidade, indica que diferenças espaciais obstaculizam o melhoramento baseado na informação temporal do som. Evitar disparidades espaciais em programas AV de reabilitação da localização visualmente guiada de alvos móveis poderá assim justificar-se não em virtude da “regra espacial” (que suporia um melhoramento baseado na coincidência espacial), mas como forma de otimizar o impacto da informação temporal veiculada pelo som na melhoria da localização. Uma segunda indicação específica diz respeito à ausência de melhoramento AV no erro de localização vertical (deslocamento O). A localização visual de alvos móveis em altura, quando o seu movimento é horizontal, não parece assim beneficiar do recurso a uma estratégia AV. No entanto, dado que o estudo apenas envolveu alvos em movimento horizontal, a questão de saber se este resultado se mantém para movimentos verticais e oblíquos permanece sem resposta.

Um debate prolongado no domínio da reabilitação da visão envolve a distinção entre abordagens compensatórias e restaurativas (Dundon, Bertini, Làdavas, Sabel, & Gall, 2015). As primeiras visam tirar partido de mecanismos não comprometidos e suscetíveis de compensar as perdas de campo visual, assentando frequentemente na aprendizagem de

estratégias oculomotoras de busca visual (ver, e.g., Roth et al., 2009). As segundas visam recuperar diretamente as perdas de visão por mobilização de funções visuais residuais, insistindo, por exemplo, no treino da detecção e discriminação de luminâncias (Sabel, Henrich-Noack, Fedorov, & Gall, 2011). O tipo de programa de reabilitação que os resultados desta dissertação poderiam suportar inscreve-se primariamente no quadro das abordagens compensatórias, como o *Audio-Visual Scanning Training (AViST)*, de que os trabalhos já citados de Bolognini, Rasi, et al. (2005) e Tinelli et al. (2015) são exemplo. No entanto, como notam Dundon et al. (2015), os conceitos de compensação e restauração não são mutuamente exclusivos, admitindo graus de sobreposição variáveis. Por exemplo, a sugestão de que os efeitos perduráveis e transferíveis de abordagens compensatórias como a de Bolognini, Russo, & Vallar (2015) possam envolver alterações da plasticidade cerebral e a promoção, mediada por estruturas como o colículo superior (cujo papel na integração audiovisual é desde há muito sublinhado: Stein & Meredith, 1993), de funções visuais residuais da visão, tem sido avançada (Grasso, Benassi, Làdavas, & Bertini, 2016; Keller & Lefin-Rank, 2010). Considerando que a magnitude das assincronias temporais utilizadas no nosso estudo é largamente superior àquelas tipicamente associadas ao mecanismo da integração sensorial (± 100 ms), não é claro que as mesmas estruturas e processos neurais classicamente apontadas como suporte da integração sensorial sustentem igualmente os efeitos de melhoramento AV que documentámos. O grau em que abordagens de reabilitação inspiradas pelos resultados da presente dissertação poderão ou não mobilizar, para além duma vertente compensatória, eventuais efeitos restaurativos, constitui por conseguinte uma questão em aberto, cujo esclarecimento depende inteiramente de investigação adicional, designadamente no âmbito da neuropsicologia e da imagiologia cerebral.

As potenciais implicações para a reabilitação compensatória de aspetos funcionais da ação visualmente guiada não se dirigem apenas a distintos quadros clínicos, mas igualmente às perdas associadas ao envelhecimento. A investigação do processamento multissensorial no contexto do envelhecimento tem revelado resultados mistos (Mozolic, Hugenschmidt, Peiffer, & Laurienti, 2012). Enquanto alguns estudos iniciais envolvendo a comparação de adultos jovens e idosos em tarefas cognitivas complexas (e.g., memória de eventos sensoriais) sugeriram maiores benefícios multissensoriais em participantes mais novos (e.g., Stine, Wingfield, & Myers, 1990), o balanço da evidência em estudos mais recentes e assentes em tarefas mais simples (e.g., tarefas de discriminação audiovisual) aponta ao contrário para um benefício multissensorial superior em participantes mais idosos (Diederich, Colonius, & Schomburg, 2008; Laurienti, Burdette, Maldjian, & Wallace, 2006; Peiffer, Mozolic,

Hugenschmidt, & Laurienti, 2007). O princípio da eficácia inversa (Tye-Murray et al., 2010) segundo o qual a integração multissensorial é tanto mais provável ou acentuada quanto maior a degradação da informação unimodal (como sucede, em geral, no curso do processo do envelhecimento, com a elevação dos limiares sensoriais), levaria na realidade a esperar uma estratégia de integração multissensorial mais marcada em participantes mais idosos. No entanto, outros fatores podem igualmente contribuir para esse resultado, entre os quais alterações nos parâmetros temporais da janela integração, como o aumento da sua duração em participantes mais idosos (e.g., Mozolic et al., 2012). Neste último cenário, a utilização de assincronias AV mais longas (como as que utilizámos no presente estudo) poderia vir a revelar-se particularmente adequada a estratégias de compensação multissensorial em populações mais idosas.

À margem dos objetivos de reabilitação da visão, o treino da acuidade e precisão de respostas de localização visualmente guiadas, em tarefas nas quais estas respostas constituam uma dimensão fundamental da eficiência da execução, pode igualmente beneficiar dos resultados obtidos nos nossos estudos. Uma área para a qual podem, em geral conjecturar-se potenciais implicações é a da ergonomia percetiva – em particular no domínio do desenvolvimento de *feedback* multissensorial (designadamente audiovisual) e da expansão intermodal da interação homem-computador (ver, e.g., Seinfeld, Feuchtner, Maselli, & Müller, 2020).

Por fim, vários dos resultados produzidos no âmbito da dissertação não foram nunca reportados, de acordo com o nosso melhor conhecimento. Entre estes contam-se os resultados relativos aos efeitos das variáveis distância percorrida e ponto de desaparecimento na localização percebida de alvos unimodais auditivos (A) em movimento horizontal – avaliados tanto na direção do movimento (momento representacional) como na direção ortogonal ao movimento (gravidade representacional). Este conjunto de resultados alarga a base de evidência disponível sobre a localização auditiva de estímulos dinâmicos. Tendo sido obtidos numa amostra limitada de 31 participantes, requerem replicações adicionais como forma de estabelecer a sua generalidade. No entanto, a verificar-se essa generalidade, uma via com potencial interesse aplicado consistirá em examinar o grau em que um ou mais destes novos efeitos dispõem de valor diagnóstico para problemas da localização de fontes sonoras em movimento (e.g., condições clínicas nos quais os efeitos estabelecidos não se verifiquem, invertam a sua direção ou apresentem desvios de magnitude significativos).

Em geral, aceitando a interpretação de uma melhoria audiovisual em tarefas de localização visual como resultado de um refinamento induzido pelo som da localização no

tempo do alvo visual, défices do processamento dos aspetos temporais da audição deveriam impedir ou limitar drasticamente estes efeitos de melhoria (Alais & Burr, 2019). O grau em que um paradigma simplificado de MR audiovisual pode, neste sentido, vir a acrescentar-se utilmente às ferramentas de diagnóstico e treino auditivo de perturbações do processamento auditivo temporal merece neste sentido ser futuramente investigado, bem como o seu eventual interesse no estudo do processamento audiovisual/multissensorial em contextos de amplificação auditiva (e.g., aparelhos auditivos) ou da (re)habilitação de utilizadores de implantes cocleares (Wallace et al., 2020: ver também, e.g., Butera et al., 2018; Isaiah, Vongpaisal, King, & Hartley, 2014).

7.3. LIMITES DO ESTUDO

Podem naturalmente apontar-se ao presente estudo alguns limites, considerados em seguida. Um primeiro tipo de limites envolve as escolhas operacionais relativas ao modo de apresentação (por auscultadores) dos sons e aos procedimentos utilizados na sua espacialização. A apresentação de sons por meio de auscultadores, impedindo designadamente as funções de filtragem acústica do pavilhão auricular (*pinnae*), específicas à morfologia do ouvinte e tidas como importantes na apreciação da distância de fontes sonoras (Baumgartner et al., 2017), dificulta reconhecidamente a externalização do som (Best, Baumgartner, Lavandier, Majdak, & Kopčo, 2020; Leclère, Lavandier, & Perrin, 2019). O modo de apresentação poderá assim ter induzido uma separação potencial entre o movimento do som percebido “no interior” da cabeça (eixo interaural) e o movimento do alvo visual observado no plano do ecrã, dificultando a integração audiovisual. A capacidade de externalizar o som apresentado por auscultadores varia de forma importante com os indivíduos e pode, em princípio, melhorar com o treino numa situação específica (Mendonça et al., 2012). Embora os participantes do estudo não tenham sido sistematicamente questionados sobre o grau em que o som lhes pareceu ter uma origem “externa”, foram sondados informalmente no final de cada tarefa sobre o grau de dificuldade que experimentaram na execução. Mesmo no caso da tarefa auxiliar com estímulos auditivos unimodais, apenas uma minoria (4 em 31 participantes) consideraram a tarefa como envolvendo alguma dificuldade. Isto parece indicar que a maioria dos participantes não experienciou a instrução de localizar externamente os sons no ecrã como desprovida de sentido ou especialmente difícil de cumprir. Por outro lado, o facto de esta tarefa auxiliar ter

produzido resultados similares aos de estudos do momento representacional auditivo em campo livre – baseados em arranjos densos de colunas ativadas consecutivamente para produzir o movimento da fonte sonora (e.g., Getzmann & Lewald, 2009) – sugere, de forma convergente, que o uso de auscultadores não terá representado um constrangimento essencial.

A banda de frequências utilizadas, centrada nos 2 kHz, corresponde por outro lado a uma região frequentemente considerada como vulnerável a ambiguidades na codificação de diferenças interaurais (Akeroyd, 2014; Middlebrooks & Green, 1991). De acordo com a teoria duplex da localização sonora (Balkany & Zeitler, 2013), as pistas de localização primárias nas baixas frequências (< 1.5 kHz) são as diferenças interaurais de tempo (ITD) com um ótimo de eficiência em torno dos 250 Hz. Acima dos 1.5 kHz, a localização dependeria primariamente de diferenças interaurais de intensidade (ILD), com um ótimo de eficiência em torno dos 4 kHz. Este quadro de referência explica a opção de alguns dos estudos anteriores do momento representacional auditivo (e.g., Schmiedchen et al., 2013) pela utilização de duas gamas de frequência, baixa (centrada nos 500 Hz) e alta (centrada nos 4 kHz), como forma de explorar eventuais diferenças devidas ao tipo de pista interaural. Aplicado ao presente estudo, poderia motivar a suspeita de uma insuficiente precisão da lateralização do som através das informações acústicas fornecidas. Na realidade, diferentemente da localização de estímulos auditivos tonais ou de banda estreita (nos quais assenta a clássica teoria duplex), a precisão da localização de ruído filtrado de banda larga, abarcando duas oitavas ou mais – como sucede no nosso estudo –, não depende da frequência central, pelo menos em normo-ouvintes (Risoud et al., 2018; Yost et al., 2013; Yost & Zhong, 2014). De modo concordante com esta indicação, os resultados obtidos na tarefa auxiliar com estímulos auditivos unimodais (ver **Figura 10** e comentário) suportam uma lateralização adequada do som e, desse modo, a eficiência informativa das pistas binaurais fornecidas.

Por fim, a utilização de uma função de transferência relacionada com a cabeça (HRTF) média, e não individualizada, poderá ter contribuído para degradar tanto a precisão como o valor ecológico da informação espacial veiculada pelo som. A vantagem da aplicação de HRTFs individualizadas ao sinal auditivo como forma de promover o realismo e exatidão da localização em espaços acústicos virtuais é conhecida (Jenny & Reuter, 2020; S.-M. Kim & Choi, 2005). No entanto, a dificuldade de implementar esta estratégia cresce com o número de participantes, tornando-a, em muitas circunstâncias, pouco praticável. Não podendo excluir que o uso de uma HRTF genérica tenha afetado negativamente as tarefas, a

semelhança (inclusive quanto às magnitudes de deslocamento M) entre os resultados da localização dos estímulos auditivos no presente estudo e no estudo de Feinkohl et al. (2014), no qual foram utilizadas HRTFs individualizadas, sugerem que, a ter existido, o impacto dessa limitação foi reduzido.

Um segundo tipo de limites envolve a reduzida dimensão da amostra de participantes ($N = 31$). Apesar do essencial das análises assentar em ANOVAs de medidas repetidas, que oferecem maior poder estatístico do que os desenhos entre-sujeitos, é possível que a extensão da amostra tenha limitado a deteção de diferenças significativas em algumas comparações. Esta limitação é particularmente óbvia no caso das comparações envolvendo apenas o subgrupo de participantes ($N = 10$) que realizou, para além da tarefa de localização de alvos auditivos, a tarefa de localização de alvos visuais na ausência de contexto audiovisual (condições visuais isoladas: V_{is}). Todos os resultados relativos ao contraste entre condições V_{is} e as restantes condições (auditivas, audiovisuais e visuais intercaladas) devem neste sentido tomar-se como sobretudo indicativos e particularmente dependentes de replicações posteriores. A extensão da amostra inviabilizou também na prática a exploração de diferenças interindividuais nos padrões de resposta. Em particular, condicionou a possibilidade do recurso a técnicas como a análise de *clusters*, com o objetivo de investigar a existência de subgrupos heterogéneos. A questão das diferenças individuais, e da possibilidade da sua organização através de *clusters* de participantes, mantém-se assim como uma importante questão em aberto e um objeto natural de futuros estudos.

Um terceiro tipo de limites diz respeito a variáveis potencialmente relevantes para os resultados que não foram incluídas como fatores no desenho experimental. Uma delas foi a direção do movimento, que ocorreu sempre da esquerda para a direita. A principal razão para a opção de não contrabalancear a direção do movimento foi a manutenção do número total de ensaios em limites praticáveis. Considerando que o desenho de medidas repetidas implica que todos os participantes passem por todas as condições experimentais em todas as tarefas, utilizar as duas direções resultaria na duplicação do número de condições por participante. Em todo o caso, não parecem existir razões *a priori* para crer que um padrão distinto de resultados ocorresse com uma alteração da direção do movimento. No contexto do momento representacional visual, não existe evidência consistente sobre um efeito sistemático da direção de movimentos horizontais (para uma revisão, ver Hubbard, 2014a). O mesmo sucede no contexto do momento representacional auditivo. Schmiedchen et al. (2013) reportaram apenas uma interação do efeito da velocidade do movimento com a sua direção, e Feinkohl et al. (2014) não encontraram diferenças devidas à direção do

movimento. Em estudos comparativos do MR visual e auditivo, Schmiedchen et al. (2012) não observaram efeitos (principais ou interações) da direção do movimento em qualquer das modalidades. No caso de sons não espacializados, Chien et al. (2013) e Teramoto, Hidaka, et al. (2010) não documentaram também efeitos de modulação do MR atribuíveis à direção do movimento visual. Apesar destas indicações, excluir em definitivo um efeito do contrabalanceamento da direção do movimento no caso concreto das nossas experiências só será possível em investigação posterior.

7.4. CONCLUSÃO PROVISÓRIA

Como qualquer trabalho de investigação, esta dissertação corresponde a um momento provisório na produção cumulativa de conhecimento e na construção de um quadro explicativo dos fenómenos que toma como objeto. Os resultados e explicações sugeridas articulam dois domínios cuja aproximação tem vindo nos últimos anos a revelar-se profícua: o dos erros de localização do movimento, em particular no quadro de tarefas de momento representacional, e o do processamento intermodal, particularmente audiovisual. As conclusões suportadas evidenciam uma estreita conexão entre processamento multissensorial, atenção intermodal e apreensão perceptiva de eventos externos, flexivelmente modulada por características da tarefa (em particular o número e fiabilidade das contingências audiovisuais).

As implicações para a compreensão do momento representacional incluem (1) o reforço do papel de componentes atencionais endógenas na modulação do erro de localização (assentes, na presente situação, na utilização das propriedades temporais do som para a produção de benefícios AV na localização), apontando para a sua penetrabilidade cognitiva, e (2) a contribuição plausível tanto de mecanismos comuns, supramodais, como de mecanismos específicos a cada modalidade para os erros de localização de alvos dinâmicos na visão e na audição. As implicações para o processamento multissensorial sugerem (1) a vantagem de explorar sistematicamente o contexto e requisitos da tarefa como variáveis independentes na investigação dos mecanismos intermodais e (2) a capacidade de uma conceção multi-mecanismos do processamento multissensorial, incluindo a modulação pela tarefa, para reconciliar resultados aparentemente contraditórios, como os “enviesamentos de atração” (documentados pela literatura anterior) e o “melhoramento AV” (documentado para o deslocamento M no presente trabalho). As principais implicações práticas resultantes

envolvem a reabilitação AV de défices visuais (particularmente da localização visualmente guiada) e de perdas associadas ao processo normal do envelhecimento. Convergindo com resultados anteriores, a programação de contingências AV envolvendo intervalos temporais superiores aos que caracterizam as típicas janelas de integração multissensorial e mobilizando formas de atenção intermodal, mostraram suportar eficazmente benefícios AV na localização de alvos visuais dinâmicos.

