UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO

MILTON AUGUSTO VENDRAMINI DE ÁVILA

Ouvidos melhores de olhos abertos: efeitos da estimulação multissensorial com estímulo visual não consciente sobre a aprendizagem auditiva

RIBEIRÃO PRETO
2019

MILTON AUGUSTO VENDRAMINI DE ÁVILA

Ouvidos melhores de olhos abertos: efeitos da estimulação multissensorial com estímulo visual não consciente sobre a aprendizagem auditiva

Versão Original

Tese apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo para a obtenção do título de Doutor em Ciências

Orientadores:

Profa. Dra. Cristina Marta Del-Bem
Prof. Dr. João Pereira Leite

Ribeirão Preto
2019

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Ávila, Milton Augusto Vendramini de

Ouvidos melhores de olhos abertos: efeitos da estimulação multissensorial com estímulo visual não consciente sobre a aprendizagem auditiva. Ribeirão Preto, 2019.

130 p. : il. ; 30 cm

Tese de Doutorado, apresentada à Faculdade de Medicina de Ribeirão Preto/USP. Área de concentração: Neurologia.

Orientadores: DelBen, Cristina Marta; Leite, João Pereira.

1. Aprendizagem auditiva. 2. Integração multissensorial. 3. música. 4. Percepção não-consciente. 5. eletroencefalografia.

Título: Ouvidos melhores de olhos abertos: efeitos da estimulação multissensorial com estímulo visual não consciente sobre a aprendizagem auditiva.

Tese apresentada à Faculdade de Medicina de Ribeirão Preto da
Universidade de São Paulo para a obtenção do título de Doutor em
Ciências

Aprovado em:

Banca Examinadora

Prof. Dr.: _____

Instituição: _____

Julgamento: _____

Prof. Dr.: _____

Instituição: _____

Julgamento: _____

Prof. Dr.: _____

Instituição: _____

Julgamento: _____

Prof. Dr.: _____

Instituição: _____

Julgamento: _____

Prof. Dr.: _____

Instituição: _____

Julgamento: _____

Ao meu pai e meu filho.
Os meus princípios.

AGRADECIMENTOS

Agradeço

à Profa. Dra. Cristina Marta Del-Ben, que desde o início abraçou a ideia, apoiou o caminho, e de maneira, sobretudo humana e muito compreensiva, me deu todo suporte e direcionamento para trilhar esta pesquisa.

ao Prof. Dr. João Pereira Leite, cuja contribuição ultrapassa os limites científicos. Um apoio essencial nas escolhas, nos direcionamentos e no suporte estratégico da pesquisa;

à Capes e ao CNPq, pelo suporte financeiro que possibilitou o desenvolvimento deste trabalho, órgãos merecedores de profundo respeito nacional, acima de tudo, dos poderes executivo e legislativo;

ao Prof. Dr. Carlos Simon Garrido e à Profa. Dra. Renata Ferranti Leoni, que abriram gentilmente um espaço no InBRAIn para que pudesse agregar novas ideias a esta pesquisa;

ao Prof. Dr. Oswaldo Baffa Filho, que com muita solicitude e gentileza, dividiu seu espaço de investigação e produção científica para permitir que esta pesquisa pudesse ocorrer;

aos amigos Dr. Lézio Soares Bueno Júnior e Dr. Rafael Naime Ruggiero, que além grandes amizades fora da ciência, contribuíram significativamente para o trabalho com o diverso conhecimento neurocientífico;

aos colegas do laboratório InBrain, que enriqueceram o trabalho com discussões e novas ideias para a pesquisa. Em especial, agradeço os colegas Dr. Danilo Maziero e Cassiano Asano, cuja experiência e presteza no auxílio às coletas foram imprescindíveis;

ao técnico Antonio Renato Meirelles e Silva, que deu todoo suporte técnico e salva-dados, impecável sempre que preciso;

à Rosana Shuhama, pelo auxílio e presteza sempre que preciso, da captação dos sujeitos a questões relacionadas ao comitê de ética.

a todos os sujeitos participantes, que, embora anônimos nesta pesquisa foram peças fundamentais para a construção deste conhecimento;

às minhas irmãs, pelo carinho desde que cheguei;

a Minha mãe, exemplo de dedicação e amor;

à Carol, minha esposa, com quem descubro o mundo a cada dia, e aprendo sobre a ciências da vida e sobre ser humano.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de
Nível Superior - Brasil (CAPES) - Código de Financiamento 001

RESUMO

Ouvidos melhores de olhos abertos: efeitos da estimulação multissensorial com estímulo visual não consciente sobre a aprendizagem auditiva