Em todas as tarefas com apresentações AV a visão constituiu a modalidade diretamente relevante para a tarefa (tratou-se de localizar a desapareção de alvos visuais) e a audição a modalidade (formalmente) irrelevante. Os resultados obtidos não informam assim sobre a situação inversa (a audição como modalidade diretamente relevante e a visão como modalidade irrelevante). Mais geralmente, o grau em que se mantêm válidos para outras formas de emparelhamentos intersensorial (e.g., tato e audição) permanece também desconhecido. Ambas as questões e particularmente a segunda, pela sua generalidade, constituem direções futuras de investigação das quais depende a ampliação da base de evidência e uma compreensão alargada dos fenómenos aqui abordados.

REFERÊNCIAS BIBLIOGRÁFICAS

- Actis-Grosso, R., Bastianelli, A., & Stucchi, N. (2008). Direction of perceptual displacement of a moving target's starting and vanishing points: The key role of velocity. *Japanese Psychological Research*, 50(4), 253–263. <https://doi.org/10.1111/j.1468-5884.2008.00381.x>
- Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the “light-from-above” prior. *Nature Neuroscience*, 7(10), 1057–1058. <https://doi.org/10.1038/nn1312>
- Akeroyd, M. A. (2014). An overview of the major phenomena of the localization of sound sources by normal-hearing, hearing-impaired, and aided listeners. *Trends in Hearing*, 18. <https://doi.org/10.1177/2331216514560442>
- Alais, D., & Burr, D. (2003). The “Flash-Lag” effect occurs in audition and cross-modally. *Current Biology*, 13(1), 59–63. [https://doi.org/10.1016/S0960-9822\(02\)01402-1](https://doi.org/10.1016/S0960-9822(02)01402-1)
- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, 14(3), 257–262. <https://doi.org/10.1016/j.cub.2004.01.029>
- Alais, D., & Burr, D. (2019). Cue combination within a Bayesian framework. In A. K. C. Lee, M. T. Wallace, A. Coffin, A. Popper, & R. Fay (Eds.), *Multisensory Processes. Springer Handbook of Auditory Research* (pp. 9–31). Springer, Cham. https://doi.org/10.1007/978-3-030-10461-0_2
- Alais, D., Newell, F., & Mamassian, P. (2010). Multisensory processing in review: From physiology to behaviour. *Seeing and Perceiving*, 23(1), 3–38. <https://doi.org/10.1163/187847510X488603>
- Albright, T. D., & Stoner, G. R. (1995). Visual motion perception. *Proceedings of the National Academy of Sciences*, 92(7), 2433–2440. <https://doi.org/10.1073/pnas.92.7.2433>
- Alsius, A., Navarra, J., & Soto-Faraco, S. (2007). Attention to touch weakens audiovisual

- speech integration. *Experimental Brain Research*, 183(3), 399–404.
<https://doi.org/10.1007/s00221-007-1110-1>
- Amorim, M. A., Siegler, I. A., Baurès, R., & Oliveira, A. M. (2015). The embodied dynamics of perceptual causality: A slippery slope? *Frontiers in Psychology*, 6(APR), 1–11.
<https://doi.org/10.3389/fpsyg.2015.00483>
- Angelaki, D. E., Gu, Y., & DeAngelis, G. C. (2009). Multisensory integration: Psychophysics, neurophysiology, and computation. *Current Opinion in Neurobiology*, 19(4), 452–458.
<https://doi.org/10.1016/j.conb.2009.06.008>
- Arnold, D. H., Petrie, K., Murray, C., & Johnston, A. (2019). Suboptimal human multisensory cue combination. *Scientific Reports*, 9(1), 5155. <https://doi.org/10.1038/s41598-018-37888-7>
- Arrighi, R., Marini, F., & Burr, D. (2009). Meaningful auditory information enhances perception of visual biological motion. *Journal of Vision*, 9(4), 25–25.
<https://doi.org/10.1167/9.4.25>
- Ashida, H. (2004). Action-specific extrapolation of target motion in human visual system. *Neuropsychologia*, 42(11), 1515–1524.
<https://doi.org/10.1016/j.neuropsychologia.2004.03.003>
- Atkins, J. E., Fiser, J., & Jacobs, R. A. (2001). Experience-dependent visual cue integration based on consistencies between visual and haptic percepts. *Vision Research*, 41(4), 449–461. [https://doi.org/10.1016/S0042-6989\(00\)00254-6](https://doi.org/10.1016/S0042-6989(00)00254-6)
- Atkins, J. E., Jacobs, R. A., & Knill, D. C. (2003). Experience-dependent visual cue recalibration based on discrepancies between visual and haptic percepts. *Vision Research*, 43(25), 2603–2613. [https://doi.org/10.1016/S0042-6989\(03\)00470-X](https://doi.org/10.1016/S0042-6989(03)00470-X)
- Attneave, F., & Olson, R. K. (1971). Pitch as a medium: A new approach to psychophysical scaling. *The American Journal of Psychology*, 84(2), 147.
<https://doi.org/10.2307/1421351>
- Audacity Team. (2012). Audacity (version 2.0.2) [Computer software]. Retrieved from <https://www.audacityteam.org/>
- Baldo, M. V. C., & Klein, S. A. (2010). Paying attention to the flash-lag effect. In R. Nijhawan & B. Khurana (Eds.), *Space and Time in Perception and Action* (pp. 396–407). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511750540.023>
- Balkany, T. J., & Zeitler, D. M. (2013). Duplex theory of sound localization. In *Encyclopedia of Otolaryngology, Head and Neck Surgery* (pp. 718–718). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-23499-6_200036

- Barnes, G. R., & Marsden, J. F. (2002). Anticipatory control of hand and eye movements in humans during oculo-manual tracking. *The Journal of Physiology*, 539(1), 317–330. <https://doi.org/10.1113/jphysiol.2001.012979>
- Barutchu, A., Crewther, D. P., & Crewther, S. G. (2009). The race that precedes coactivation: Development of multisensory facilitation in children. *Developmental Science*, 12(3), 464–473. <https://doi.org/10.1111/j.1467-7687.2008.00782.x>
- Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A*, 20(7), 1391–1397. <https://doi.org/10.1364/JOSAA.20.001391>
- Baumgartner, R., Reed, D. K., Tóth, B., Best, V., Majdak, P., Colburn, H. S., & Shinn-Cunningham, B. (2017). Asymmetries in behavioral and neural responses to spectral cues demonstrate the generality of auditory looming bias. *Proceedings of the National Academy of Sciences of the United States of America*, 114(36), 9743–9748. <https://doi.org/10.1073/pnas.1703247114>
- Beierholm, U. R., Quartz, S. R., & Shams, L. (2009). Bayesian priors are encoded independently from likelihoods in human multisensory perception. *Journal of Vision*, 9(5), 23.1-9. <https://doi.org/10.1167/9.5.23>
- Bertamini, M. (2002). Representational momentum, internalized dynamics, and perceptual adaptation. *Visual Cognition*, 9(1–2), 195–216. <https://doi.org/10.1080/13506280143000395>
- Bertelson, P. (1999). Ventriloquism: A case of crossmodal perceptual grouping. In G. Aschersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 347–362). North-Holland/Elsevier Science Publishers. [https://doi.org/10.1016/S0166-4115\(99\)80034-X](https://doi.org/10.1016/S0166-4115(99)80034-X)
- Bertelson, P., & Aschersleben, G. (2003). Temporal ventriloquism: crossmodal interaction on the time dimension: 1. Evidence from auditory-visual temporal order judgment. *International Journal of Psychophysiology*, 50(1–2), 147–155. [https://doi.org/10.1016/S0167-8760\(03\)00130-2](https://doi.org/10.1016/S0167-8760(03)00130-2)
- Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & Psychophysics*, 29(6), 578–584. <https://doi.org/10.3758/BF03207374>
- Bertelson, P., Vroomen, J., de Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics*, 62(2), 321–332. <https://doi.org/10.3758/BF03205552>

- Best, V., Baumgartner, R., Lavandier, M., Majdak, P., & Kopčo, N. (2020). Sound externalization: A review of recent research. *Trends in Hearing, 24*, 233121652094839. <https://doi.org/10.1177/2331216520948390>
- Bizley, J. K., Jones, G. P., & Town, S. M. (2016). Where are multisensory signals combined for perceptual decision-making? *Current Opinion in Neurobiology, 40*, 31–37. <https://doi.org/10.1016/j.conb.2016.06.003>
- Blake, R., & Lee, S. H. (2005). The role of temporal structure in human vision. *Behavioral and Cognitive Neuroscience Reviews, 4*(1), 21–42. <https://doi.org/10.1177/1534582305276839>
- Blättler, C., Ferrari, V., Didierjean, A., & Marmèche, E. (2011). Representational momentum in aviation. *Journal of Experimental Psychology: Human Perception and Performance, 37*(5), 1569–1577. <https://doi.org/10.1037/a0023512>
- Blättler, C., Ferrari, V., Didierjean, A., van Elslande, P., & Marmèche, E. (2010). Can expertise modulate representational momentum? *Visual Cognition, 18*(9), 1253–1273. <https://doi.org/10.1080/13506281003737119>
- Blurton, S. P., Greenlee, M. W., & Gondan, M. (2015). Cross-modal cueing in audiovisual spatial attention. *Attention, Perception, & Psychophysics, 77*(7), 2356–2376. <https://doi.org/10.3758/s13414-015-0920-7>
- Boersma, P., & Weenink, D. (2013). Praat (version 5.3.60) [Computer software]. Retrieved from <http://www.fon.hum.uva.nl/praat/>
- Bolognini, N., Frassinetti, F., Serino, A., & Làdavvas, E. (2005). “Acoustical vision” of below threshold stimuli: Interaction among spatially converging audiovisual inputs. *Experimental Brain Research, 160*(3), 273–282. <https://doi.org/10.1007/s00221-004-2005-z>
- Bolognini, N., Rasi, F., Coccia, M., & Làdavvas, E. (2005). Visual search improvement in hemianopic patients after audio-visual stimulation. *Brain, 128*(12), 2830–2842. <https://doi.org/10.1093/brain/awh656>
- Bolognini, N., Russo, C., & Vallar, G. (2015). Crossmodal illusions in neurorehabilitation. *Frontiers in Behavioral Neuroscience, 9*(August), 1–6. <https://doi.org/10.3389/fnbeh.2015.00212>
- Bolognini, N., Senna, I., Maravita, A., Pascual-Leone, A., & Merabet, L. B. (2010). Auditory enhancement of visual phosphene perception: The effect of temporal and spatial factors and of stimulus intensity. *Neuroscience Letters, 477*(3), 109–114. <https://doi.org/10.1016/j.neulet.2010.04.044>

- Borst, A. (2000). Models of motion detection. *Nature Neuroscience*, 3(S11), 1168–1168. <https://doi.org/10.1038/81435>
- Bregman, A. S. (Ed.). (1990). *Auditory scene analysis: The perceptual organization of sound*. The MIT Press.
- Brehaut, J. C., & Tipper, S. P. (1996). Representational momentum and memory for luminance. *Journal of Experimental Psychology: Human Perception and Performance*, 22(2), 480–501. <https://doi.org/10.1037/0096-1523.22.2.480>
- Bresciani, J.-P., Dammeier, F., & Ernst, M. O. (2006). Vision and touch are automatically integrated for the perception of sequences of events. *Journal of Vision*, 6(5), 2. <https://doi.org/10.1167/6.5.2>
- Bresciani, J.-P., Ernst, M. O., Drewing, K., Bouyer, G., Maury, V., & Kheddar, A. (2005). Feeling what you hear: auditory signals can modulate tactile tap perception. *Experimental Brain Research*, 162(2), 172–180. <https://doi.org/10.1007/s00221-004-2128-2>
- Brouwer, A.-M., Franz, V. H., & Thornton, I. M. (2004). Representational momentum in perception and grasping: Translating versus transforming objects. *Journal of Vision*, 4(7), 5. <https://doi.org/10.1167/4.7.5>
- Brown, S. W. (1995). Time, change, and motion: The effects of stimulus movement on temporal perception. *Perception & Psychophysics*, 57(1), 105–116. <https://doi.org/10.3758/BF03211853>
- Bruns, P., & Getzmann, S. (2008). Audiovisual influences on the perception of visual apparent motion: Exploring the effect of a single sound. *Acta Psychologica*, 129(2), 273–283. <https://doi.org/10.1016/j.actpsy.2008.08.002>
- Burr, D., & Alais, D. (2006). Combining visual and auditory information. *Progress in Brain Research*, 155, 243–258. [https://doi.org/10.1016/S0079-6123\(06\)55014-9](https://doi.org/10.1016/S0079-6123(06)55014-9)
- Burton, H., Snyder, A. Z., Diamond, J. B., & Raichle, M. E. (2002). Adaptive changes in early and late blind: A fMRI study of verb generation to heard nouns. *Journal of Neurophysiology*, 88(6), 3359–3371. <https://doi.org/10.1152/jn.00129.2002>
- Busse, L., Roberts, K. C., Crist, R. E., Weissman, D. H., & Woldorff, M. G. (2005). The spread of attention across modalities and space in a multisensory object. *Proceedings of the National Academy of Sciences of the United States of America*, 102(51), 18751–18756. <https://doi.org/10.1073/pnas.0507704102>
- Butera, I. M., Stevenson, R. A., Mangus, B. D., Woynaroski, T. G., Gifford, R. H., & Wallace, M. T. (2018). Audiovisual temporal processing in postlingually deafened adults with cochlear implants. *Scientific Reports*, 8(1), 1–12. <https://doi.org/10.1038/s41598-018-29598-x>

- Calvert, G. A., Spence, C., & Stein, B. E. (Eds.). (2004). *The Handbook of Multisensory Processes*. *The Handbook of Multisensory Processes*. Cambridge, MA, US: MIT Press.
- Calvert, G. A., & Thesen, T. (2004). Multisensory integration: Methodological approaches and emerging principles in the human brain. *Journal of Physiology Paris*, *98*(1-3 SPEC. ISS.), 191–205. <https://doi.org/10.1016/j.jphysparis.2004.03.018>
- Carlile, S., & Leung, J. (2016). The perception of auditory motion. *Trends in Hearing*, *20*, 233121651664425. <https://doi.org/10.1177/2331216516644254>
- Cecere, R., Gross, J., & Thut, G. (2016). Behavioural evidence for separate mechanisms of audiovisual temporal binding as a function of leading sensory modality. *European Journal of Neuroscience*, *43*(12), 1561–1568. <https://doi.org/10.1111/ejn.13242>
- Cedrus Corporation. (2008). SuperLab (version 4.0.7b) [Computer software]. San Pedro, California: Cedrus Corporation. Retrieved from <https://cedrus.com/superlab/>
- Chan, J. S., Connolly, S. K., & Setti, A. (2018). The number of stimulus-onset asynchronies affects the perception of the sound-induced flash illusion in young and older adults. *Multisensory Research*, *31*(3–4), 175–190. <https://doi.org/10.1163/22134808-00002605>
- Chandler, D. W., & Grantham, D. W. (1992). Minimum audible movement angle in the horizontal plane as a function of stimulus frequency and bandwidth, source azimuth, and velocity. *The Journal of the Acoustical Society of America*, *91*(3), 1624–1636. <https://doi.org/10.1121/1.402443>
- Changizi, M. A., Hsieh, A., Nijhawan, R., Kanai, R., & Shimojo, S. (2010). Perceiving-the-present and a unifying theory of illusions. In R. Nijhawan & B. Khurana (Eds.), *Space and Time in Perception and Action* (pp. 441–476). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511750540.026>
- Chen, L., Shi, Z., & Müller, H. J. (2010). Influences of intra- and crossmodal grouping on visual and tactile Ternus apparent motion. *Brain Research*, *1354*, 152–162. <https://doi.org/10.1016/j.brainres.2010.07.064>
- Chen, L., & Vroomen, J. (2013). Intersensory binding across space and time: A tutorial review. *Attention, Perception, and Psychophysics*, *75*(5), 790–811. <https://doi.org/10.3758/s13414-013-0475-4>
- 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. <https://doi.org/10.1016/j.cognition.2009.10.012>
- Chen, Y.-C., & Spence, C. (2017). Assessing the role of the 'Unity Assumption' on multisensory

- integration: A review. *Frontiers in Psychology*, 8(MAR), 1–22. <https://doi.org/10.3389/fpsyg.2017.00445>
- Chien, S., Ono, F., & Watanabe, K. (2013). A transient auditory signal shifts the perceived offset position of a moving visual object. *Frontiers in Psychology*, 4(February), 1–10. <https://doi.org/10.3389/fpsyg.2013.00070>
- Choi, H., & Scholl, B. J. (2006). Measuring causal perception: Connections to representational momentum? *Acta Psychologica*, 123(1–2), 91–111. <https://doi.org/10.1016/j.actpsy.2006.06.001>
- Choi, I., Lee, J. Y., & Lee, S. H. (2018). Bottom-up and top-down modulation of multisensory integration. *Current Opinion in Neurobiology*, 52, 115–122. <https://doi.org/10.1016/j.conb.2018.05.002>
- Clark, J. J., & Yuille, A. L. (1990). *Data fusion for sensory information processing systems*. Boston: Kluwer Academic Publishers. <https://doi.org/10.1007/978-1-4757-2076-1>
- Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics*, 16(2), 409–412. <https://doi.org/10.3758/BF03203962>
- Colonus, H., & Diederich, A. (2010). The optimal time window of visual–auditory integration: a reaction time analysis. *Frontiers in Integrative Neuroscience*, 4(MAY 2010), 1–8. <https://doi.org/10.3389/fnint.2010.00011>
- Colonus, H., & Diederich, A. (2017). Measuring multisensory integration: From reaction times to spike counts. *Scientific Reports*, 7(1), 1–11. <https://doi.org/10.1038/s41598-017-03219-5>
- Conrey, B., & Pisoni, D. B. (2006). Auditory-visual speech perception and synchrony detection for speech and nonspeech signals. *The Journal of the Acoustical Society of America*, 119(6), 4065–4073. <https://doi.org/10.1121/1.2195091>
- Cook, L. A., & Van Valkenburg, D. L. (2009). Audio-visual organisation and the temporal ventriloquism effect between grouped sequences: Evidence that unimodal grouping precedes cross-modal integration. *Perception*, 38(8), 1220–1233. <https://doi.org/10.1068/p6344>
- Cooper, L. A., & Munger, M. P. (1993). Extrapolating and remembering positions along cognitive trajectories: Use and limitations of analogies to physical motion. In *Spatial representation: Problems in philosophy and psychology*. (pp. 112–131). Malden: Blackwell Publishing.
- Coull, J. T., & Nobre, A. C. (1998). Where and when to pay attention: The neural systems for directing attention to spatial locations and to time intervals as revealed by both PET

- and fMRI. *Journal of Neuroscience*, 18(18), 7426–7435.
<https://doi.org/10.1523/jneurosci.18-18-07426.1998>
- Cui, Q. N., Razavi, B., O'Neill, W. E., & Paige, G. D. (2010). Perception of auditory, visual, and egocentric spatial alignment adapts differently to changes in eye position. *Journal of Neurophysiology*, 103(2), 1020–1035. <https://doi.org/10.1152/jn.00500.2009>
- Culling, J. F., & Stone, M. A. (2017). Energetic Masking and Masking Release. In J. C. Middlebrooks, J. Z. Simon, A. N. Popper, & R. R. Fay (Eds.), *The Auditory System at the Cocktail Party. Springer Handbook of Auditory Research* (pp. 41–73). Springer, Cham. https://doi.org/10.1007/978-3-319-51662-2_3
- Darwin, C. J. (2008). Spatial hearing and perceiving sources. In W. A. Yost, R. R. Fay, & A. N. Popper (Eds.), *Auditory Perception of Sound Sources. Springer Handbook of Auditory Research* (pp. 215–232). Boston, MA: Springer US. https://doi.org/10.1007/978-0-387-71305-2_8
- de Haas, B., & Rees, G. (2010). Multiple stages of cross-modal integration in visual processing. Comment on “Crossmodal influences on visual perception” by L. Shams & R. Kim. *Physics of Life Reviews*, 7(3), 287–288. <https://doi.org/10.1016/j.plrev.2010.06.007>
- De Sá Teixeira, N. A. (2014). Fourier decomposition of spatial localization errors reveals an idiotropic dominance of an internal model of gravity. *Vision Research*, 105, 177–188. <https://doi.org/10.1016/j.visres.2014.10.024>
- De Sá Teixeira, N. A. (2016). The visual representations of motion and of gravity are functionally independent: Evidence of a differential effect of smooth pursuit eye movements. *Experimental Brain Research*, 234(9), 2491–2504. <https://doi.org/10.1007/s00221-016-4654-0>
- De Sá Teixeira, N. A., Hecht, H., & Oliveira, A. M. (2013). The representational dynamics of remembered projectile locations. *Journal of Experimental Psychology: Human Perception and Performance*, 39(6), 1690–1699. <https://doi.org/10.1037/a0031777>
- De Sá Teixeira, N. A., Kerzel, D., Hecht, H., & Lacquaniti, F. (2019). A novel dissociation between representational momentum and representational gravity through response modality. *Psychological Research*, 83(6), 1223–1236. <https://doi.org/10.1007/s00426-017-0949-4>
- De Sá Teixeira, N. A., & Oliveira, A. M. (2011). Disambiguating the effects of target travelled distance and target vanishing point upon representational momentum. *Journal of Cognitive Psychology*, 23(5), 650–658. <https://doi.org/10.1080/20445911.2011.557357>

- De Sá Teixeira, N. A., & Oliveira, A. M. (2013). Explorando a trajetória espaço-temporal da representação dinâmica de projéteis. *Psicologia: Reflexão e Crítica*, 26(4), 721–729. <https://doi.org/10.1590/S0102-79722013000400012>
- De Sá Teixeira, N. A., & Oliveira, A. M. (2014). Spatial and foveal biases, not perceived mass or heaviness, explain the effect of target size on representational momentum and representational gravity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(6), 1664–1679. <https://doi.org/10.1037/xlm0000011>
- De Sá Teixeira, N. A., Oliveira, A. M., & Amorim, M.-A. (2010). Combined effects of mass and velocity on forward displacement and phenomenological ratings: a functional measurement approach to the momentum metaphor. *Psicológica*, 31, 659–676.
- De Sá Teixeira, N. A., Oliveira, A. M., & Viegas, R. (2008). Integration of kinematics and dynamics functional approach to the integration of kinematic and dynamic variables in causal perception: Is there a link between phenomenology and behavioral responses? *Japanese Psychological Research*, 50(4), 232–241. <https://doi.org/10.1111/j.1468-5884.2008.00379.x>
- Degerman, A., Rinne, T., Pekkola, J., Autti, T., Jääskeläinen, I. P., Sams, M., & Alho, K. (2007). Human brain activity associated with audiovisual perception and attention. *NeuroImage*, 34(4), 1683–1691. <https://doi.org/10.1016/j.neuroimage.2006.11.019>
- Deneve, S., & Pouget, A. (2004). Bayesian multisensory integration and cross-modal spatial links. *Journal of Physiology-Paris*, 98(1–3), 249–258. <https://doi.org/10.1016/j.jphysparis.2004.03.011>
- Di Luca, M., Machulla, T. K., & Ernst, M. O. (2009). Recalibration of multisensory simultaneity: Cross-modal transfer coincides with a change in perceptual latency. *Journal of Vision*, 9(12), 7–7. <https://doi.org/10.1167/9.12.7>
- Dick, M., Ullman, S., & Sagi, D. (1987). Parallel and serial processes in motion detection. *Science*, 237(4813), 400–402. <https://doi.org/10.1126/science.3603025>
- Diederich, A., & Colonius, H. (1987). Intersensory facilitation in the motor component? *Psychological Research*, 49(1), 23–29. <https://doi.org/10.1007/BF00309199>
- Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics*, 66(8), 1388–1404. <https://doi.org/10.3758/BF03195006>
- Diederich, A., Colonius, H., & Schomburg, A. (2008). Assessing age-related multisensory enhancement with the time-window-of-integration model. *Neuropsychologia*, 46(10), 2556–2562. <https://doi.org/10.1016/j.neuropsychologia.2008.03.026>

- Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception*, 9(6), 719–721. <https://doi.org/10.1068/p090719>
- Doehrmann, O., & Naumer, M. J. (2008). Semantics and the multisensory brain: How meaning modulates processes of audio-visual integration. *Brain Research*, 1242, 136–150. <https://doi.org/10.1016/j.brainres.2008.03.071>
- Donohue, S. E., Green, J. J., & Woldorff, M. G. (2015). The effects of attention on the temporal integration of multisensory stimuli. *Frontiers in Integrative Neuroscience*, 9(APR), 1–14. <https://doi.org/10.3389/fnint.2015.00032>
- Donohue, S. E., Roberts, K. C., Grent-'t-Jong, T., & Woldorff, M. G. (2011). The cross-modal spread of attention reveals differential constraints for the temporal and spatial linking of visual and auditory stimulus events. *Journal of Neuroscience*, 31(22), 7982–7990. <https://doi.org/10.1523/JNEUROSCI.5298-10.2011>
- Drewing, K., Hitzel, E., & Scocchia, L. (2018). The haptic and the visual flash-lag effect and the role of flash characteristics. *PLOS ONE*, 13(1), e0189291. <https://doi.org/10.1371/journal.pone.0189291>
- Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. *Neuron*, 57(1), 11–23. <https://doi.org/10.1016/j.neuron.2007.12.013>
- Driver, J., & Spence, C. (2000). Multisensory perception: Beyond modularity and convergence. *Current Biology*, 10(20), R731–R735. [https://doi.org/10.1016/S0960-9822\(00\)00740-5](https://doi.org/10.1016/S0960-9822(00)00740-5)
- Dundon, N. M., Bertini, C., Làdavas, E., Sabel, B. A., & Gall, C. (2015). Visual rehabilitation: Visual scanning, multisensory stimulation and vision restoration trainings. *Frontiers in Behavioral Neuroscience*, 9(JULY), 1–14. <https://doi.org/10.3389/fnbeh.2015.00192>
- Dupierriex, E., Alleysson, D., Ohlmann, T., & Chokron, S. (2008). Spatial bias induced by a non-conflictual task reveals the nature of space perception. *Brain Research*, 1214, 127–135. <https://doi.org/10.1016/j.brainres.2008.01.021>
- Eg, R., & Behne, D. M. (2015). Perceived synchrony for realistic and dynamic audiovisual events. *Frontiers in Psychology*, 6(JUN). <https://doi.org/10.3389/fpsyg.2015.00736>
- Eg, R., Griwodz, C., Halvorsen, P., & Behne, D. (2015). Audiovisual robustness: exploring perceptual tolerance to asynchrony and quality distortion. *Multimedia Tools and Applications*, 74(2), 345–365. <https://doi.org/10.1007/s11042-014-2136-6>
- Eimer, M., & Driver, J. (2001). Crossmodal links in endogenous and exogenous spatial attention: evidence from event-related brain potential studies. *Neuroscience &*

- Biobehavioral Reviews*, 25(6), 497–511. [https://doi.org/10.1016/S0149-7634\(01\)00029-X](https://doi.org/10.1016/S0149-7634(01)00029-X)
- Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision*, 7(5), 1–14. <https://doi.org/10.1167/7.5.7>
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870), 429–433. <https://doi.org/10.1038/415429a>
- Ernst, M. O., Banks, M. S., & Bühlhoff, H. H. (2000). Touch can change visual slant perception. *Nature Neuroscience*, 3(1), 69–73. <https://doi.org/10.1038/71140>
- Ernst, M. O., & Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, 8(4), 162–169. <https://doi.org/10.1016/j.tics.2004.02.002>
- Ernst, M. O., & Di Luca, M. (2011). Multisensory perception: From integration to remapping. In J. Trommershäuser, K. Kording, & M. S. Landy (Eds.), *Sensory Cue Integration* (Vol. 15, pp. 224–250). Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195387247.003.0012>
- Fairhall, S. L., & Macaluso, E. (2009). Spatial attention can modulate audiovisual integration at multiple cortical and subcortical sites. *European Journal of Neuroscience*, 29(6), 1247–1257. <https://doi.org/10.1111/j.1460-9568.2009.06688.x>
- Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *Journal of Neuroscience*, 22(13), 5749–5759. <https://doi.org/10.1523/jneurosci.22-13-05749.2002>
- Faust, M. (1990). *Representational momentum: A dual process perspective*. Unpublished doctoral dissertation. University of Oregon, Eugene.
- Feinkohl, A., Locke, S. M., Leung, J., & Carlile, S. (2014). The effect of velocity on auditory representational momentum. *The Journal of the Acoustical Society of America*, 136(1), EL20–EL25. <https://doi.org/10.1121/1.4881318>
- Feldman, J. (2015). Bayesian inference and “truth”: a comment on Hoffman, Singh, and Prakash. *Psychonomic Bulletin & Review*, 22(6), 1523–1525. <https://doi.org/10.3758/s13423-014-0795-y>
- Feng, W., Störmer, V. S., Martinez, A., McDonald, J. J., & Hillyard, S. A. (2014). Sounds activate visual cortex and improve visual discrimination. *Journal of Neuroscience*, 34(29), 9817–9824. <https://doi.org/10.1523/JNEUROSCI.4869-13.2014>
- Fernandez-Duque, D., & Posner, M. I. (1997). Relating the mechanisms of orienting and alerting. *Neuropsychologia*, 35(4), 477–486. <https://doi.org/10.1016/S0028->

3932(96)00103-0

- Fetsch, C. R., DeAngelis, G. C., & Angelaki, D. E. (2013). Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. *Nature Reviews Neuroscience*, *14*(6), 429–442. <https://doi.org/10.1038/nrn3503>
- Fiebelkorn, I. C., Foxe, J. J., Butler, J. S., & Molholm, S. (2011). Auditory facilitation of visual-target detection persists regardless of retinal eccentricity and despite wide audiovisual misalignments. *Experimental Brain Research*, *213*(2–3), 167–174. <https://doi.org/10.1007/s00221-011-2670-7>
- Fiebelkorn, I. C., Foxe, J. J., & Molholm, S. (2010). Dual mechanisms for the cross-sensory spread of attention: How much do learned associations matter? *Cerebral Cortex*, *20*(1), 109–120. <https://doi.org/10.1093/cercor/bhp083>
- Finke, R. A., & Freyd, J. J. (1985). Transformations of visual memory induced by implied motions of pattern elements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 780–794. <https://doi.org/10.1037/0278-7393.11.1-4.780>
- Finke, R. A., & Freyd, J. J. (1989). Mental extrapolation and cognitive penetrability: Reply to Ranney and proposals for evaluative criteria. *Journal of Experimental Psychology: General*, *118*(4), 403–408. <https://doi.org/10.1037/0096-3445.118.4.403>
- Finke, R. A., Freyd, J. J., & Shyi, G. C. (1986). Implied velocity and acceleration induce transformations of visual memory. *Journal of Experimental Psychology: General*, *115*(2), 175–188. <https://doi.org/10.1037/0096-3445.115.2.175>
- Fiser, J., Chiu, C., & Weliky, M. (2004). Small modulation of ongoing cortical dynamics by sensory input during natural vision. *Nature*, *431*(7008), 573–578. <https://doi.org/10.1038/nature02907>
- Fisher, G. H. (1968). Agreement between the spatial senses. *Perceptual and Motor Skills*, *26*(3), 849–850. <https://doi.org/10.2466/pms.1968.26.3.849>
- Frassinetti, F., Bolognini, N., Bottari, D., Bonora, A., & Làdavas, E. (2005). Audiovisual integration in patients with visual deficit. *Journal of Cognitive Neuroscience*, *17*(9), 1442–1452. <https://doi.org/10.1162/0898929054985446>
- Frassinetti, F., Bolognini, N., & Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research*, *147*(3), 332–343. <https://doi.org/10.1007/s00221-002-1262-y>
- Freeman, E., & Driver, J. (2008). Direction of visual apparent motion driven solely by timing of a static sound. *Current Biology*, *18*(16), 1262–1266. <https://doi.org/10.1016/j.cub.2008.07.066>

- Freeman, T. C. A., Leung, J., Wufong, E., Orchard-Mills, E., Carlile, S., & Alais, D. (2014). Discrimination contours for moving sounds reveal duration and distance cues dominate auditory speed perception. *PLoS ONE*, 9(7), e102864. <https://doi.org/10.1371/journal.pone.0102864>
- Frens, M. A., van Opstal, A. J., & Van der Willigen, R. F. (1995). Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Perception & Psychophysics*, 57(6), 802–816. <https://doi.org/10.3758/BF03206796>
- Freyd, J. J. (1983). The mental representation of movement when static stimuli are viewed. *Perception & Psychophysics*, 33(6), 575–581. <https://doi.org/10.3758/BF03202940>
- Freyd, J. J. (1987). Dynamic mental representations. *Psychological Review*, 94(4), 427–438. <https://doi.org/10.1037/0033-295X.94.4.427>
- Freyd, J. J. (1993). Five hunches about perceptual processes and dynamic representations. In D. Meyer & S. Kornblum (Eds.), *Attention and Performance XIV: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience* (pp. 99–119). Cambridge, MA: MIT Press.
- Freyd, J. J., & Finke, R. A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(1), 126–132. <https://doi.org/10.1037/0278-7393.10.1.126>
- Freyd, J. J., & Johnson, J. Q. (1987). Probing the time course of representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(2), 259–268. <https://doi.org/10.1037/0278-7393.13.2.259>
- Freyd, J. J., & Jones, K. T. (1994). Representational momentum for a spiral path. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4), 968–976. <https://doi.org/10.1037/0278-7393.20.4.968>
- Freyd, J. J., Kelly, M. H., & DeKay, M. L. (1990). Representational momentum in memory for pitch. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(6), 1107–1117. <https://doi.org/10.1037/0278-7393.16.6.1107>
- Fröhlich, F. W. (1924). Über die messung der empfindungszeit. *Pflügers Archiv Für Die Gesamte Physiologie Des Menschen Und Der Tiere*, 202(1), 566–572. <https://doi.org/10.1007/BF01723521>
- Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, 7(7), 773–778. <https://doi.org/10.1038/nn1268>
- Futterweit, L. R., & Beilin, H. (1994). Recognition memory for movement in photographs: A developmental study. *Journal of Experimental Child Psychology*, 57(2), 163–179.