A integração audiovisual é capaz de melhorar a performance e a aprendizagem unissensorial. Esta integração ocorre mesmo quando uma das informações sensoriais não é consciente ao indivíduo, e.g. informação semântica auditiva pode impactar percepção visual inconsciente. No âmbito da percepção visual não consciente, o fluxo da informação fica restrito ao início do processamento cortical, não alcançando áreas de alta ordem, como a rede parietal-frontal. Considerando que as interações multissensoriais no córtex podem ocorrer desde estágios iniciais do processamento, levantamos a hipótese que a apresentação de estímulos visuais não conscientes pode beneficiar a aprendizagem auditiva de percepção de notas musicais. Neste estudo, testamos sujeitos em um paradigma de aprendizagem auditiva. Os indivíduos tinham como tarefa identificar seis notas diferentes separadas por 50 cents, ou $\frac{1}{4}$ de tom entre si. Os grupos foram divididos de acordo com o treinamento: auditivo (A), Auditivo + visual não consciente congruente (AV), e Auditivo + visual não consciente incongruente (AVi). Eles passaram por testes pré- e pós-treinamento de escolha forçada dentre seis alternativas somente com informação auditiva. Registramos dados de Eletroencefalografia (EEG) ao longo do experimento. A performance foi calculada por meio do Desvio Médio Absoluto das respostas com relação às notas corretas. Resultados comportamentais mostraram que o grupo AV não somente teve uma melhor performance durante o treinamento, quando a informação visual não consciente estava presente, como também apresentou um maior efeito do treinamento (melhora de pré- para pós-treinamento). Ambos os controles não diferiram entre si. Também mostramos que estes efeitos são devidos às notas da extremidade da escala. Os resultados de EEG mostram que a banda theta tem um papel fundamental na melhora da performance através do aumento da sincronização em áreas centrais e temporais, e da conectividade funcional entre córtices visuais e auditivos. Estes resultados são uma demonstração que as

interações multissensoriais ocorridas sem consciente são fortes o suficiente para aumentar a aprendizagem auditiva. Concluimos também que a conectividade funcional entre áreas corticais de processamento inicial aumenta depois de poucos minutos de treinamento. Além disso, esta tese levanta novas possibilidades para protocolos de aprendizagem perceptual auditiva, especialmente aqueles aplicado para aprendizagem de reconhecimento de notas musicais.

Palavras-chave: integração multissensorial, aprendizagem auditiva, percepção musical, córtex sensorial.

ABSTRACT

Better ears with eyes open: effects of multisensory stimulation with nonconscious visual stimulus on auditory learning

Audiovisual integration may improve unisensory perceptual performance and learning. Interestingly, this integration may occur even when one of the sensory modalities is not conscious to the subject, e.g., semantic auditory information may impact nonconscious visual perception. Studies have shown that the flow of nonconscious visual information is mostly restricted to early cortical processing, without reaching higher-order areas, such as the parieto-frontal network. Thus, because multisensory cortical interactions may already occur in early stages of processing, we hypothesized that nonconscious visual stimulation without semantic information might facilitate auditory pitch learning. In this study we used a pitch learning paradigm, in which individuals had to identify six pitches in a scale with constant intervals of 50 cents. Subjects were assigned to one of three training groups: the test group (Auditory + congruent unconscious visual, AV), and two control groups (Auditory only, A, and Auditory + incongruent unconscious visual, AVi). Auditory-only tests were done before and after training in all groups. Electroencephalography (EEG) was recorded throughout the experiment. Results show that the test group (AV, with congruent nonconscious visual stimuli) performed better during the training, and showed a greater improvement from pre- to post-test. Control groups did not differ from one another. Changes in the AV group were mainly due to performances in the first and last pitches of the scale. We also observed consistent EEG patterns associated with this performance improvement in the AV group, especially higher theta-band synchronization among central and temporal areas, and stronger theta-band functional connectivity between visual and auditory cortices. Therefore, we show that nonconscious multisensory interactions are powerful enough to boost auditory perceptual learning, and that increased functional connectivity between early visual and auditory cortices after training might play a role in this effect. Moreover, we

provide a methodological contribution for future studies on auditory perceptual learning, particularly those applied to relative and absolute pitch training.

keywords: multisensory integration, auditory learning, pitch, musical perception, early cortex.

SUMÁRIO

1. INTRODUÇÃO	18
1.1. Percepção musical e aprendizagem auditiva	18
1.2. Integração multissensorial	22
1.2.1. Bases Neurais	28
1.2.2. A aprendizagem perceptual e a integração multissensorial	29
1.3. O processamento não consciente	30
1.4. Proposta	32
2. OBJETIVOS	35
3. Experimento 1 - Definição dos parâmetros	37
3.1. Descrição	37
3.2. Materiais e Métodos	37
3.2.1. Sujeitos	37
3.2.2. Estímulos e equipamentos	38
3.2.3. Protocolo experimental	38
3.2.3.1. <i>Staircase</i>	38
3.2.4. Análise comportamental	45
3.3. Resultados	46
4. Experimento 2 - Aprendizagem auditiva	51
4.1. Descrição	51
4.2. Materiais e métodos	53
4.2.1. Sujeitos	53
4.2.2. Equipamentos	54
4.2.3. Procedimento e estímulos	55
4.2.4. Análise comportamental	65
4.2.5. Aquisição de EEG	66
4.2.6. Pré-processamento do sinal de EEG	68
4.2.7. Análise EEG	69
4.2.7.1. <i>Event Related Potentials</i> (ERP)	69
4.2.7.2. Tempo-frequência	70
4.2.7.3. Conectividade	71
4.3. Resultados	73
4.3.1. Comportamental	73
4.3.2. Eletrofisiologia	79
5. Discussão	91
5.1 Os efeitos da estimulação multissensorial na aprendizagem auditiva	92
5.2 A integração multissensorial nesta tarefa	95
5.3. A plasticidade nesta tarefa	99