- <https://doi.org/10.1006/jecp.1994.1008>
- García-Pérez, M. A., & Alcalá-Quintana, R. (2012). On the discrepant results in synchrony judgment and temporal-order judgment tasks: a quantitative model. *Psychonomic Bulletin & Review*, 19(5), 820–846. <https://doi.org/10.3758/s13423-012-0278-y>
- Gebhard, J. W., & Mowbray, G. H. (1959). On discriminating the rate of visual flicker and auditory flutter. *The American Journal of Psychology*, 72, 521–529. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/13827044>
- Getzmann, S. (2005a). Representational momentum in spatial hearing does not depend on eye movements. *Experimental Brain Research*, 165(2), 229–238. <https://doi.org/10.1007/s00221-005-2291-0>
- Getzmann, S. (2005b). Shifting the onset of a moving sound source: A Fröhlich effect in spatial hearing. *Hearing Research*, 210(1–2), 104–111. <https://doi.org/10.1016/j.heares.2005.08.003>
- Getzmann, S. (2007). The effect of brief auditory stimuli on visual apparent motion. *Perception*, 36(7), 1089–1103. <https://doi.org/10.1068/p5741>
- Getzmann, S., & Lewald, J. (2007). Localization of moving sound. *Perception & Psychophysics*, 69(6), 1022–1034. <https://doi.org/10.3758/BF03193940>
- Getzmann, S., & Lewald, J. (2009). Constancy of target velocity as a critical factor in the emergence of auditory and visual representational momentum. *Experimental Brain Research*, 193(3), 437–443. <https://doi.org/10.1007/s00221-008-1641-0>
- Getzmann, S., Lewald, J., & Guski, R. (2004). Representational momentum in spatial hearing. *Perception*, 33(5), 591–599. <https://doi.org/10.1068/p5093>
- Ghahramani, Z., Wolpert, D. M., & Jordan, M. I. (1997). Computational models of sensorimotor integration. In P. Morasso & V. Sanguineti (Eds.), *Self-Organization, Computational Maps, and Motor Control* (Vol. 119, pp. 117–147). North-Holland. [https://doi.org/10.1016/S0166-4115\(97\)80006-4](https://doi.org/10.1016/S0166-4115(97)80006-4)
- Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, 10(6), 278–285. <https://doi.org/10.1016/j.tics.2006.04.008>
- Girard, S., Pelland, M., Lepore, F., & Collignon, O. (2013). Impact of the spatial congruence of redundant targets on within-modal and cross-modal integration. *Experimental Brain Research*, 224(2), 275–285. <https://doi.org/10.1007/s00221-012-3308-0>
- Gobara, A., Yamada, Y., & Miura, K. (2016). Crossmodal modulation of spatial localization by mimetic words. *I-Perception*, 7(6), 204166951668424. <https://doi.org/10.1177/2041669516684244>

- Godfroy-Cooper, M., Sandor, P. M. B., Miller, J. D., & Welch, R. B. (2015). The interaction of vision and audition in two-dimensional space. *Frontiers in Neuroscience*, 9(SEP). <https://doi.org/10.3389/fnins.2015.00311>
- Godfroy, M., Roumes, C., & Dauchy, P. (2003). Spatial variations of visual - auditory fusion areas. *Perception*, 32(10), 1233–1245. <https://doi.org/10.1068/p3344>
- Gori, M., Del Viva, M., Sandini, G., & Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Current Biology*, 18(9), 694–698. <https://doi.org/10.1016/j.cub.2008.04.036>
- Gori, M., Sandini, G., Martinoli, C., & Burr, D. (2010). Poor haptic orientation discrimination in nonsighted children may reflect disruption of cross-sensory calibration. *Current Biology*, 20(3), 223–225. <https://doi.org/10.1016/j.cub.2009.11.069>
- Grasso, P. A., Benassi, M., Làdavas, E., & Bertini, C. (2016). Audio-visual multisensory training enhances visual processing of motion stimuli in healthy participants: an electrophysiological study. *European Journal of Neuroscience*, 44(10), 2748–2758. <https://doi.org/10.1111/ejn.13221>
- Gregory, R. L. (1997). Knowledge in perception and illusion. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 352(1358), 1121–1127. <https://doi.org/10.1098/rstb.1997.0095>
- Guttman, S. E., Gilroy, L. A., & Blake, R. (2007). Spatial grouping in human vision: Temporal structure trumps temporal synchrony. *Vision Research*, 47(2), 219–230. <https://doi.org/10.1016/j.visres.2006.09.012>
- Hackett, T. A., De La Mothe, L. A., Ulbert, I., Karmos, G., Smiley, J., & Schroeder, C. E. (2007). Multisensory convergence in auditory cortex, II. Thalamocortical connections of the caudal superior temporal plane. *The Journal of Comparative Neurology*, 502(6), 924–952. <https://doi.org/10.1002/cne.21326>
- Hadad, B., Schwartz, S., Maurer, D., & Lewis, T. L. (2015). Motion perception: a review of developmental changes and the role of early visual experience. *Frontiers in Integrative Neuroscience*, 9(SEP), 1–18. <https://doi.org/10.3389/fnint.2015.00049>
- Hairston, W. D., Laurienti, P. J., Mishra, G., Burdette, J. H., & Wallace, M. T. (2003). Multisensory enhancement of localization under conditions of induced myopia. *Experimental Brain Research*, 152(3), 404–408. <https://doi.org/10.1007/s00221-003-1646-7>
- Hairston, W. D., Wallace, M. T., Vaughan, J. W., Stein, B. E., Norris, J. L., & Schirillo, J. A. (2003). Visual localization ability influences cross-modal bias. *Journal of Cognitive*

- Neuroscience*, 15(1), 20–29. <https://doi.org/10.1162/089892903321107792>
- Halpern, A. R., & Kelly, M. H. (1993). Memory biases in left versus right implied motion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(2), 471–484. <https://doi.org/10.1037/0278-7393.19.2.471>
- Harrington, L. K., & Peck, C. K. (1998). Spatial disparity affects visual-auditory interactions in human sensorimotor processing. *Experimental Brain Research*, 122(2), 247–252. <https://doi.org/10.1007/s002210050512>
- Hasson, U., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *Journal of Neuroscience*, 28(10), 2539–2550. <https://doi.org/10.1523/JNEUROSCI.5487-07.2008>
- Hayes, A. E., & Freyd, J. J. (2002). Representational momentum when attention is divided. *Visual Cognition*, 9(1–2), 8–27. <https://doi.org/10.1080/13506280143000296>
- Helmholtz, H. (1866). *Handbuch der Physiologischen Optik*. Hamburg: Voss.
- Heron, J., Roach, N. W., Hanson, J. V. M., McGraw, P. V., & Whitaker, D. (2012). Audiovisual time perception is spatially specific. *Experimental Brain Research*, 218(3), 477–485. <https://doi.org/10.1007/s00221-012-3038-3>
- Heron, J., Roach, N. W., Whitaker, D., & Hanson, J. V. M. (2010). Attention regulates the plasticity of multisensory timing. *European Journal of Neuroscience*, 31(10), 1755–1762. <https://doi.org/10.1111/j.1460-9568.2010.07194.x>
- Heron, J., Whitaker, D., & McGraw, P. V. (2004). Sensory uncertainty governs the extent of audio-visual interaction. *Vision Research*, 44(25), 2875–2884. <https://doi.org/10.1016/j.visres.2004.07.001>
- Hidaka, S., Kawachi, Y., & Gyoba, J. (2009). The representation of moving 3-D objects in apparent motion perception. *Attention, Perception, & Psychophysics*, 71(6), 1294–1304. <https://doi.org/10.3758/APP.71.6.1294>
- Hidaka, S., Manaka, Y., Teramoto, W., Sugita, Y., Miyauchi, R., Gyoba, J., ... Iwaya, Y. (2009). Alternation of sound location induces visual motion perception of a static object. *PLoS ONE*, 4(12), e8188. <https://doi.org/10.1371/journal.pone.0008188>
- Hidaka, S., Teramoto, W., Keetels, M., & Vroomen, J. (2013). Effect of pitch-space correspondence on sound-induced visual motion perception. *Experimental Brain Research*, 231(1), 117–126. <https://doi.org/10.1007/s00221-013-3674-2>
- Hidaka, S., Teramoto, W., & Sugita, Y. (2015). Spatiotemporal processing in crossmodal interactions for perception of the external world: A review. *Frontiers in Integrative Neuroscience*, 9(DEC), 1–13. <https://doi.org/10.3389/fnint.2015.00062>

- Hidaka, S., Teramoto, W., Sugita, Y., Manaka, Y., Sakamoto, S., & Suzuki, Y. (2011). Auditory motion information drives visual motion perception. *PLoS ONE*, *6*(3), e17499. <https://doi.org/10.1371/journal.pone.0017499>
- Hill, N. I., & Darwin, C. J. (1996). Lateralization of a perturbed harmonic: Effects of onset asynchrony and mistuning. *The Journal of the Acoustical Society of America*, *100*(4), 2352–2364. <https://doi.org/10.1121/1.417945>
- Hillis, J. M., Ernst, M. O., Banks, M. S., & Landy, M. S. (2002). Combining sensory information: Mandatory fusion within, but not between, senses. *Science*, *298*(5598), 1627–1630. <https://doi.org/10.1126/science.1075396>
- Holmes, N. P. (2007). The law of inverse effectiveness in neurons and behaviour: Multisensory integration versus normal variability. *Neuropsychologia*, *45*(14), 3340–3345. <https://doi.org/10.1016/j.neuropsychologia.2007.05.025>
- Holmes, N. P. (2009). The principle of inverse effectiveness in multisensory integration: Some statistical considerations. *Brain Topography*, *21*(3–4), 168–176. <https://doi.org/10.1007/s10548-009-0097-2>
- Holmes, N. P., & Spence, C. (2005). Multisensory integration: Space, time and superadditivity. *Current Biology*, *15*(18), R762–764. <https://doi.org/10.1016/j.cub.2005.08.058>
- Holmes, N. P., & Spence, C. (2006). Beyond the body schema: Visual, prosthetic, and technological contributions to bodily perception and awareness. In I. M. Thornton, M. Grosjean, & M. Shiffrin (Eds.), *Human Body Perception From the Inside Out* (pp. 15–64). Oxford University Press.
- Hommel, B. (2011). The Simon effect as tool and heuristic. *Acta Psychologica*, *136*(2), 189–202. <https://doi.org/10.1016/j.actpsy.2010.04.011>
- Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, *24*(05), 849–878. <https://doi.org/10.1017/S0140525X01000103>
- Hubbard, T. L. (1990). Cognitive representation of linear motion: Possible direction and gravity effects in judged displacement. *Memory & Cognition*, *18*(3), 299–309. <https://doi.org/10.3758/BF03213883>
- Hubbard, T. L. (1993). Auditory representational momentum: Musical schemata and modularity. *Bulletin of the Psychonomic Society*, *31*(3), 201–204. <https://doi.org/10.3758/BF03337324>
- Hubbard, T. L. (1995a). Auditory representational momentum: Surface form, direction, and velocity effects. *The American Journal of Psychology*, *108*(2), 255.

- <https://doi.org/10.2307/1423131>
- Hubbard, T. L. (1995b). Cognitive representation of motion: Evidence for friction and gravity analogues. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(1), 241–254. <https://doi.org/10.1037/0278-7393.21.1.241>
- Hubbard, T. L. (1995c). Environmental invariants in the representation of motion: Implied dynamics and representational momentum, gravity, friction, and centripetal force. *Psychonomic Bulletin & Review*, 2(3), 322–338. <https://doi.org/10.3758/BF03210971>
- Hubbard, T. L. (1996). Representational momentum, centripetal force, and curvilinear impetus. *Journal of Experimental Psychology: Learning Memory and Cognition*, 22(4), 1049–1060. <https://doi.org/10.1037/0278-7393.22.4.1049>
- Hubbard, T. L. (1997). Target size and displacement along the axis of implied gravitational attraction: Effects of implied weight and evidence of representational gravity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(6), 1484–1493. <https://doi.org/10.1037/0278-7393.23.6.1484>
- Hubbard, T. L. (1998a). Representational momentum and other displacements in memory as evidence for nonconscious knowledge of physical principles. In S. R. Hameroff, A. W. Kaszniak, & A. Scott (Eds.), *Towards a science of consciousness II: The second Tucson discussions and debates* (pp. 505–512). Cambridge, MA: MIT Press.
- Hubbard, T. L. (1998b). Some effects of representational friction, target size, and memory averaging on memory for vertically moving targets. *Canadian Journal of Experimental Psychology*, 52(1), 44–49.
- Hubbard, T. L. (2005). Representational momentum and related displacements in spatial memory: A review of the findings. *Psychonomic Bulletin & Review*, 12(5), 822–851. <https://doi.org/10.3758/BF03196775>
- Hubbard, T. L. (2006). Bridging the gap: Possible roles and contributions of representational momentum. *Psicológica*, 27(1), 1–34.
- Hubbard, T. L. (2008). Representational momentum contributes to motion induced mislocalization of stationary objects. *Visual Cognition*, 16(1), 44–67. <https://doi.org/10.1080/13506280601155468>
- Hubbard, T. L. (2010). Approaches to representational momentum: theories and models. In R. Nijhawan & B. Khurana (Eds.), *Space and Time in Perception and Action* (pp. 338–365). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511750540.020>
- Hubbard, T. L. (2013). Do the flash-lag effect and representational momentum involve

- similar extrapolations? *Frontiers in Psychology*, 4(MAY), 1–6.
<https://doi.org/10.3389/fpsyg.2013.00290>
- Hubbard, T. L. (2014a). Forms of momentum across space: Representational, operational, and attentional. *Psychonomic Bulletin & Review*, 21(6), 1371–1403.
<https://doi.org/10.3758/s13423-014-0624-3>
- Hubbard, T. L. (2014b). The flash-lag effect and related mislocalizations: Findings, properties, and theories. *Psychological Bulletin*, 140(1), 308–338.
<https://doi.org/10.1037/a0032899>
- Hubbard, T. L. (2015). Forms of momentum across time: Behavioral and psychological. *The Journal of Mind and Behavior*, 36(1/2), 47–82.
- Hubbard, T. L. (2017). Toward a general theory of momentum-like effects. *Behavioural Processes*, 141(Part 1), 50–66. <https://doi.org/10.1016/j.beproc.2017.02.019>
- Hubbard, T. L. (2018a). Influences on representational momentum. In T. L. Hubbard (Ed.), *Spatial Biases in Perception and Cognition* (pp. 121–138). Cambridge: Cambridge University Press.
- Hubbard, T. L. (Ed.). (2018b). *Spatial Biases in Perception and Cognition*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781316651247>
- Hubbard, T. L. (2019). Momentum-like effects and the dynamics of perception, cognition, and action. *Attention, Perception, and Psychophysics*. <https://doi.org/10.3758/s13414-019-01770-z>
- Hubbard, T. L. (2020). Representational gravity: Empirical findings and theoretical implications. *Psychonomic Bulletin and Review*, 27(1), 36–55.
<https://doi.org/10.3758/s13423-019-01660-3>
- Hubbard, T. L., & Bharucha, J. J. (1988). Judged displacement in apparent vertical and horizontal motion. *Perception & Psychophysics*, 44(3), 211–221.
<https://doi.org/10.3758/BF03206290>
- Hubbard, T. L., & Courtney, J. R. (2010). Cross-Modal Influences on Representational Momentum and Representational Gravity. *Perception*, 39(6), 851–862.
<https://doi.org/10.1068/p6538>
- Hubbard, T. L., & Favretto, A. (2003). Naïve impetus and Michotte’s “tool effect”: evidence from representational momentum. *Psychological Research*, 67(2), 134–152.
<https://doi.org/10.1007/s00426-002-0122-5>
- Hubbard, T. L., Kumar, A. M., & Carp, C. L. (2009). Effects of spatial cueing on representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition*,

- 35(3), 666–677. <https://doi.org/10.1037/a0014870>
- Hubbard, T. L., & Lange, M. (2010). Prior probabilities and representational momentum: A signal detection analysis. *Journal of Vision*, 9(8), 692–692. <https://doi.org/10.1167/9.8.692>
- Hubbard, T. L., & Motes, M. A. (2002). Does representational momentum reflect a distortion of the length or the endpoint of a trajectory? In *Cognition* (Vol. 82, pp. 89–99). [https://doi.org/10.1016/S0010-0277\(01\)00156-1](https://doi.org/10.1016/S0010-0277(01)00156-1)
- Hubbard, T. L., & Ruppel, S. E. (1999). Representational momentum and the landmark attraction effect. *Canadian Journal of Experimental Psychology*, 53(3), 242–256. <https://doi.org/10.1037/h0087313>
- Hubbard, T. L., & Ruppel, S. E. (2002). A possible role of naïve impetus in Michotte’s “launching effect” Evidence from representational momentum. *Visual Cognition*, 9(1–2), 153–176. <https://doi.org/10.1080/13506280143000377>
- Hubbard, T. L., & Ruppel, S. E. (2013). A Fröhlich effect and representational gravity in memory for auditory pitch. *Journal of Experimental Psychology: Human Perception and Performance*, 39(4), 1153–1164. <https://doi.org/10.1037/a0031103>
- Hubbard, T. L., & Ruppel, S. E. (2014). An effect of contrast and luminance on visual representational momentum for location. *Perception*, 43(8), 754–766. <https://doi.org/10.1068/p7714>
- Huddleston, W. E., Lewis, J. W., Phinney, R. E., & DeYoe, E. A. (2008). Auditory and visual attention-based apparent motion share functional parallels. *Perception & Psychophysics*, 70(7), 1207–1216. <https://doi.org/10.3758/PP.70.7.1207>
- Hudson, M., Liu, C. H., & Jellema, T. (2009). Anticipating intentional actions: The effect of eye gaze direction on the judgment of head rotation. *Cognition*, 112(3), 423–434. <https://doi.org/10.1016/j.cognition.2009.06.011>
- Ikumi, N., & Soto-Faraco, S. (2014). Selective attention modulates the direction of audio-visual temporal recalibration. *PLoS ONE*, 9(7). <https://doi.org/10.1371/journal.pone.0099311>
- Isaiah, A., Vongpaisal, T., King, A. J., & Hartley, D. E. H. (2014). Multisensory training improves auditory spatial processing following bilateral cochlear implantation. *Journal of Neuroscience*, 34(33), 11119–11130. <https://doi.org/10.1523/JNEUROSCI.4767-13.2014>
- Jack, C. E., & Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the “ventriloquism” effect. *Perceptual and Motor Skills*, 37(3), 967–