6. Limitações do trabalho e direções futuras	106
7. Conclusão	108
8. Referências Bibliográficas	110

1. INTRODUÇÃO

1.1. Percepção musical e aprendizagem auditiva

Após ser apresentado à plateia, o pianista entra no palco. Reverencia o público. Senta-se ao piano. Organiza sua partitura e, concentrado e pronto para executar a peça, fecha a tampa do piano sobre as teclas. Nada mais faz. Após 4 movimentos semelhantes, e nenhuma tecla pressionada, finda-se a peça intitulada 4'33", de John Cage. Uma obra de um artista brilhante, que pode ser minimamente compreendido por sua frase: "Everything we do is music". Esta peça ilustra como o conceito do que é música pode assumir profundidades infinitas. No entanto, dentro da imensidão deste fenômeno, existem elementos chaves para sua construção, apreciação e estudo, por exemplo o ritmo, o timbre, a harmonia, a melodia.

Dentro da percepção auditiva, a melodia, definida como uma sucessão rítmica de sons e silêncios, é estudada de diferentes maneiras. A apreciação da melodia é baseada na capacidade do indivíduo em distinguir uma nota de outra, ou seja, de perceber as diferentes frequências e a relação entre elas. Este desenho formado pela sequência das notas, o contorno melódico, pode ser reproduzido em diferentes tons, mais graves ou mais agudos, e ainda assim ser percebido como um mesmo desenho. Perceber este contorno está relacionado a uma capacidade de identificar a relação entre as notas, mas não necessariamente identificar cada nota isoladamente. Em outras palavras, um indivíduo pode identificar de maneira inequívoca uma sequência de notas que tenha o mesmo contorno que outra sequência, apontando com precisão uma eventual diferença, mas não ser capaz de nomear cada nota. Estudos mostram que a percepção de contorno tem um caráter ontologicamente mais básico que de intervalo, evidenciado pelo fato de que crianças e não

músicos têm maior facilidade em processar a informação de contorno que identificar intervalos de canções não familiares (Trehub, Trainor & Unyk, 1993; Bartlett & Dowling, 1980; Dowling, 1978; Cuddy & Cohen, 1976).

Essa é uma característica essencial daqueles que possuem o chamado ouvido relativo (OR). Uma vez nomeada uma nota de referência, indivíduos possuidores do OR são capazes de nomear outra nota tocada em sequência (Ward & Burns, 1982). Esta capacidade está baseada na identificação de intervalos musicais, cujas características sonoras dentro da escala se mantém independente de qual nota é a referência.

Ainda que o OR configure uma capacidade muito importante e útil para musicistas, a identificação e nomeação de uma única nota qualquer tocada isoladamente, mesmo na ausência do contexto musical ou de uma nota de referência é uma habilidade rara de indivíduos que dizemos ter um ouvido absoluto (OA) (Baggaley, 1974; Takeuchi & Hulse, 1993). Estudo nas últimas duas décadas mostram que esta capacidade está relacionada a alterações funcionais e anatômicas do córtex, como aumento do volume cortical de regiões auditivas (Wengenroth et al., 2013), assimetria entre hemisférios direito e esquerdo no *Planum temporale* (Keenan et al., 2001), um aumento na efetividade da conectividade funcional (Loui et al., 2012), e menor atividade de regiões frontais em julgamento de intervalos musicais (Zatorre et al., 1998).

É discussão antiga se as pessoas são capazes de desenvolver o OA ou se esta é uma condição inata. Uma das questões que desfavorecem a identificação de indivíduos que possuem OA sem o treinamento formal, é a incapacidade teórica de nomeação das notas musicais. Até o início dos anos 2000, os testes se baseavam na identificação da nota tocada dentro de escalas com notas encontradas na música ocidental. Na tentativa de desenvolver um paradigma que fosse independente do treinamento, Ross e colaboradores (2003) propuseram o seguinte método: os sujeitos ouviam uma determinada nota, e com um gerador de ondas, deveriam ajustar a frequência de modo a se igualar à nota ouvida anteriormente. Eles testaram 21 sujeitos, sendo 6 possuidores de OA como atestado por métodos tradicionais, e 15 não possuidores de OA. Eles mostraram que os indivíduos com

OA tinham uma acurácia significativamente maior que outros indivíduos. Em seguida testaram se, mesmo após ouvir uma série de tons distratores subsequentes ao que deveriam identificar, conseguiriam reproduzir. Este teste foi realizado pois evidências apontavam para uma interferência na manutenção destes estímulos na memória operacional, especialmente em indivíduos não possuidores de OA (Pechmann, 1992, Deutsch, 1970). O resultado foi uma performance severamente afetada em indivíduos não possuidores de OA, enquanto que no grupo OA, a performance continuou semelhante à de antes da interferência. De maneira interessante, neste trabalho, eles testaram um indivíduo que se considerava possuidor de OA, mas que praticamente não teve treinamento musical ao longo da vida. O sujeito em questão teve um desempenho semelhante ao grupo com OA, tanto no primeiro, como no segundo experimento, indicando que podem haver casos de ouvido absoluto independente de treinamento no início da vida, ou seja, pode estar relacionado com condições inatas e hereditárias (Ross et al., 2003).

Contudo, é inegável que o processo de aprendizagem é um fator capaz de deflagrar uma capacidade inata, sendo este o caso, ou atingir níveis de excelência em performance, derivados exclusivamente da experiência. Um trabalho interessante que mostra este processo de aprendizagem em diferentes faixas etárias foi realizado por Russo e colaboradores (2005). Eles propuseram um método para avaliar a identificação de uma nota especial, expondo crianças de diferentes idades e adultos a um treinamento. A tarefa consistia em identificar uma nota dentre sete outras, sinalizando sempre que a nota fosse ouvida. O protocolo de treinamento durou seis semanas. Em seus resultados, identificaram que crianças com 5 e 6 anos de idade, obtiveram uma maior performance que crianças mais novas e adultos a partir da terceira semana. Os autores argumentam a favor de uma janela de aprendizagem para o desenvolvimento do ouvido absoluto. Embora este trabalho não seja capaz de dar conta de indivíduos como aquele descrito no trabalho anterior, sem nenhum treino formal, ao menos indica que favorece o desenvolvimento do ouvido absoluto quando trabalhado desde a infância.

Fato é que o treinamento do ouvido musical no início da vida tem um grande impacto futuro, e naturalmente refletirá sua capacidade de identificar e nomear os sons. Os efeitos de um treinamento musical por um prazo moderado a longo podem afetar estruturas cerebrais desde o início do processamento auditivo no tronco encefálico. Um trabalho bastante curioso levanta a questão se o treinamento durante a infância pode se refletir no processamento neural na vida adulta, uma vez que é esperado um declínio na velocidade desta resposta com o envelhecimento (Schwoch et al., 2013). O trabalho indicou que sujeitos que tiveram treinamento por um período entre 4 e 14 anos de música, exibiram uma resposta mais rápida ao processamento neural da fala, mesmo após mais de 40 anos do encerramento de suas atividades musicais.

George & Coch (2001) estudaram o impacto do treinamento musical na memória de trabalho. Eles testaram a memória de músicos e não-músicos e identificaram um melhor desempenho de músicos em memória visual, auditiva, fonológica e executiva. Além disso, identificaram reflexos deste desempenho no sinal de Eletroencefalografia (EEG), especificamente maior amplitude no componente P300 e maior sensibilidade para estímulos infrequentes em paradigma de *oddball*, no qual, dentro de uma série de estímulos frequentes, são apresentados estímulos raros que evocam respostas corticais características (George & Coch, 2011).

De fato, a música está intimamente relacionada à plasticidade neural, evidenciada por marcadores eletrofisiológicos, como *Mismatch Negativity* (MMN) (Fujioka et al., 2004), e trabalhos com imagem por ressonância magnética, revelando toda uma rede distribuída, incluindo áreas auditivas, motoras, e regiões multissensoriais (e.g. Gaser & Schlaug, 2003; Zatorre et al., 2007; Schlaug et al., 2010; Schlaug, 2015). Contudo, estes estudos e revisões, baseiam-se em análises realizadas em indivíduos após anos de prática musical.

Por outro lado, estudos controlados, utilizando um acompanhamento de indivíduos ao longo de um processo de aprendizagem, são capazes de delinear modificações neurais decorrentes deste processo. Recentemente, Schlaug (2015) propôs a música como um modelo para estudar e promover a neuroplasticidade, tal o efeito de sua prática no cérebro.

Em estudo utilizando magnetoencefalografia, Menning e colaboradores (2000), mostraram efeitos do treinamento em marcadores de resposta neural a notas musicais. Utilizando-se de um protocolo de reconhecimento de *deviant* (um estímulo diferente de um estímulo padrão definido na tarefa), eles observaram alterações no MMN (mais precisamente, no componente de campo magnético do MMN, o MMF) e no componente N1, uma voltagem negativa no potencial evocado por volta de 100ms. Em outro trabalho, Seppanen et al., (2012) mostraram a rápida plasticidade do complexo eletrofisiológico N1-P2 (voltagem negativa em torno de 100ms e positiva em torno de 200ms).

Não é por acaso que a música recruta tantas áreas sensoriais e de associação por todo o córtex (Zimmerman & Lahav, 2012). A música não é um fenômeno unissensorial, e, seja na apreciação ou na execução, questões audiovisuais e somatossensoriais estão presentes. Tsay (2013) utiliza o julgamento da performance para mostrar que a apreciação musical está sujeita a fortes influências dos estímulos visuais, a ponto de, em alguns casos, ser mais determinantes que o próprio som praticado pelo(s) músico(s). Na mesma linha, Vines e colaboradores (2010) mostraram os efeitos multissensoriais nas emoções vivenciadas ao apreciar músicas, embora outros trabalhos tenham falhado em demonstrar esta influência nas emoções (Vuoskoski et al., 2016). Outro exemplo da influência por questões visuais, no entanto neste caso em níveis perceptuais, foi mostrado por Schutz (2007), ao apresentar notas longas e curtas tocadas com gestuais diferentes. O trabalho evidencia que a percepção da duração da nota é favorecida e fortemente influenciada pelo gestual.

1.2. Integração multissensorial

Não somente a música. O mundo em que vivemos não é unissensorial. Diariamente somos expostos a estímulos que possuem características multissensoriais quase indissociáveis. Faça você mesmo o exercício de fechar os olhos (após ler esta sentença) e

imaginar a seguinte cena: uma forte chuva caindo do céu, batendo nas folhas das árvores e molhando a terra em seu quintal. Ao fundo, uma lata, cada vez menos vazia recolhendo gota a gota a água caída do telhado. Visualize esta cena nas suas memórias de infância. Faça isso e volte a ler em seguida.