979. <https://doi.org/10.2466/pms.1973.37.3.967>
- Jay, M. F., & Sparks, D. L. (1984). Auditory receptive fields in primate superior colliculus shift with changes in eye position. *Nature*, 309(5966), 345–347. <https://doi.org/10.1038/309345a0>
- Jenny, C., & Reuter, C. (2020). Usability of individualized head-related transfer functions in virtual reality: Empirical study with perceptual attributes in sagittal plane sound localization. *JMIR Serious Games*, 8(3), e17576. <https://doi.org/10.2196/17576>
- Johnston, H. M., & Jones, M. R. (2006). Higher order pattern structure influences auditory representational momentum. *Journal of Experimental Psychology: Human Perception and Performance*, 32(1), 2–17. <https://doi.org/10.1037/0096-1523.32.1.2>
- Joordens, S., Spalek, T. M., Razmy, S., & Van Duijn, M. (2004). A clockwork orange: Compensation opposing momentum in memory for location. *Memory & Cognition*, 32(1), 39–50. <https://doi.org/10.3758/BF03195819>
- Jordan, J. S., & Hunsinger, M. (2008). Learned patterns of action-effect anticipation contribute to the spatial displacement of continuously moving stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, 34(1), 113–124. <https://doi.org/10.1037/0096-1523.34.1.113>
- Jordan, J. S., & Knoblich, G. (2004). Spatial perception and control. *Psychonomic Bulletin & Review*, 11(1), 54–59. <https://doi.org/10.3758/BF03206460>
- Jordan, J. S., Stork, S., Knuf, L., Kerzel, D., & Musseler, J. (2002). Action planning affects spatial localization. *Common Mechanisms in Perception and Action (Attention and Performance XIX)*, (1), 158–176.
- Jordan, K. E., Clark, K., & Mitroff, S. R. (2010). See an object, hear an object file: Object correspondence transcends sensory modality. *Visual Cognition*, 18(4), 492–503. <https://doi.org/10.1080/13506280903338911>
- Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, 24(2), 175–219. [https://doi.org/10.1016/0010-0285\(92\)90007-0](https://doi.org/10.1016/0010-0285(92)90007-0)
- Kanaya, S., & Yokosawa, K. (2011). Perceptual congruency of audio-visual speech affects ventriloquism with bilateral visual stimuli. *Psychonomic Bulletin and Review*, 18(1), 123–128. <https://doi.org/10.3758/s13423-010-0027-z>
- Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron*, 22(4), 751–761. [https://doi.org/10.1016/S0896-6273\(00\)80734-](https://doi.org/10.1016/S0896-6273(00)80734-)

5

- Kawachi, Y. (2016). Visual mislocalization of moving objects in an audiovisual event. *PLOS ONE*, *11*(4), e0154147. <https://doi.org/10.1371/journal.pone.0154147>
- Kayser, C., & Shams, L. (2015). Multisensory causal inference in the brain. *PLOS Biology*, *13*(2), e1002075. <https://doi.org/10.1371/journal.pbio.1002075>
- Keller, I., & Lefin-Rank, G. (2010). Improvement of visual search after audiovisual exploration training in hemianopic patients. *Neurorehabilitation and Neural Repair*, *24*(7), 666–673. <https://doi.org/10.1177/1545968310372774>
- Kelly, M. H., & Freyd, J. J. (1987). Explorations of representational momentum. *Cognitive Psychology*, *19*(3), 369–401. [https://doi.org/10.1016/0010-0285\(87\)90009-0](https://doi.org/10.1016/0010-0285(87)90009-0)
- Kerzel, D. (2000). Eye movements and visible persistence explain the mislocalization of the final position of a moving target. *Vision Research*, *40*(27), 3703–3715. [https://doi.org/10.1016/S0042-6989\(00\)00226-1](https://doi.org/10.1016/S0042-6989(00)00226-1)
- Kerzel, D. (2002a). A matter of design: No representational momentum without predictability. *Visual Cognition*, *9*(1–2), 66–80. <https://doi.org/10.1080/13506280143000322>
- Kerzel, D. (2002b). Attention shifts and memory averaging. *The Quarterly Journal of Experimental Psychology Section A*, *55*(2), 425–443. <https://doi.org/10.1080/02724980143000424>
- Kerzel, D. (2002c). The locus of “memory displacement” is at least partially perceptual: Effects of velocity, expectation, friction, memory averaging, and weight. *Perception & Psychophysics*, *64*(4), 680–692. <https://doi.org/10.3758/BF03194735>
- Kerzel, D. (2003a). Attention maintains mental extrapolation of target position: irrelevant distractors eliminate forward displacement after implied motion. *Cognition*, *88*(1), 109–131. [https://doi.org/10.1016/S0010-0277\(03\)00018-0](https://doi.org/10.1016/S0010-0277(03)00018-0)
- Kerzel, D. (2003b). Centripetal force draws the eyes, not memory of the target, toward the center. *Journal of Experimental Psychology: Learning Memory and Cognition*, *29*(3), 458–466. <https://doi.org/10.1037/0278-7393.29.3.458>
- Kerzel, D. (2003c). Mental extrapolation of target position is strongest with weak motion signals and motor responses. *Vision Research*, *43*(25), 2623–2635. [https://doi.org/10.1016/S0042-6989\(03\)00466-8](https://doi.org/10.1016/S0042-6989(03)00466-8)
- Kerzel, D. (2005). Representational momentum beyond internalized physics. *Current Directions in Psychological Science*, *14*(4), 180–184. <https://doi.org/10.1111/j.0963-7214.2005.00360.x>

- Kerzel, D. (2006). Comment and Reply Why eye movements and perceptual factors have to be controlled in studies on “representational momentum.” *Psychonomic Bulletin & Review*, 13(1), 166–173. <https://doi.org/10.3758/BF03193829>
- Kerzel, D., & Gegenfurtner, K. R. (2003). Neuronal processing delays are compensated in the sensorimotor branch of the visual system. *Current Biology*, 13(22), 1975–1978. <https://doi.org/10.1016/j.cub.2003.10.054>
- Kerzel, D., Jordan, J. S., & Müsseler, J. (2001). The role of perception in the mislocalization of the final position of a moving target. *Journal of Experimental Psychology: Human Perception and Performance*, 27(4), 829–840. <https://doi.org/10.1037/0096-1523.27.4.829>
- Kilteni, K., Maselli, A., Körding, K. P., & Slater, M. (2015). Over my fake body: body ownership illusions for studying the multisensory basis of own-body perception. *Frontiers in Human Neuroscience*, 9(MAR). <https://doi.org/10.3389/fnhum.2015.00141>
- Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: Is it long-term and implicit? *Neuroscience Letters*, 461(2), 145–149. <https://doi.org/10.1016/j.neulet.2009.06.030>
- Kim, R., Seitz, A. R., & Shams, L. (2008). Benefits of stimulus congruency for multisensory facilitation of visual learning. *PLoS ONE*, 3(1). <https://doi.org/10.1371/journal.pone.0001532>
- Kim, S.-M., & Choi, W. (2005). On the externalization of virtual sound images in headphone reproduction: A Wiener filter approach. *The Journal of the Acoustical Society of America*, 117(6), 3657–3665. <https://doi.org/10.1121/1.1921548>
- King, A. J. (2009). Visual influences on auditory spatial learning. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 364(1515), 331–339. <https://doi.org/10.1098/rstb.2008.0230>
- Kitagawa, N., & Ichihara, S. (2002). Hearing visual motion in depth. *Nature*, 416(6877), 172–174. <https://doi.org/10.1038/416172a>
- Kitajima, N., & Yamashita, Y. (1999). Dynamic capture of sound motion by light stimuli moving in three-dimensional space. *Perceptual and Motor Skills*, 89(3_suppl), 1139–1158. <https://doi.org/10.2466/pms.1999.89.3f.1139>
- Klein, R. M. (2000). Inhibition of return. *Trends in Cognitive Sciences*, 4(4), 138–147. [https://doi.org/10.1016/S1364-6613\(00\)01452-2](https://doi.org/10.1016/S1364-6613(00)01452-2)
- Knill, D. C. (2007). Robust cue integration: A Bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant. *Journal of Vision*, 7(7), 5.

- <https://doi.org/10.1167/7.7.5>
- Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2009). Competition between auditory and visual spatial cues during visual task performance. *Experimental Brain Research*, *195*(4), 593–602. <https://doi.org/10.1007/s00221-009-1829-y>
- Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta Psychologica*, *134*(3), 372–384. <https://doi.org/10.1016/j.actpsy.2010.03.010>
- König, P., & Krüger, N. (2006). Symbols as self-emergent entities in an optimization process of feature extraction and predictions. *Biological Cybernetics*, *94*(4), 325–334. <https://doi.org/10.1007/s00422-006-0050-3>
- König, P., Wilming, N., Kaspar, K., Nagel, S. K., & Onat, S. (2013). Predictions in the light of your own action repertoire as a general computational principle. *Behavioral and Brain Sciences*, *36*(3), 219–220. <https://doi.org/10.1017/S0140525X12002294>
- Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS ONE*, *2*(9), e943. <https://doi.org/10.1371/journal.pone.0000943>
- Kourtzi, Z., & Kanwisher, N. (2000). Activation in human MT/MST by static images with implied motion. *Journal of Cognitive Neuroscience*, *12*(1), 48–55. <https://doi.org/10.1162/08989290051137594>
- Krauzlis, R. J. (2005). The control of voluntary eye movements: New perspectives. *The Neuroscientist*, *11*(2), 124–137. <https://doi.org/10.1177/1073858404271196>
- Krekelberg, B., & Lappe, M. (2001). Neuronal latencies and the position of moving objects. *Trends in Neurosciences*, *24*(6), 335–339. [https://doi.org/10.1016/S0166-2236\(00\)01795-1](https://doi.org/10.1016/S0166-2236(00)01795-1)
- Kubovy, M., & Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition*, *80*, 97–126. [https://doi.org/10.1016/S0010-0277\(00\)00155-4](https://doi.org/10.1016/S0010-0277(00)00155-4)
- Kwon, O.-S., Tadin, D., & Knill, D. C. (2015). Unifying account of visual motion and position perception. *Proceedings of the National Academy of Sciences*, *112*(26), 8142–8147. <https://doi.org/10.1073/pnas.1500361112>
- Lamme, V. A. F. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *Journal of Neuroscience*, *15*(2), 1605–1615. <https://doi.org/10.1523/jneurosci.15-02-01605.1995>
- Landy, M. S., & Kojima, H. (2001). Ideal cue combination for localizing texture-defined edges. *Journal of the Optical Society of America A*, *18*(9), 2307.

- <https://doi.org/10.1364/JOSAA.18.002307>
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research*, 35(3), 389–412. [https://doi.org/10.1016/0042-6989\(94\)00176-M](https://doi.org/10.1016/0042-6989(94)00176-M)
- Laurienti, P. J., Burdette, J. H., Maldjian, J. A., & Wallace, M. T. (2006). Enhanced multisensory integration in older adults. *Neurobiology of Aging*, 27(8), 1155–1163. <https://doi.org/10.1016/j.neurobiolaging.2005.05.024>
- Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental Brain Research*, 158(4), 405–414. <https://doi.org/10.1007/s00221-004-1913-2>
- Leclère, T., Lavandier, M., & Perrin, F. (2019). On the externalization of sound sources with headphones without reference to a real source. *The Journal of the Acoustical Society of America*, 146(4), 2309–2320. <https://doi.org/10.1121/1.5128325>
- Lee, A. K. C., Maddox, R. K., & Bizley, J. K. (2019). An object-based interpretation of audiovisual processing. In A. K. C. Lee, M. T. Wallace, A. B. Coffin, A. N. Popper, & R. R. Fay (Eds.), *Multisensory Processes. Springer Handbook of Auditory Research* (pp. 59–83). Cham: Springer, Cham. https://doi.org/10.1007/978-3-030-10461-0_4
- Lee, A. K. C., & Wallace, M. T. (2019). Visual influence on auditory perception. In A. K. C. Lee, M. T. Wallace, A. B. Coffin, A. N. Popper, & R. R. Fay (Eds.), *Multisensory Processes. Springer Handbook of Auditory Research* (Vol. 68, pp. 1–8). Cham: Springer, Cham. https://doi.org/10.1007/978-3-030-10461-0_1
- Leone, L. M., & McCourt, M. E. (2013). The roles of physical and physiological simultaneity in audiovisual multisensory facilitation. *I-Perception*, 4(4), 213–228. <https://doi.org/10.1068/i0532>
- Lewald, J., Ehrenstein, W. H., & Guski, R. (2001). Spatio-temporal constraints for auditory-visual integration. *Behavioural Brain Research*, 121(1–2), 69–79. [https://doi.org/10.1016/S0166-4328\(00\)00386-7](https://doi.org/10.1016/S0166-4328(00)00386-7)
- Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research*, 16(3), 468–478. [https://doi.org/10.1016/S0926-6410\(03\)00074-0](https://doi.org/10.1016/S0926-6410(03)00074-0)
- Lewis, R., & Noppeney, U. (2010). Audiovisual synchrony improves motion discrimination via enhanced connectivity between early visual and auditory areas. *Journal of Neuroscience*, 30(37), 12329–12339. <https://doi.org/10.1523/JNEUROSCI.5745-09.2010>

- Lewkowicz, D. J. (2003). Learning and discrimination of audiovisual events in human infants: The hierarchical relation between intersensory temporal synchrony and rhythmic pattern cues. *Developmental Psychology*, 39(5), 795–804. <https://doi.org/10.1037/0012-1649.39.5.795>
- Li, Q., Yang, H., Sun, F., & Wu, J. (2015). Spatiotemporal relationships among audiovisual stimuli modulate auditory facilitation of visual target discrimination. *Perception*, 44(3), 232–242. <https://doi.org/10.1068/p7846>
- Lippert, M., Logothetis, N. K., & Kayser, C. (2007). Improvement of visual contrast detection by a simultaneous sound. *Brain Research*, 1173(1), 102–109. <https://doi.org/10.1016/j.brainres.2007.07.050>
- Liu, T., Stevens, S. T., & Carrasco, M. (2007). Comparing the time course and efficacy of spatial and feature-based attention. *Vision Research*, 47(1), 108–113. <https://doi.org/10.1016/j.visres.2006.09.017>
- Los, S. A., & Van der Burg, E. (2013). Sound speeds vision through preparation, not integration. *Journal of Experimental Psychology: Human Perception and Performance*, 39(6), 1612–1624. <https://doi.org/10.1037/a0032183>
- Ma, W. J., Zhou, X., Ross, L. A., Foxe, J. J., & Parra, L. C. (2009). Lip-reading aids word recognition most in moderate noise: A Bayesian explanation using high-dimensional feature space. *PLoS ONE*, 4(3). <https://doi.org/10.1371/journal.pone.0004638>
- Macaluso, E., & Driver, J. (2005). Multisensory spatial interactions: A window onto functional integration in the human brain. *Trends in Neurosciences*, 28(5), 264–271. <https://doi.org/10.1016/j.tins.2005.03.008>
- Macaluso, E., Frith, C. D., & Driver, J. (2002). Supramodal effects of covert spatial orienting triggered by visual or tactile events. *Journal of Cognitive Neuroscience*, 14(3), 389–401. <https://doi.org/10.1162/089892902317361912>
- Macaluso, E., Noppeney, U., Talsma, D., Vercillo, T., Hartcher-O'Brien, J., & Adam, R. (2016). The curious incident of attention in multisensory integration: Bottom-up vs. Top-down. *Multisensory Research*, 29(6–7), 557–583. <https://doi.org/10.1163/22134808-00002528>
- MacKay, D. M. (1958). Perceptual stability of a stroboscopically lit visual field containing self-luminous objects. *Nature*, 181(4607), 507–508. <https://doi.org/10.1038/181507a0>
- Maddox, R. K., & Shinn-Cunningham, B. G. (2012). Influence of task-relevant and task-irrelevant feature continuity on selective auditory attention. *JARO - Journal of the Association for Research in Otolaryngology*, 13(1), 119–129.

- <https://doi.org/10.1007/s10162-011-0299-7>
- Maeda, F., Kanai, R., & Shimojo, S. (2004). Changing pitch induced visual motion illusion. *Current Biology*, *14*(23), R990–R991. <https://doi.org/10.1016/j.cub.2004.11.018>
- Martins, I. C. (2015). *Rotações mentais implícitas e representações dinâmicas em pessoas com paralisia cerebral: o impacto das limitações da actividade*. Tese de doutoramento não-publicada. Faculdade de Psicologia e de Ciências da Educação da Universidade de Coimbra, Coimbra.
- Mateeff, S., & Hohnsbein, J. (1988). Dynamic auditory localization: Perceived position of a moving sound source. *Acta Physiologica et Pharmacologica Bulgarica*, *14*(3), 32–38.
- Matthews, N., Welch, L., Achtman, R., Fenton, R., & FitzGerald, B. (2016). Simultaneity and temporal order judgments exhibit distinct reaction times and training effects. *PLoS ONE*, *11*(1), e0145926. <https://doi.org/10.1371/journal.pone.0145926>
- Maus, G. W., & Nijhawan, R. (2009). Going, going, gone: Localizing abrupt offsets of moving objects. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 611–626. <https://doi.org/10.1037/a0012317>
- McCloskey, M., Washburn, A., & Felch, L. (1983). Intuitive physics: The straight-down belief and its origin. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(4), 636–649. <https://doi.org/10.1037/0278-7393.9.4.636>
- McGeorge, P., Beschin, N., & Della Sala, S. (2006). Representing target motion: The role of the right hemisphere in the forward displacement bias. *Neuropsychology*, *20*(6), 708–715. <https://doi.org/10.1037/0894-4105.20.6.708>
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. <https://doi.org/10.1038/264746a0>
- Mckee, S. P., & Nakayama, K. (1984). The detection of motion in the peripheral visual field. *Vision Research*, *24*(1), 25–32. [https://doi.org/10.1016/0042-6989\(84\)90140-8](https://doi.org/10.1016/0042-6989(84)90140-8)
- Mégevand, P., Molholm, S., Nayak, A., & Foxe, J. J. (2013). Recalibration of the multisensory temporal window of integration results from changing task demands. *PLoS ONE*, *8*(8), e71608. <https://doi.org/10.1371/journal.pone.0071608>
- Mehta, A. D. (2000). Intermodal selective attention in monkeys. II: Physiological mechanisms of modulation. *Cerebral Cortex*, *10*(4), 359–370. <https://doi.org/10.1093/cercor/10.4.359>
- Meijer, D., Veselič, S., Calafiore, C., & Noppeney, U. (2019). Integration of audiovisual spatial signals is not consistent with maximum likelihood estimation. *Cortex*, *119*, 74–88. <https://doi.org/10.1016/j.cortex.2019.03.026>

- Melcher, D., Wutz, A., Drewes, J., & Fairhall, S. (2014). The role of temporal integration windows in visual perception. *Procedia - Social and Behavioral Sciences*, 126, 92–93. <https://doi.org/10.1016/j.sbspro.2014.02.323>
- Mendonça, C., Campos, G., Dias, P., Vieira, J., Ferreira, J. P., & Santos, J. A. (2012). On the improvement of localization accuracy with non-individualized HRTF-based sounds. *Journal of the Audio Engineering Society*, 60(10), 821–830.
- Mendonça, C., Escher, A., van de Par, S., & Colonius, H. (2015). Predicting auditory space calibration from recent multisensory experience. *Experimental Brain Research*, 233(7), 1983–1991. <https://doi.org/10.1007/s00221-015-4259-z>
- Mendonça, C., Mandelli, P., & Pulkki, V. (2016). Modeling the perception of audiovisual distance: Bayesian causal inference and other models. *PLOS ONE*, 11(12), e0165391. <https://doi.org/10.1371/journal.pone.0165391>
- Meredith, M. A. (2002). On the neuronal basis for multisensory convergence: a brief overview. *Cognitive Brain Research*, 14(1), 31–40. [https://doi.org/10.1016/S0926-6410\(02\)00059-9](https://doi.org/10.1016/S0926-6410(02)00059-9)
- Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *The Journal of Neuroscience*, 7(10), 3215–3229. <https://doi.org/10.1523/jneurosci.07-10-03215.1987>
- Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science*, 221(4608), 389–391. <https://doi.org/10.1126/science.6867718>
- Meredith, M. A., & Stein, B. E. (1986). Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Research*, 365(2), 350–354. [https://doi.org/10.1016/0006-8993\(86\)91648-3](https://doi.org/10.1016/0006-8993(86)91648-3)
- Meredith, M. A., & Stein, B. E. (1996). Spatial determinants of multisensory integration in cat superior colliculus neurons. *Journal of Neurophysiology*, 75(5), 1843–1857. <https://doi.org/10.1152/jn.1996.75.5.1843>
- Merz, S., Meyerhoff, H. S., Frings, C., & Spence, C. (2020). Representational momentum in vision and touch: Visual motion information biases tactile spatial localization. *Attention, Perception, and Psychophysics*, 82(5), 2618–2629. <https://doi.org/10.3758/s13414-020-01989-1>
- Merz, S., Meyerhoff, H. S., Spence, C., & Frings, C. (2019). Implied tactile motion: Localizing dynamic stimulations on the skin. *Attention, Perception, and Psychophysics*, 81(3), 794–808. <https://doi.org/10.3758/s13414-018-01645-9>

- Meyer, G. F., Wuerger, S. M., Röhrbein, F., & Zetsche, C. (2005). Low-level integration of auditory and visual motion signals requires spatial co-localisation. *Experimental Brain Research*, 166(3–4), 538–547. <https://doi.org/10.1007/s00221-005-2394-7>
- Middlebrooks, J. C., & Green, D. M. (1991). Sound localization by human listeners. *Annual Review of Psychology*, 42(1), 135–159. <https://doi.org/10.1146/annurev.ps.42.020191.001031>
- Mihalik, A., & Noppeney, U. (2020). Causal inference in audiovisual perception. *The Journal of Neuroscience*, 40(34), 6600–6612. <https://doi.org/10.1523/JNEUROSCI.0051-20.2020>
- Miller, J. D., & Wenzel, E. M. (2002). Recent developments in SLAB: A software-based system for interactive spatial sound synthesis. In *Proceedings of the 2002 International Conference on Auditory Display* (pp. 403–408). Kyoto, Japan.
- Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. Oxford: Oxford University Press.
- Misceo, G. F., Jackson, S. V. S., & Perdue, J. R. (2014). Again, knowledge of common source fails to promote visual-haptic integration. *Perceptual and Motor Skills*, 118(1), 183–194. <https://doi.org/10.2466/24.23.PMS.118k11w0>
- Mitroff, S. R., Scholl, B. J., & Noies, N. S. (2007). Object files can be purely episodic. *Perception*, 36(12), 1730–1735. <https://doi.org/10.1068/p5804>
- Molholm, S., Martinez, A., Shpaner, M., & Foxe, J. J. (2007). Object-based attention is multisensory: Co-activation of an object's representations in ignored sensory modalities. *European Journal of Neuroscience*, 26(2), 499–509. <https://doi.org/10.1111/j.1460-9568.2007.05668.x>
- Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. *Cognitive Brain Research*, 17(1), 154–163. [https://doi.org/10.1016/S0926-6410\(03\)00089-2](https://doi.org/10.1016/S0926-6410(03)00089-2)
- Morgan, W. A. (1939). A test for the significance of the difference between the two variances in a sample from a normal bivariate population. *Biometrika*, 31(1/2), 13. <https://doi.org/10.2307/2334972>
- Mozolic, J. L., Hugenschmidt, C. E., Peiffer, A. M., & Laurienti, P. J. (2008). Modality-specific selective attention attenuates multisensory integration. *Experimental Brain Research*, 184(1), 39–52. <https://doi.org/10.1007/s00221-007-1080-3>
- Mozolic, J. L., Hugenschmidt, C. E., Peiffer, A. M., & Laurienti, P. J. (2012). Multisensory integration and aging. In M. M. Murray & M. T. Wallace (Eds.), *The Neural Bases of Multisensory Processes*. Boca Raton, FL: CRC Press/Taylor & Francis.

- Mozolic, J. L., Joyner, D., Hugenschmidt, C. E., Peiffer, A. M., Kraft, R. A., Maldjian, J. A., & Laurienti, P. J. (2008). Cross-modal deactivations during modality-specific selective attention. *BMC Neurology*, *8*, 1–11. <https://doi.org/10.1186/1471-2377-8-35>
- MSC.Software Corporation. (2000). Interactive Physics 2000 (version 5.2.1.1) [Computer software]. San Mateo, CA: MSC.Software. Retrieved from <https://www.design-simulation.com/IP/index.php>
- Munger, M. P., Dellinger, M. C., Lloyd, T. G., Johnson-Reid, K., Tonelli, N. J., Wolf, K., & Scott, J. M. (2006). Representational momentum in scenes: Learning spatial layout. *Memory & Cognition*, *34*(7), 1557–1568. <https://doi.org/10.3758/BF03195919>
- Munger, M. P., & Minchew, J. H. (2002). Parallels between remembering and predicting an object's location. *Visual Cognition*, *9*(1-2), 177–194. <https://doi.org/10.1080/13506280143000386>
- Munger, M. P., & Owens, T. R. (2004). Representational momentum and the flash-lag effect. *Visual Cognition*, *11*(1), 81–103. <https://doi.org/10.1080/13506280344000257>
- Munger, M. P., Solberg, J. L., & Horrocks, K. K. (1999). The relationship between mental rotation and representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. US: American Psychological Association. <https://doi.org/10.1037/0278-7393.25.6.1557>
- Munger, M. P., Solberg, J. L., Horrocks, K. K., & Preston, A. S. (1999). Representational momentum for rotations in depth: Effects of shading and axis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(1), 157–171. <https://doi.org/10.1037/0278-7393.25.1.157>
- Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics*, *58*(3), 351–362. <https://doi.org/10.3758/BF03206811>
- Murai, Y., & Murakami, I. (2016). The flash-lag effect and the flash-drag effect in the same display. *Journal of Vision*, *16*(11), 31. <https://doi.org/10.1167/16.11.31>
- Müsseler, J., Stork, S., & Kerzel, D. (2002). Comparing mislocalizations with moving stimuli: The Fröhlich effect, the flash-lag, and representational momentum. *Visual Cognition*, *9*(1-2), 120–138. <https://doi.org/10.1080/13506280143000359>
- Nagai, M., & Yagi, A. (2001). The pointedness effect on representational momentum. *Memory and Cognition*, *29*(1), 91–99. <https://doi.org/10.3758/BF03195744>
- Nakamoto, H., Mori, S., Ikudome, S., Unenaka, S., & Imanaka, K. (2015). Effects of sport expertise on representational momentum during timing control. *Attention, Perception,*

- & *Psychophysics*, 77(3), 961–971. <https://doi.org/10.3758/s13414-014-0818-9>
- NASA, & Miller, J. D. (2012). slab3d (version 6.6.1) [Computer software]. Retrieved from <http://slab3d.sonisphere.com/>
- Niemi, P., & Näätänen, R. (1981). Foreperiod and simple reaction time. *Psychological Bulletin*, 89(1), 133–162. <https://doi.org/10.1037/0033-2909.89.1.133>
- Nijhawan, R. (1994). Motion extrapolation in catching. *Nature*, 370(6487), 256–257. <https://doi.org/10.1038/370256b0>
- Nijhawan, R. (2008). Visual prediction: Psychophysics and neurophysiology of compensation for time delays. *Behavioral and Brain Sciences*, 31(2), 179–198. <https://doi.org/10.1017/S0140525X08003804>
- Nijhawan, R., & Khurana, B. (Eds.). (2010). *Space and Time in Perception and Action*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511750540>
- Noesselt, T., Bergmann, D., Hake, M., Heinze, H.-J., & Fendrich, R. (2008). Sound increases the saliency of visual events. *Brain Research*, 1220, 157–163. <https://doi.org/10.1016/j.brainres.2007.12.060>
- Noppeney, U. (2021). Perceptual inference, learning, and attention in a multisensory world. *Annual Review of Neuroscience*, 44(1), annurev-neuro-100120-085519. <https://doi.org/10.1146/annurev-neuro-100120-085519>
- O’Callaghan, C. (2008). Object perception: Vision and audition. *Philosophy Compass*, 3(4), 803–829. <https://doi.org/10.1111/j.1747-9991.2008.00145.x>
- Occelli, V., Spence, C., & Zampini, M. (2011). Audiotactile interactions in front and rear space. *Neuroscience and Biobehavioral Reviews*, 35(3), 589–598. <https://doi.org/10.1016/j.neubiorev.2010.07.004>
- Odegaard, B., & Shams, L. (2016). The brain’s tendency to bind audiovisual signals is stable but not general. *Psychological Science*, 27(4), 583–591. <https://doi.org/10.1177/0956797616628860>
- Odegaard, B., Wozny, D. R., & Shams, L. (2016). The effects of selective and divided attention on sensory precision and integration. *Neuroscience Letters*, 614, 24–28. <https://doi.org/10.1016/j.neulet.2015.12.039>
- Oruc, İ., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research*, 43(23), 2451–2468. [https://doi.org/10.1016/S0042-6989\(03\)00435-8](https://doi.org/10.1016/S0042-6989(03)00435-8)
- Oruc, I., Sinnett, S., Bischof, W. F., Soto-Faraco, S., Lock, K., & Kingstone, A. (2008). The effect

- of attention on the illusory capture of motion in bimodal stimuli. *Brain Research*, 1242, 200–208. <https://doi.org/10.1016/j.brainres.2008.04.014>
- Parise, C. V., & Ernst, M. O. (2016). Correlation detection as a general mechanism for multisensory integration. *Nature Communications*, 7(1), 11543. <https://doi.org/10.1038/ncomms11543>
- Parise, C. V., Harrar, V., Ernst, M. O., & Spence, C. (2013). Cross-correlation between auditory and visual signals promotes multisensory integration. *Multisensory Research*, 26(3), 307–316. <https://doi.org/10.1163/22134808-00002417>
- Parise, C. V., & Spence, C. (2009). “When birds of a feather flock together”: Synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLoS ONE*, 4(5). <https://doi.org/10.1371/journal.pone.0005664>
- Parise, C. V., Spence, C., & Ernst, M. O. (2012). When correlation implies causation in multisensory integration. *Current Biology*, 22(1), 46–49. <https://doi.org/10.1016/j.cub.2011.11.039>
- Passamonti, C., Bertini, C., & Làdavas, E. (2009). Audio-visual stimulation improves oculomotor patterns in patients with hemianopia. *Neuropsychologia*, 47(2), 546–555. <https://doi.org/10.1016/j.neuropsychologia.2008.10.008>
- Peiffer, A. M., Mozolic, J. L., Hugenschmidt, C. E., & Laurienti, P. J. (2007). Age-related multisensory enhancement in a simple audiovisual detection task. *NeuroReport*, 18(10), 1077–1081. <https://doi.org/10.1097/WNR.0b013e3281e72ae7>
- Perrott, D. R., Costantino, B., & Ball, J. (1993). Discrimination of moving events which accelerate or decelerate over the listening interval. *The Journal of the Acoustical Society of America*, 93(2), 1053–1057. <https://doi.org/10.1121/1.405553>
- Perrott, D. R., Costantino, B., & Cisneros, J. (1993). Auditory and visual localization performance in a sequential discrimination task. *The Journal of the Acoustical Society of America*, 93(4), 2134–2138. <https://doi.org/10.1121/1.406675>
- Perrott, D. R., & Musicant, A. D. (1977). Minimum auditory movement angle: Binaural localization of moving sound sources. *The Journal of the Acoustical Society of America*, 62(6), 1463. <https://doi.org/10.1121/1.381675>
- Perrott, D. R., & Saberi, K. (1990). Minimum audible angle thresholds for sources varying in both elevation and azimuth. *The Journal of the Acoustical Society of America*, 87(4), 1728–1731. <https://doi.org/10.1121/1.399421>
- Piotrowski, A. S., & Jakobson, L. S. (2011). Representational momentum in older adults. *Brain and Cognition*, 77(1), 106–112. <https://doi.org/10.1016/j.bandc.2011.05.002>

- Pitman, E. J. G. (1939). A note on normal correlation. *Biometrika*, 31(1-2), 9-12.
<https://doi.org/10.1093/biomet/31.1-2.9>
- Poljanšek, A. (2002). The effect of motion acceleration on displacement of continuous and staircase motion in the frontoparallel plane. *Psihološka Obzorja/Horizons of Psychology*, 11(2), 7-21.
- Pöppel, E. (2009). Pre-semantically defined temporal windows for cognitive processing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1525), 1887-1896. <https://doi.org/10.1098/rstb.2009.0015>
- Posner, M. I., & Boies, S. J. (1971). Components of attention. *Psychological Review*, 78(5), 391-408. <https://doi.org/10.1037/h0031333>
- Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: An information-processing account of its origins and significance. *Psychological Review*, 83(2), 157-171. <https://doi.org/10.1037/0033-295X.83.2.157>
- Prigent, E., Amorim, M.-A., & de Oliveira, A. M. (2018). Representational momentum in dynamic facial expressions is modulated by the level of expressed pain: Amplitude and direction effects. *Attention, Perception, & Psychophysics*, 80(1), 82-93. <https://doi.org/10.3758/s13414-017-1422-6>
- Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology*, 9(2), 129-154. <https://doi.org/10.1080/713752551>
- Radeau, M., & Bertelson, P. (1987). Auditory-visual interaction and the timing of inputs. *Psychological Research*, 49(1), 17-22. <https://doi.org/10.1007/BF00309198>
- Recanzone, G. H. (1998). Rapidly induced auditory plasticity: The ventriloquism aftereffect. *Proceedings of the National Academy of Sciences*, 95(3), 869-875. <https://doi.org/10.1073/pnas.95.3.869>
- Recanzone, G. H. (2009). Interactions of auditory and visual stimuli in space and time. *Hearing Research*, 258(1-2), 89-99. <https://doi.org/10.1016/j.heares.2009.04.009>
- Reed, C. L., & Vinson, N. G. (1996). Conceptual effects on representational momentum. *Journal of Experimental Psychology: Human Perception and Performance*, 22(4), 839-850. <https://doi.org/10.1037/0096-1523.22.4.839>
- Reynolds, J. H., & Chelazzi, L. (2004). Attentional modulation of visual processing. *Annual Review of Neuroscience*, 27, 611-647. <https://doi.org/10.1146/annurev.neuro.26.041002.131039>
- Risoud, M., Hanson, J.-N., Gauvrit, F., Renard, C., Lemesre, P.-E., Bonne, N.-X., & Vincent, C. (2018). Sound source localization. *European Annals of Otorhinolaryngology, Head and*