Se você está lendo esta tese com a rara calma que seus dias permitem, é possível que você tenha visto as folhas balançando, ouvido os pingos d'água na lata, sentido o cheiro da terra molhada. Tudo provocado por palavras somadas às suas memórias. O que é mais interessante é que suas memórias foram criadas a partir destas vivências, e a experiência com estas cenas foram naturalmente multissensoriais.

Historicamente, o estudo da integração multissensorial, em especial, a integração audiovisual, tem marcos interessantes. Um deles envolve a percepção de fala: o efeito McGurk, que embora leve o nome de somente um dos pesquisadores, foi descrito por McGurk & MacDonald (1976). Este efeito revela não uma integração, mas a influência do estímulo visual na percepção auditiva. Nele, a apresentação do som de um fonema, por exemplo /ba/ concomitante à apresentação visual de lábios pronunciando o fonema /ga/ resulta na percepção do fonema /da/.

Outro fenômeno estudado décadas atrás é o ventriloquismo (Howard & Templeton, 1966). O ventriloquismo é a arte de fazer com que a voz produzida por um indivíduo pareça ter uma origem diferente, e é datada desde civilizações Romana e Grega (Connor, 2000). Desde seus primeiros estudos no campo da cognição, foi compreendido como um efeito do estímulo visual “capturando” o auditivo (Pick et al., 1969; Wareen et al., 1981). Contudo, como sabemos, existem bons e maus ventríloquos. O que vai diferenciá-los é basicamente a sua capacidade de gerar a ilusão, e esta não é tarefa fácil. É preciso ter um elemento chave: a congruência. Neste caso, primordialmente é a congruência temporal. Se o ventríloquo fizer a voz do boneco sem movimentar a boca da personagem de maneira congruente, dificilmente teremos a ilusão da voz estar vindo do boneco. A congruência temporal é sabidamente um elemento chave para a integração multissensorial, sendo mais

provável a integração quando elementos são apresentados temporalmente alinhados, ou muito próximos (Jones & Jarick, 2006; Shore et al., 2006; van Wassenhove et al., 2007).

Estas são situações que ilustram como nossa relação com o mundo está intimamente ligada à capacidade do cérebro em processar as informações sensoriais provenientes do ambiente. Quando você olha para uma placa de trânsito e, a uma longa distância, consegue ler a que destino leva aquela saída na rodovia, tenha certeza que isso só acontece pela capacidade de precisão de nossos olhos refletida na resolução espacial de frações de um grau de ângulo visual.

Nosso ouvido não possui esta precisão espacial. Com grande dificuldade, seres humanos com capacidades auditivas normais conseguem distinguir fontes sonoras muito próximas uma da outra a uma longa distância. Por outro lado, temporalmente temos uma capacidade excelente de distinguir dois sons.

O que podemos concluir disto é que temos capacidades diferentes de percepção dos elementos da natureza de acordo com cada sentido. O mais interessante, é que trabalhos têm mostrado que nosso cérebro se apoia nas modalidades sensoriais que apresentam capacidades mais apuradas de percepção (Welch & Warren, 1980, Alais & Burr, 2004). Alais e colaboradores estudaram o efeito de ventriloquismo, variando as condições visuais e auditivas de modo que em determinada tarefa, a condição de localização visual era clara. Neste caso, a visão dominava a audição como fonte informativa para a resposta de localização. A outra condição era com o estímulo visual borrado. A audição passava a dominar a visão como principal fonte de confiança para resposta a partir de um alto índice de distorção. Isso indica uma clara dominância da visão para localização espacial sobre a audição. De fato, este fenômeno é facilmente compreendido quando pensamos na capacidade de localização de objetos da visão comparada a precisão encontrada pela audição para fontes sonoras localizadas no espaço (Alais et al., 2004). Por outro lado, a audição tem sido demonstrada “capturar” a visão no efeito do ventriloquismo quando a questão crítica é temporal, e não espacial, revelando uma espécie de “ventriloquismo temporal” liderado pela audição (Morein-Zamir et al., 2003). De fato, estudos

mostram que a congruência espacial, apesar de facilitar a integração (Frens et al., 1995; Slutsky & Recanzone, 2001), não é uma condição *sine qua non* (e.g., Bertelson et al., 1994; Innes-Brown & Crewther, 2009; Jones & Jarick, 2006; Jones & Munhall, 1997; Recanzone, 2003; Vroomen & Keetels, 2006).

Uma maneira clara de perceber isso é irmos ao cinema e não darmos a mínima para o fato de que o som da fala dos atores e atrizes não vem diretamente dos respectivos lábios. O efeito de ventriloquismo. O som sempre vem de caixas de som, em geral nas laterais das salas de cinema ou da sua televisão. Contudo, quando a incongruência é tão pequena quanto 100ms, estudos mostram que a média da população é capaz de perceber a diferença entre som e visual (Vatakis & Spence, 2006), trazendo um incômodo profundo com a fala “atrasada”.

Com tantos anos de evolução em um mundo cuja física temos fortes evidências ter sido a mesma desde nosso surgimento como espécie, e ontologicamente, desde o nascimento vivenciando um mundo multissensorial, nosso cérebro trata alguns elementos como correspondentes de maneira bastante clara. E curiosa. Observe as imagens da Figura 1:

Figura 1. Bouba e Kiki, ou Maluma e Takete. Figuras utilizadas desde a década de 1920 para estudo das correspondências entre visão e audição. A maioria da população associa a bouba ou taluma à figura da esquerda. Fonte: Köhler, W, 1929.

Em 1929, o pesquisador alemão Wolfgang Kohler apresentou a clara associação das figuras a e b às palavras “takete” e “maluma”. Décadas mais tarde, figuras semelhantes e outras pseudopalavras (“bouba” e “kiki”) foram estudadas e encontrou-se que entre 95% e 98% das pessoas associam a palavra “Bouba” à figura curva, da esquerda; e a palavra “kiki” à figura da direita, angulada (Ramachandran & Hubbard’s (2001, 2003)). Estes estudos, feitos com falantes nativos de diversas línguas tem até o momento majoritariamente assumido como uma associação natural entre som e forma. Embora estudo recente revele em análise mais minuciosa uma diferença entre falantes no ocidente e oriente nesta percepção (Chen et al., 2016). Contudo, você, falante nativo do português, deve ter empiricamente vivenciado esta correspondência.

Mais uma demonstração que a língua não é o que determina esta associação, temos que crianças de 4 meses tendem a ter a mesma experiência (Ozturk et al., 2013). Os autores deste trabalho sugerem ainda que, ao contrário, estas correspondências pré-aquisição da fala podem beneficiar no próprio processo.

Estas associações auditivas e visuais são comuns e, como mostrado, espalhadas por toda a população. Indivíduos sinestésicos são grandes expoentes de associações entre os sentidos. Sairíamos do escopo desta tese ao tentar esgotar as associações sinestésicas, mesmo restringindo-nos àquelas audiovisuais. Mas estes indivíduos possuem habilidades realmente incomuns. Com o intuito de compreender com mais detalhes associações audiovisuais nestes indivíduos, diversos trabalhos foram conduzidos estudando elementos de associação entre estas modalidades. Lacey e seus colaboradores (2016) testaram associações entre sons e símbolos (lumu e kike), portanto de alta ordem, e associações “*low-level*” como nota e tamanho do objeto, e nota e elevação do objeto. Os resultados mostraram que os indivíduos sinestésicos estudados possuem uma associação maior entre sons e símbolos, que entre características low-level.

O que é rico ressaltar aqui é a diferença entre os processos de associação. Quando tratamos de associações de alta ordem, implicamos elementos ligados a uma

representação dos objetos para além de suas características básicas. Por exemplo, a representação derivada da forma do objeto da Figura 1, se estende além de características básicas visuais como contraste, inclinação, tamanho ou posição. Por outro lado, a associação de elementos auditivos e visuais básicos, como frequência do som e elevação do objeto no espaço, não implicam necessariamente em uma representação simbólica destes elementos. E estas associações também estão sujeitas à congruência, e além disso, evidenciam quais características visuais encontram pares correspondentes auditivos.

Este conceito de correspondência é bastante importante, pois ela favorece a percepção dos elementos a serem pareados. Utilizarei aqui o conceito como apresentado por Spence (2011): correspondência deve ser entendida como a compatibilidade entre atributos ou dimensões de um estímulo em diferentes modalidades sensoriais.

Restringindo estas características às modalidades visual e auditiva, a gama de correspondências estudadas é extensa. Uma maneira clássica de avaliar este tipo de correspondência é por meio de tarefas em que a resposta deve ser dada o mais rápido possível (*speeded*). A medida neste caso para se compreender o efeito da correspondência, portanto, é o tempo de reação. Em geral, estas tarefas demandam do participante responder para combinações de estímulos congruentes e incongruentes (e.g. Evans & Treisman, 2010). Em Evans e Treisman os participantes tinham como tarefa indicar se a nota era de frequência alta ou baixa (dentre duas opções possíveis). O som era apresentado pareado com o estímulo visual abaixo ou acima da linha horizontal da tela, ou seja, ora congruente, ora incongruente. Os resultados mostram que quando os estímulos estavam pareados de maneira congruente, os TRs eram menores. No trabalho, as autoras testaram os pares “frequência-tamanho”, “frequência-contraste”, “frequência-elevação”, “frequência-frequência espacial”. Os resultados mostraram que houve um mapeamento espontâneo entre todos os pares, com exceção de contraste. Mais detalhadamente, frequências altas são naturalmente pareadas com objetos de tamanhos menores, com posições elevadas no espaço e com alta frequência espacial (Evans & Treisman, 2010).

Outros trabalhos, buscando se esta sorte de pareamentos acontece de maneira inata ou por influências de elementos culturais, testaram algumas correspondências em bebês, de até três meses de idade. As medidas com este grupo de sujeitos foram indiretas, como preferências de olhar, habituação e desabituação dos batimentos cardíacos. Os resultados foram indicativos de um pareamento natural de frequência e elevação, por exemplo (Walker et al., 2010).

De fato, esta é uma das correspondências mais estudadas (Melara & O'Brien, 1987; Ben-Artzi & Marks, 1995; Patching & Quinlan, 2002; Maeda et al., 2004; Widmann et al., 2004; Salgado-Montejo et al., 2016). Contudo, apesar destas demonstrações, inclusive em bebês, Parise (2016) levanta dúvidas sobre este mapeamento natural, que pode ter relações bayesianas de uma vida sendo exposta ao mesmo tipo de relação (Parise, 2016). De qualquer forma, ainda que não haja consenso sobre este mapeamento, tem sido fruto de muitos avanços recentes.