- Neck Diseases*, 135(4), 259–264. <https://doi.org/10.1016/j.anorl.2018.04.009>
- Roach, N. W., Heron, J., & McGraw, P. V. (2006). Resolving multisensory conflict : a strategy for balancing the costs and benefits of audio-visual integration, (June), 2159–2168. <https://doi.org/10.1098/rspb.2006.3578>
- Rock, I. (1966). *The nature of perceptual adaptation*. New York: Basic Books, Inc.
- Rock, I., & Victor, J. (1964). Vision and touch: An experimentally created conflict between the two senses. *Science*, 143(3606), 594–596. <https://doi.org/10.1126/science.143.3606.594>
- Rockland, K. S., & Ojima, H. (2003). Multisensory convergence in calcarine visual areas in macaque monkey. *International Journal of Psychophysiology*, 50(1–2), 19–26. [https://doi.org/10.1016/S0167-8760\(03\)00121-1](https://doi.org/10.1016/S0167-8760(03)00121-1)
- Röder, B., Stock, O., Bien, S., Neville, H., & Rösler, F. (2002). Speech processing activates visual cortex in congenitally blind humans. *European Journal of Neuroscience*, 16(5), 930–936. <https://doi.org/10.1046/j.1460-9568.2002.02147.x>
- Roggerone, V., Vacher, J., Tarlao, C., & Guastavino, C. (2019). Auditory motion perception emerges from successive sound localizations integrated over time. *Scientific Reports*, 9(1), 16437. <https://doi.org/10.1038/s41598-019-52742-0>
- Rohe, T., Ehlis, A. C., & Noppeney, U. (2019). The neural dynamics of hierarchical Bayesian causal inference in multisensory perception. *Nature Communications*, 10(1), 1–17. <https://doi.org/10.1038/s41467-019-09664-2>
- Rohe, T., & Noppeney, U. (2018). Reliability-weighted integration of audiovisual signals can be modulated by top-down attention. *Eneuro*, 5(1), ENEURO.0315-17.2018. <https://doi.org/10.1523/eneuro.0315-17.2018>
- Rohenkohl, G., Coull, J. T., & Nobre, A. C. (2011). Behavioural dissociation between exogenous and endogenous temporal orienting of attention. *PLoS ONE*, 6(1), 1–5. <https://doi.org/10.1371/journal.pone.0014620>
- Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex*, 17(5), 1147–1153. <https://doi.org/10.1093/cercor/bhl024>
- Roth, T., Sokolov, A. N., Messias, A., Roth, P., Weller, M., & Trauzettel-Klosinski, S. (2009). Comparing explorative saccade and flicker training in hemianopia: A randomized controlled study. *Neurology*, 72(4), 324–331. <https://doi.org/10.1212/01.wnl.0000341276.65721.f2>

- Sabel, B. A., Henrich-Noack, P., Fedorov, A., & Gall, C. (2011). Chapter 13 - Vision restoration after brain and retina damage: The “residual vision activation theory.” In A. Green, C. E. Chapman, J. F. Kalaska, & F. Lepore (Eds.), *Enhancing performance for action and perception* (Vol. 192, pp. 199–262). Elsevier. <https://doi.org/https://doi.org/10.1016/B978-0-444-53355-5.00013-0>
- Sambo, C. F., & Forster, B. (2009). An ERP investigation on visuotactile interactions in peripersonal and extrapersonal space: Evidence for the spatial rule. *Journal of Cognitive Neuroscience*, *21*(8), 1550–1559. <https://doi.org/10.1162/jocn.2009.21109>
- Santangelo, V., Belardinelli, M. O., Spence, C., & Macaluso, E. (2009). Interactions between voluntary and stimulus-driven spatial attention mechanisms across sensory modalities. *Journal of Cognitive Neuroscience*, *21*(12), 2384–2397. <https://doi.org/10.1162/jocn.2008.21178>
- Sarko, D. K., Nidiffer, A. R., Powers III, A. R., Ghose, D., Hillock-dunn, A., Fister, M. C., ... Wallace, M. T. (2012). Spatial and temporal features of multisensory processes: Bridging between animal and human studies. In M. M. Murray & M. T. Wallace (Eds.), *The Neural Bases of Multisensory Processes* (pp. 191–216). Boca Raton, FL: CRC Press/Taylor & Francis. <https://doi.org/10.1201/b11092-15>
- Saygin, A. P., Driver, J., & de Sa, V. R. (2008). In the footsteps of biological motion and multisensory perception: Judgments of audiovisual temporal relations are enhanced for upright walkers. *Psychological Science*, *19*(5), 469–475. <https://doi.org/10.1111/j.1467-9280.2008.02111.x>
- Scheier, C. R., Nijhawan, R., & Shimojo, S. (1999). Sound alters visual temporal resolution. *Investigative Ophthalmology & Visual Science*, *40*, 4169.
- Schmiedchen, K., Freigang, C., Nitsche, I., & RübSamen, R. (2012). Crossmodal interactions and multisensory integration in the perception of audio-visual motion - A free-field study. *Brain Research*, *1466*, 99–111. <https://doi.org/10.1016/j.brainres.2012.05.015>
- Schmiedchen, K., Freigang, C., RübSamen, R., & Richter, N. (2013). A comparison of visual and auditory representational momentum in spatial tasks. *Attention, Perception, & Psychophysics*, *75*(7), 1507–1519. <https://doi.org/10.3758/s13414-013-0495-0>
- Schorr, E. A., Fox, N. A., Van Wassenhove, V., & Knudsen, E. I. (2005). Auditory-visual fusion in speech perception in children with cochlear implants. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(51), 18748–18750. <https://doi.org/10.1073/pnas.0508862102>
- Schouten, B., Troje, N. F., Vroomen, J., & Verfaillie, K. (2011). The effect of looming and

- receding sounds on the perceived in-depth orientation of depth-ambiguous biological motion figures. *PLoS ONE*, 6(2), e14725. <https://doi.org/10.1371/journal.pone.0014725>
- Schroeder, C. E., & Foxe, J. (2005). Multisensory contributions to low-level, “unisensory” processing. *Current Opinion in Neurobiology*, 15(4), 454–458. <https://doi.org/10.1016/j.conb.2005.06.008>
- Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences*, 12(3), 106–113. <https://doi.org/10.1016/j.tics.2008.01.002>
- Seinfeld, S., Feuchtner, T., Maselli, A., & Müller, J. (2020). User representations in human-computer interaction. *Human-Computer Interaction*, 1–39. <https://doi.org/10.1080/07370024.2020.1724790>
- Seitz, A. R., Kim, R., & Shams, L. (2006). Sound facilitates visual learning. *Current Biology*, 16(14), 1422–1427. <https://doi.org/10.1016/j.cub.2006.05.048>
- Seitz, A. R., Kim, R., van Wassenhove, V., & Shams, L. (2007). Simultaneous and independent acquisition of multisensory and unisensory associations. *Perception*, 36(10), 1445–1453. <https://doi.org/10.1068/p5843>
- Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature*, 385(6614), 308–308. <https://doi.org/10.1038/385308a0>
- Senior, C., Barnes, J., & David, A. S. (2001). Mental imagery increases representational momentum: preliminary findings. *Journal of Mental Imagery*, 25, 177–184.
- Senior, C., Barnes, J., Giampietroc, V., Simmons, A., Bullmore, E. T., Brammer, M., & David, A. S. (2000). The functional neuroanatomy of implicit-motion perception or ‘representational momentum.’ *Current Biology*, 10(1), 16–22. [https://doi.org/10.1016/S0960-9822\(99\)00259-6](https://doi.org/10.1016/S0960-9822(99)00259-6)
- Shamma, S. A., Elhilali, M., & Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences*, 34(3), 114–123. <https://doi.org/10.1016/j.tins.2010.11.002>
- Shams, L. (2012). Early integration and Bayesian causal inference in multisensory perception. In M. M. Murray & M. T. Wallace (Eds.), *The Neural Bases of Multisensory Processes* (pp. 217–231). Boca Raton, FL: CRC Press/Taylor & Francis.
- Shams, L., & Beierholm, U. R. (2010). Causal inference in perception. *Trends in Cognitive Sciences*, 14(9), 425–432. <https://doi.org/10.1016/j.tics.2010.07.001>
- Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature*,

- 408(6814), 788–788. <https://doi.org/10.1038/35048669>
- Shams, L., & Kim, R. (2010). Crossmodal influences on visual perception. *Physics of Life Reviews*, 7(3), 269–284. <https://doi.org/10.1016/j.plrev.2010.04.006>
- Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences*, 12(11), 411–417. <https://doi.org/10.1016/j.tics.2008.07.006>
- Shams, L., Wozny, D. R., Kim, R., & Seitz, A. R. (2011). Influences of multisensory experience on subsequent unisensory processing. *Frontiers in Psychology*, 2(OCT), 1–9. <https://doi.org/10.3389/fpsyg.2011.00264>
- Shangari, T. A., Falahi, M., & Bakouie, F. (2015). Multisensory integration using dynamical Bayesian networks, 9(May), 9–11. <https://doi.org/10.3389/fncom.2015.00058>
- Shepard, R. N., & Chipman, S. (1970). Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology*, 1(1), 1–17. [https://doi.org/10.1016/0010-0285\(70\)90002-2](https://doi.org/10.1016/0010-0285(70)90002-2)
- Shepard, R. N., & Cooper, L. A. (Eds.). (1982). *Mental images and their transformations*. Cambridge, MA: MIT Press.
- Shipley, T. (1964). Auditory flutter-driving of visual flicker. *Science*, 145(3638), 1328–1330. <https://doi.org/10.1126/science.145.3638.1328>
- Shomstein, S., & Behrmann, M. (2008). Object-based attention: Strength of object representation and attentional guidance. *Perception & Psychophysics*, 70(1), 132–144. <https://doi.org/10.3758/PP.70.1.132>
- Šimko, J., Aalto, D., Lippus, P., Włodarczak, M., & Vainio, M. (2015). Pitch, perceived duration and auditory biases: Comparison among languages. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*. Glasgow: University of Glasgow. Retrieved from <http://www.icphs2015.info/pdfs/Papers/ICPHS0575.pdf>
- Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport*, 12(1), 7–10. <https://doi.org/10.1097/00001756-200101220-00009>
- Snedecor, G. W., & Cochran, W. G. (1989). *Statistical Methods* (8th ed.). Ames, IA: Iowa State University Press.
- Soto-Faraco, S., Lyons, J., Gazzaniga, M., Spence, C., & Kingstone, A. (2002). The ventriloquist in motion: Illusory capture of dynamic information across sensory modalities. *Cognitive Brain Research*, 14(1), 139–146. [https://doi.org/10.1016/S0926-6410\(02\)00068-X](https://doi.org/10.1016/S0926-6410(02)00068-X)
- Soto-Faraco, S., Spence, C., & Kingstone, A. (2004). Cross-modal dynamic capture:

- Congruency effects in the perception of motion across sensory modalities. *Journal of Experimental Psychology: Human Perception and Performance*, 30(2), 330–345. <https://doi.org/10.1037/0096-1523.30.2.330>
- SourceForge. (2010). VirtualDub (version 1.9.11) [Computer software]. Retrieved from <http://virtualdub.sourceforge.net/>
- Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science and Technology*, 28(2), 61–70. <https://doi.org/10.1250/ast.28.61>
- Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, 73(4), 971–995. <https://doi.org/10.3758/s13414-010-0073-7>
- Spence, C. (2013). Just how important is spatial coincidence to multisensory integration? Evaluating the spatial rule. *Annals of the New York Academy of Sciences*, 1296(1), 31–49. <https://doi.org/10.1111/nyas.12121>
- Spence, C. (2015). Cross-modal perceptual organization. In J. Wagemans (Ed.), *The Oxford Handbook of Perceptual Organization* (pp. 639–654). Oxford, England: Oxford University Press.
- Spence, C., & Driver, J. (1998). Auditory and audiovisual inhibition of return. *Perception & Psychophysics*, 60(1), 125–139. <https://doi.org/10.3758/BF03211923>
- Spence, C., & Driver, J. (Eds.). (2004). *Crossmodal space and crossmodal attention*. Oxford: Oxford University Press.
- Spence, C., & Frings, C. (2020). Multisensory feature integration in (and out) of the focus of spatial attention. *Attention, Perception, & Psychophysics*, 82(1), 363–376. <https://doi.org/10.3758/s13414-019-01813-5>
- Spence, C., Senkowski, D., & Röder, B. (2009). Crossmodal processing. *Experimental Brain Research*, 198(2–3), 107–111. <https://doi.org/10.1007/s00221-009-1973-4>
- Spence, C., Shore, D. I., & Klein, R. M. (2001). Multisensory prior entry. *Journal of Experimental Psychology: General*, 130(4), 799–832. <https://doi.org/10.1037/0096-3445.130.4.799>
- Stanford, T. R., Quessy, S., & Stein, B. E. (2005). Evaluating the operations underlying multisensory integration in the cat superior colliculus. *Journal of Neuroscience*, 25(28), 6499–6508. <https://doi.org/10.1523/JNEUROSCI.5095-04.2005>
- Stanford, T. R., & Stein, B. E. (2007). Superadditivity in multisensory integration: putting the computation in context. *NeuroReport*, 18(8), 787–792. <https://doi.org/10.1097/WNR.0b013e3280c1e315>
- Staufenbiel, S. M., van der Lubbe, R. H. J., & Talsma, D. (2011). Spatially uninformative sounds

- increase sensitivity for visual motion change. *Experimental Brain Research*, 213(4), 457–464. <https://doi.org/10.1007/s00221-011-2797-6>
- Stein, B. E. (1998). Neural mechanisms for synthesizing sensory information and producing adaptive behaviors. *Experimental Brain Research*, 123(1–2), 124–135. <https://doi.org/10.1007/s002210050553>
- Stein, B. E. (Ed.). (2012). *The New Handbook of Multisensory Processing*. Cambridge, MA: MIT Press.
- Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. *Journal of Cognitive Neuroscience*, 8(6), 497–506. <https://doi.org/10.1162/jocn.1996.8.6.497>
- Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cognitive neuroscience. Cambridge, MA, US: The MIT Press.
- Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, 9(4), 255–266. <https://doi.org/10.1038/nrn2331>
- Stein, B. E., Stanford, T. R., Ramachandran, R., Perrault, T. J., & Rowland, B. A. (2009). Challenges in quantifying multisensory integration: Alternative criteria, models, and inverse effectiveness. *Experimental Brain Research*, 198(2–3), 113. <https://doi.org/10.1007/s00221-009-1880-8>
- Stein, B. E., Stanford, T. R., & Rowland, B. A. (2014). Development of multisensory integration from the perspective of the individual neuron. *Nature Reviews Neuroscience*, 15(8), 520–535. <https://doi.org/10.1038/nrn3742>
- Stein, B. E., Stanford, T. R., Wallace, M. T., Vaughan, J. W., & Jiang, W. (2004). Crossmodal spatial interactions in subcortical and cortical circuits. In C. Spence & J. Driver (Eds.), *Crossmodal Space and Crossmodal Attention* (pp. 25–50). Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198524861.003.0002>
- Stekelenburg, J. J., & Keetels, M. (2016). The effect of synesthetic associations between the visual and auditory modalities on the Colavita effect. *Experimental Brain Research*, 234(5), 1209–1219. <https://doi.org/10.1007/s00221-015-4363-0>
- Stekelenburg, J. J., & Vroomen, J. (2009). Neural correlates of audiovisual motion capture. *Experimental Brain Research*, 198(2–3), 383–390. <https://doi.org/10.1007/s00221-009-1763-z>
- Stekelenburg, J. J., Vroomen, J., & de Gelder, B. (2004). Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity. *Neuroscience Letters*, 357(3), 163–