1.2.1. Bases Neurais

Quanto aos aspectos neurais, teorias mais tradicionais em torno das redes envolvidas sugerem que a integração multissensorial ocorre somente em áreas corticais de associação, distinguindo claramente regiões unissensoriais de regiões de associação (e.g. Mesulam, 1998). Este pensamento é derivado de uma visão modular da percepção, que considera áreas unimodais e áreas multissensoriais do córtex. Dentre as últimas, destacam-se o Sulco Temporal Superior (STS) (Beauchamp et al., 2004; Powers et al., 2012) e o Sulco Intraparietal (SIP) (Grefkes & Fink, 2005), além de regiões frontais (Calvert, 2001). Entretanto, é crescente o número de evidências em favor de uma visão do córtex com uma natureza multissensorial, ou seja, áreas tidas como unimodais, em verdade respondem a estímulos sensoriais de outras modalidades (Ghazanfar & Schroeder, 2006; Liang et al., 2013), ou mesmo que há comunicação entre áreas corticais primárias em eventos multissensoriais. (Schroeder et al., 2004; Cappe et al., 2005; Foxe & Schroeder, 2005;

Ghazanfar & Schroeder, 2006; Lewis & Noppeney, 2010). Diversos estudos tratam anatomicamente em outros animais a comunicação direta entre córtices unissensoriais visuais e auditivos (Clarke & Innocenti, 1990; Falchier et al., 2002, 2009; Rockland & Ojima, 2003; Cappe & Barone, 2005; Clemo et al., 2008; Ibrahim et al., 2016 (evidências em gatos); Vaudano et al., 1991 (em ratos); Laramée et al., 2011, 2013 (camundongos)). Um destaque interessante é o fato destas conexões serem mais densamente distribuídas em regiões que processam informações da periferia do campo visual (Falchier et al., 2002; Rockland & Ojima, 2003). Entretanto, esta comunicação pode se dar tanto por relações diretas, subcorticais ou por feedback (Tyll, 2011; Proulx et al., 2012, para revisão). Este crescente interesse pela integração multissensorial e a rede neural envolvida, vem trazendo não somente avanços nesta compreensão, mas também acompanhando o desenvolvimento de novos protocolos para o estudo desta integração e suas aplicações na aprendizagem perceptual.

1.2.2. A aprendizagem perceptual e a integração multissensorial

A aprendizagem perceptual consiste no aprimoramento, dependente da experiência, da capacidade de compreender o que ouvimos, vemos, sentimos, saboreamos ou cheiramos (Gold & Watanabe, 2010). A aprendizagem perceptual acontece durante nosso desenvolvimento, ao desafiarmos nossos sentidos, seja de maneira cotidiana, seja em atividades de alta demanda, como o esporte e a música. A maioria da literatura aplicando estimulação multissensorial na aprendizagem perceptual concentra-se no sentido da visão (e.g. King, 2009; Kim et al., 2010; Powers et al., 2012). Um dos primeiros trabalhos a ganhar destaque foi realizado por Seitz e colaboradores (2006). Os autores treinaram dois grupos de indivíduos a identificar o movimento de um conjunto de pontos apresentados em uma tela de computador. Em linhas gerais, a tarefa exigia a percepção do movimento de uma população de pontos, dos quais uma pequena percentagem se movia coerentemente para a

direita ou esquerda. Os grupos foram separados pela presença ou ausência de estímulo auditivo durante os dias de treinamento. Nos resultados, os autores constataram que a estimulação audiovisual beneficiou o processo de aprendizagem, gerando melhor performance ao comparar medidas pré- e pós-treinamento (Seitz et al., 2006).

Sendo a música um fenômeno multissensorial, é de se esperar que a integração multissensorial esteja intimamente relacionada com a aprendizagem. Lappe e seus colaboradores (2008) compararam a plasticidade cortical induzida pela aprendizagem unimodal (auditiva) e multimodal (sensorimotora + auditiva). Os resultados indicaram que o treinamento multimodal refletiu em alterações em marcadores eletrofisiológicos auditivos (MMNm) mais acentuadas que o grupo treinado somente com estímulo auditivo. Já em outro trabalho, investigou-se a plasticidade comparada entre grupos treinados com estímulos audiovisuais e audiovisuais + somatossensoriais. Após o período de treinamento, os autores concluíram que as alterações observadas foram decorrentes especialmente de áreas corticais multissensoriais (Paraskevopoulos et al., 2012).