166. <https://doi.org/10.1016/j.neulet.2003.12.085>
- Stevenson, R. A., Zemtsov, R. K., & Wallace, M. T. (2012). Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(6), 1517–1529. <https://doi.org/10.1037/a0027339>
- Stine, E. A. L., Wingfield, A., & Myers, S. D. (1990). Age differences in processing information from television news: The effects of bisensory augmentation. *Journal of Gerontology*, *45*(1), P1–P8. <https://doi.org/10.1093/geronj/45.1.P1>
- Stork, S., & Müsseler, J. (2004). Perceived localizations and eye movements with action-generated and computer-generated vanishing points of moving stimuli. *Visual Cognition*, *11*(2–3), 299–314. <https://doi.org/10.1080/13506280344000365>
- Sugita, Y., & Suzuki, Y. (2003). Implicit estimation of sound-arrival time. *Nature*, *421*(6926), 911–911. <https://doi.org/10.1038/421911a>
- Talsma, D. (2015). Predictive coding and multisensory integration: An attentional account of the multisensory mind. *Frontiers in Integrative Neuroscience*, *9*(MAR), 1–13. <https://doi.org/10.3389/fnint.2015.00019>
- Talsma, D., Doty, T. J., & Woldorff, M. G. (2007). Selective attention and audiovisual integration: Is attending to both modalities a prerequisite for early integration? *Cerebral Cortex*, *17*(3), 679–690. <https://doi.org/10.1093/cercor/bhk016>
- Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, *14*(9), 400–410. <https://doi.org/10.1016/j.tics.2010.06.008>
- Talsma, D., Senkowski, D., & Woldorff, M. G. (2009). Intermodal attention affects the processing of the temporal alignment of audiovisual stimuli. *Experimental Brain Research*, *198*(2–3), 313–328. <https://doi.org/10.1007/s00221-009-1858-6>
- Talsma, D., & Woldorff, M. G. (2005). Selective attention and multisensory integration: Multiple phases of effects on the evoked brain activity. *Journal of Cognitive Neuroscience*, *17*(7), 1098–1114. <https://doi.org/10.1162/0898929054475172>
- Tang, X., Wu, J., & Shen, Y. (2016). The interactions of multisensory integration with endogenous and exogenous attention. *Neuroscience & Biobehavioral Reviews*, *61*, 208–224. <https://doi.org/10.1016/j.neubiorev.2015.11.002>
- Targher, S., Micciolo, R., Occelli, V., & Zampini, M. (2017). The role of temporal disparity on audiovisual integration in low-vision individuals. *Perception*, *46*(12), 1356–1370. <https://doi.org/10.1177/0301006617720124>

- Targher, S., Occelli, V., & Zampini, M. (2012). Audiovisual integration in low vision individuals. *Neuropsychologia*, 50(5), 576–582. <https://doi.org/10.1016/j.neuropsychologia.2011.10.021>
- Teramoto, W., Hidaka, S., Gyoba, J., & Suzuki, Y. (2010). Auditory temporal cues can modulate visual representational momentum. *Attention, Perception & Psychophysics*, 72(8), 2215–2226. <https://doi.org/10.3758/APP.72.8.2215>
- Teramoto, W., Hidaka, S., & Sugita, Y. (2018). Auditory biases in visual motion perception. In T. L. Hubbard (Ed.), *Spatial Biases in Perception and Cognition* (pp. 167–179). Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781316651247.012>
- Teramoto, W., Hidaka, S., Sugita, Y., Sakamoto, S., Gyoba, J., Iwaya, Y., & Suzuki, Y. (2012). Sounds can alter the perceived direction of a moving visual object. *Journal of Vision*, 12(3), 11–11. <https://doi.org/10.1167/12.3.11>
- Teramoto, W., Manaka, Y., Hidaka, S., Sugita, Y., Miyauchi, R., Sakamoto, S., ... Suzuki, Y. (2010). Visual motion perception induced by sounds in vertical plane. *Neuroscience Letters*, 479(3), 221–225. <https://doi.org/10.1016/j.neulet.2010.05.065>
- Thornton, I. M. (2002). The onset repulsion effect. *Spatial Vision*, 15(2), 219–243. <https://doi.org/10.1163/15685680252875183>
- Thornton, I. M., & Hayes, A. E. (2004). Anticipating action in complex scenes. *Visual Cognition*, 11(2–3), 341–370. <https://doi.org/10.1080/13506280344000374>
- Thornton, I. M., & Hubbard, T. L. (2002). Representational momentum: New findings, new directions. *Visual Cognition*, 9(1–2), 1–7. <https://doi.org/10.1080/13506280143000430>
- Tinelli, F., Purpura, G., & Cioni, G. (2015). Audio-visual stimulation improves visual search abilities in hemianopia due to childhood acquired brain lesions. *Multisensory Research*, 28(1–2), 153–171. <https://doi.org/10.1163/22134808-00002484>
- Turatto, M., Mazza, V., & Umiltà, C. (2005). Crossmodal object-based attention: Auditory objects affect visual processing. *Cognition*, 96(2), B55–B64. <https://doi.org/10.1016/j.cognition.2004.12.001>
- Tye-Murray, N., Sommers, M., Spehar, B., Myerson, J., & Hale, S. (2010). Aging, audiovisual integration, and the principle of inverse effectiveness. *Ear and Hearing*, 31(5), 636–644. <https://doi.org/10.1097/AUD.0b013e3181dd7ff>
- Uono, S., Sato, W., & Toichi, M. (2014). Reduced representational momentum for subtle dynamic facial expressions in individuals with autism spectrum disorders. *Research in*

- Autism Spectrum Disorders*, 8(9), 1090–1099.
<https://doi.org/10.1016/j.rasd.2014.05.018>
- van Beers, R. J., Sittig, A. C., & Gon, J. J. D. van der. (1999). Integration of proprioceptive and visual position-information: An experimentally supported model. *Journal of Neurophysiology*, 81(3), 1355–1364. <https://doi.org/10.1152/jn.1999.81.3.1355>
- van de Rijt, L. P. H., Roye, A., Mylanus, E. A. M., van Opstal, A. J., & van Wanrooij, M. M. (2019). The Principle of inverse effectiveness in audiovisual speech perception. *Frontiers in Human Neuroscience*, 13(September), 1–15.
<https://doi.org/10.3389/fnhum.2019.00335>
- Van der Burg, E., Cass, J., Olivers, C. N. L., Theeuwes, J., & Alais, D. (2010). Efficient visual search from synchronized auditory signals requires transient audiovisual events. *PLoS ONE*, 5(5), e10664. <https://doi.org/10.1371/journal.pone.0010664>
- Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: Nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 34(5), 1053–1065.
<https://doi.org/10.1037/0096-1523.34.5.1053>
- Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2009). Poke and pop: Tactile-visual synchrony increases visual saliency. *Neuroscience Letters*, 450(1), 60–64.
<https://doi.org/10.1016/j.neulet.2008.11.002>
- van Ee, R., van Boxtel, J. J. A., Parker, A. L., & Alais, D. (2009). Multisensory congruency as a mechanism for attentional control over perceptual selection. *Journal of Neuroscience*, 29(37), 11641–11649. <https://doi.org/10.1523/JNEUROSCI.0873-09.2009>
- van Eijk, R. L. J., Kohlrausch, A., Juola, J. F., & van de Par, S. (2008). Audiovisual synchrony and temporal order judgments: Effects of experimental method and stimulus type. *Perception & Psychophysics*, 70(6), 955–968. <https://doi.org/10.3758/PP.70.6.955>
- van Wassenhove, V., Buonomano, D. V., Shimojo, S., & Shams, L. (2008). Distortions of subjective time perception within and across senses. *PLoS ONE*, 3(1).
<https://doi.org/10.1371/journal.pone.0001437>
- van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, 45(3), 598–607.
<https://doi.org/10.1016/j.neuropsychologia.2006.01.001>
- Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the “unity assumption” using audiovisual speech stimuli. *Perception & Psychophysics*, 69(5), 744–756.
<https://doi.org/10.3758/BF03193776>

- Vatakis, A., & Spence, C. (2008). Evaluating the influence of the “unity assumption” on the temporal perception of realistic audiovisual stimuli. *Acta Psychologica*, *127*(1), 12–23. <https://doi.org/10.1016/j.actpsy.2006.12.002>
- Vercillo, T., & Gori, M. (2015). Attention to sound improves auditory reliability in audio-tactile spatial optimal integration. *Frontiers in Integrative Neuroscience*, *9*(MAY), 1–8. <https://doi.org/10.3389/fnint.2015.00034>
- Verfaillie, K., & D’Ydewalle, G. (1991). Representational momentum and event course anticipation in the perception of implied periodical motions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(2), 302–313. <https://doi.org/10.1037/0278-7393.17.2.302>
- Verfaillie, K., & Daems, A. (2002). Representing and anticipating human actions in vision. *Visual Cognition*, *9*(1–2), 217–232. <https://doi.org/10.1080/13506280143000403>
- Verfaillie, K., De Troy, A., & Van Rensbergen, J. (1994). Transsaccadic integration of biological motion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(3), 649–670. <https://doi.org/10.1037/0278-7393.20.3.649>
- Vilares, I., & Körding, K. P. (2011). Bayesian models: the structure of the world, uncertainty, behavior, and the brain. *Annals of the New York Academy of Sciences*, *1224*(1), 22–39. <https://doi.org/10.1111/j.1749-6632.2011.05965.x>
- Vinson, N. G., & Reed, C. L. (2002). Sources of object-specific effects in representational momentum. *Visual Cognition*, *9*(1–2), 41–65. <https://doi.org/10.1080/13506280143000313>
- von der Heydt, R. (2015). Figure-ground organization and the emergence of proto-objects in the visual cortex. *Frontiers in Psychology*, *6*(NOV), 1–10. <https://doi.org/10.3389/fpsyg.2015.01695>
- Vroomen, J., Bertelson, P., & De Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception and Psychophysics*, *63*(4), 651–659. <https://doi.org/10.3758/BF03194427>
- 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. <https://doi.org/10.1037/0096-1523.26.5.1583>
- Vroomen, J., & de Gelder, B. (2004). Temporal ventriloquism: sound modulates the flash-lag effect. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(3), 513–518. <https://doi.org/10.1037/0096-1523.30.3.513>
- Vroomen, J., & Keetels, M. (2006). The spatial constraint in intersensory pairing: No role in

- temporal ventriloquism. *Journal of Experimental Psychology: Human Perception and Performance*, 32(4), 1063–1071. <https://doi.org/10.1037/0096-1523.32.4.1063>
- Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. *Attention, Perception, & Psychophysics*, 72(4), 871–884. <https://doi.org/10.3758/APP.72.4.871>
- Vroomen, J., Keetels, M., de Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive Brain Research*, 22(1), 32–35. <https://doi.org/10.1016/j.cogbrainres.2004.07.003>
- Wallace, M. T., Perrault Jr, T. J., Hairston, W. D., & Stein, B. E. (2004). Visual experience is necessary for the development of multisensory integration. *Journal of Neuroscience*, 24(43), 9580–9584. <https://doi.org/10.1523/JNEUROSCI.2535-04.2004>
- Wallace, M. T., & Stein, B. E. (1996). Sensory organization of the superior colliculus in cat and monkey (pp. 301–311). [https://doi.org/10.1016/S0079-6123\(08\)63337-3](https://doi.org/10.1016/S0079-6123(08)63337-3)
- Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia*, 64, 105–123. <https://doi.org/10.1016/j.neuropsychologia.2014.08.005>
- Wallace, M. T., Wilkinson, L. K., & Stein, B. E. (1996). Representation and integration of multiple sensory inputs in primate superior colliculus. *Journal of Neurophysiology*, 76(2), 1246–1266. <https://doi.org/10.1152/jn.1996.76.2.1246>
- Wallace, M. T., Woynaroski, T. G., & Stevenson, R. A. (2020). Multisensory integration as a window into orderly and disrupted cognition and communication. *Annual Review of Psychology*, 71(1), 193–219. <https://doi.org/10.1146/annurev-psych-010419-051112>
- Warren, D. H., & Cleaves, W. T. (1971). Visual-proprioceptive interaction under large amounts of conflict. *Journal of Experimental Psychology*, 90(2), 206–214. <https://doi.org/10.1037/h0031545>
- Warren, D. H., Welch, R. B., & McCarthy, T. J. (1981). The role of visual-auditory “compellingness” in the ventriloquism effect: Implications for transitivity among the spatial senses. *Perception & Psychophysics*, 30(6), 557–564. <https://doi.org/10.3758/BF03202010>
- Watanabe, K., & Shimojo, S. (2001). When sound affects vision: Effects of auditory grouping on visual motion perception. *Psychological Science*, 12(2), 109–116. <https://doi.org/10.1111/1467-9280.00319>
- Watkins, S., Shams, L., Tanaka, S., Haynes, J.-D., & Rees, G. (2006). Sound alters activity in human V1 in association with illusory visual perception. *NeuroImage*, 31(3), 1247–

1256. <https://doi.org/10.1016/j.neuroimage.2006.01.016>
- Weinbach, N., & Henik, A. (2012). Temporal orienting and alerting – The same or different? *Frontiers in Psychology, 3*(JUL), 1–3. <https://doi.org/10.3389/fpsyg.2012.00236>
- Welch, R. B. (1999). Meaning, attention, and the “unity assumption” in the intersensory bias of spatial and temporal perceptions. In G. Aschersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive Contributions to the Perception of Spatial and Temporal Events* (Vol. 129, pp. 371–387). North-Holland. [https://doi.org/https://doi.org/10.1016/S0166-4115\(99\)80036-3](https://doi.org/https://doi.org/10.1016/S0166-4115(99)80036-3)
- Welch, R. B., DuttonHurt, L. D., & Warren, D. H. (1986). Contributions of audition and vision to temporal rate perception. *Perception & Psychophysics, 39*(4), 294–300. <https://doi.org/10.3758/BF03204939>
- Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88*(3), 638–667. <https://doi.org/10.1037/0033-2909.88.3.638>
- Welch, R. B., & Warren, D. H. (1986). Intersensory interactions. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of Perception and Human Performance* (Vol. 1, pp. 25 – 36). New York: Wiley.
- Werner, S., & Noppeney, U. (2010). Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. *Journal of Neuroscience, 30*(7), 2662–2675. <https://doi.org/10.1523/JNEUROSCI.5091-09.2010>
- White, P. A. (2012). The experience of force: The role of haptic experience of forces in visual perception of object motion and interactions, mental simulation, and motion-related judgments. *Psychological Bulletin, 138*(4), 589–615. <https://doi.org/10.1037/a0025587>
- Whitney, D. (2002). The influence of visual motion on perceived position. *Trends in Cognitive Sciences, 6*(5), 211–216. [https://doi.org/10.1016/S1364-6613\(02\)01887-9](https://doi.org/10.1016/S1364-6613(02)01887-9)
- Willems, R. M. (2009). Body-specific motor imagery of hand actions: neural evidence from right- and left-handers. *Frontiers in Human Neuroscience, 3*(NOV), 1–9. <https://doi.org/10.3389/neuro.09.039.2009>
- Wilson, W. W., & O'Neill, W. E. (1998). Auditory motion induces directionally dependent receptive field shifts in inferior colliculus neurons. *Journal of Neurophysiology, 79*(4), 2040–2062. <https://doi.org/10.1152/jn.1998.79.4.2040>
- Wozny, D. R., Beierholm, U. R., & Shams, L. (2010). Probability matching as a computational strategy used in perception. *PLoS Computational Biology, 6*(8), e1000871.

- <https://doi.org/10.1371/journal.pcbi.1000871>
- Yost, W. A. (1997). The cocktail party problem: Forty years later. In R. H. Gilkey & T. R. Anderson (Eds.), *Binaural and spatial hearing in real and virtual environments* (pp. 329–347). Lawrence Erlbaum Associates, Inc.
- Yost, W. A., Loisel, L., Dorman, M., Burns, J., & Brown, C. A. (2013). Sound source localization of filtered noises by listeners with normal hearing: A statistical analysis. *The Journal of the Acoustical Society of America*, *133*(5), 2876–2882. <https://doi.org/10.1121/1.4799803>
- Yost, W. A., & Zhong, X. (2014). Sound source localization identification accuracy: Bandwidth dependencies. *The Journal of the Acoustical Society of America*, *136*(5), 2737–2746. <https://doi.org/10.1121/1.4898045>
- Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005). Audio-visual simultaneity judgments. *Perception & Psychophysics*, *67*(3), 531–544. <https://doi.org/10.3758/BF03193329>
- Zampini, M., Shore, D. I., & Spence, C. (2003). Audiovisual temporal order judgments. *Experimental Brain Research*, *152*(2), 198–210. <https://doi.org/10.1007/s00221-003-1536-z>
- Zhou, H., Cheung, E. F. C., & Chan, R. C. K. (2020). Audiovisual temporal integration: Cognitive processing, neural mechanisms, developmental trajectory and potential interventions. *Neuropsychologia*, *140*(February), 107396. <https://doi.org/10.1016/j.neuropsychologia.2020.107396>
- Zimmer, U., Itthipanyanan, S., Grent-'t-Jong, T., & Woldorff, M. G. (2010). The electrophysiological time course of the interaction of stimulus conflict and the multisensory spread of attention. *European Journal of Neuroscience*, *31*(10), 1744–1754. <https://doi.org/10.1111/j.1460-9568.2010.07229.x>
- Zimmer, U., Roberts, K. C., Harshbarger, T. B., & Woldorff, M. G. (2010). Multisensory conflict modulates the spread of visual attention across a multisensory object. *NeuroImage*, *52*(2), 606–616. <https://doi.org/10.1016/j.neuroimage.2010.04.245>
- Zmigrod, S., & Hommel, B. (2009). Auditory event files: Integrating auditory perception and action planning. *Attention, Perception & Psychophysics*, *71*(2), 352–362. <https://doi.org/10.3758/APP.71.2.352>
- Zwiers, M. P., Van Opstal, A. J., & Paige, G. D. (2003). Plasticity in human sound localization induced by compressed spatial vision. *Nature Neuroscience*, *6*(2), 175–181. <https://doi.org/10.1038/nn999>