1.3. O processamento não consciente

Ainda no contexto da aprendizagem perceptual, apesar de ser um tema controverso (Holender & Duscherer, 2004, Kanai et al., 2006; Lin & He, 2009; Opstal et al., 2010), existem evidências em favor da percepção não consciente, como por aspectos emocionais (Tamietto & de Gelder, 2011), semânticos (Greenwald et al., 1996), ou aritméticos (Rusconi et al., 2006). Uma questão foi levantada e respondida de maneira muito elegante por Watanabe et al., (2001). Haveria aprendizagem perceptual mesmo quando o estímulo alvo da aprendizagem não é percebido conscientemente? Neste trabalho, os autores manipulam a consciência visual dos participantes da pesquisa ao oferecer estímulos que competem a atenção do indivíduo. Em uma medida pré-intervenção, o estímulo alvo foi um grupo de pontos em movimento, e a tarefa foi semelhante à apresentada em Seitz et al. (2006). No

entanto, na apresentação entre os testes, os indivíduos deveriam desempenhar uma tarefa atencional com estímulos visuais no centro da tela, enquanto ao redor eram apresentados pontos em movimento, sem que tivessem a consciência reportada da presença dos mesmos. Ainda que não fizesse parte da tarefa principal, esta apresentação melhorou o desempenho nos testes subsequentes, nos quais o estímulo alvo novamente eram os pontos em movimento. Estes resultados evidenciaram ter havido uma aprendizagem perceptual, mesmo que os estímulos não estivessem na consciência visual dos participantes durante as apresentações entre os testes (Watanabe et al., 2001).

Muitos esforços têm sido feitos em busca de esclarecer os mecanismos neurais da consciência visual, entretanto, um consenso parece longe de ser encontrado (Crick & Koch, 2003; Lamme, 2006; Cohen & Dennet, 2011; Fahrenfort & Lamme, 2012). Há estudos que sugerem que este fenômeno ocorre em decorrência de estruturas subcorticais (e.g. Paus, 2000), outros em áreas corticais primárias (Zeki, 2003; Lamme 2003) e outros ainda em áreas de associação como córtex parietal e pré-frontal (Dehaene & Naccache, 2001; Rees et al., 2002). Segundo um modelo conciliador advogado por Dehaene et al. (2006) baseando-se na hipótese do espaço de trabalho global (Baars, 1988), que sugere uma rede neural central responsável por gerir os *inputs* sensoriais de modo a permitir processamentos cognitivos de ordem superior, como memória e consciência, existem três diferentes estágios do processamento: 1) subliminal, 2) pré-consciente, e 3) consciente. No subliminal, ocorre a ativação de áreas primárias, no entanto esta ativação é insuficiente para ingressar em áreas associativas superiores, através das conexões *bottom-up*. Pré-consciente é o processo neural que tem potencialmente a ativação necessária para um evento consciente, mas não o desencadeia por razões como a atenção, e.g. *inattentional blindness* (Mack & Rock, 1998) ou *attentional blink* (Raymond et al., 1992), ou seja não há influências *top-down* suficientes. Mesmo assim, a rede ativada é mais extensa e inclui áreas associativas no córtex temporal. No processamento consciente, estas áreas são ativadas com maior intensidade e a informação flui de áreas sensoriais a outras áreas associativas, atingindo a rede parietal-frontal e córtex cingulado, que posteriormente enviam *feedback* para as áreas do

processamento inicial. A transição entre pré-consciente e consciente é um processo não linear (Dehaene et al., 2006). Com base na dinâmica de ativação do córtex visual, portanto, é possível distinguir duas etapas de ativação cortical diferentes em um estímulo: uma inicial, que em um evento consciente é mais intensa que em um não-consciente (Fisch et al., 2009), e uma tardia, que ocorre somente em um evento consciente (Dehaene et al., 2001; Lamme, 2003). Dessa forma, ao se apresentar um estímulo visual a um indivíduo de maneira degradada de modo a obter seu limiar de detecção consciente, monitorando através de EEG a dinâmica da atividade cortical, é possível diferenciar a ativação do córtex visual para os eventos conscientes e não-conscientes.

1.4. Proposta

No presente trabalho, propomos um protocolo de aprendizagem multissensorial, utilizando como modelo o desenvolvimento do ouvido musical. Mais especificamente, denominamos aqui ouvido musical a habilidade em perceber e distinguir notas musicais. Bermudez & Zatorre (2008) propuseram um método de avaliação de desempenho que consiste em encontrar o desvio médio absoluto (DMA) da nota alvo, desconsiderando-se oitavas, ou seja, quanto a resposta do sujeito está distante da resposta correta, em semitons. Deste modo, em uma performance perfeita o DMA é 0, e em uma randômica, ~ 3 , definindo um score para o ouvido musical. Contudo, o que é uma dificuldade inerente a esta avaliação é que ela depende do treinamento musical, pois sujeitos devem nomear as notas dentre as 12 possíveis na escala ocidental, o que certamente leva a um viés negativo na performance de sujeitos sem o estudo musical formal. Por esta razão, desenvolvemos um protocolo de treinamento para o desenvolvimento do reconhecimento de notas musicais, sem a necessidade de estudo formal. Aos estímulos auditivos, pareamos temporalmente estímulos visuais, de modo a explorar a aprendizagem multissensorial. Entretanto, como expusemos anteriormente, o cérebro tende a se apoiar no sentido de maior precisão,

fazendo com que informações temporais sejam mais confiáveis quando provindas da audição, e espaciais quando da visão. Em uma tarefa difícil, como de estimação de notas musicais, ao ser apresentado um estímulo visual, o participante tenderá a confiar mais no que vê, e destinar sua atenção à visão mais que à audição. Para evitar que isso ocorra, propomos um estímulo visual não consciente, abaixo do limiar de detecção.

Dessa forma, através de uma aprendizagem multissensorial, pudemos desenvolver um paradigma relevante tanto para o treinamento do ouvido musical, quanto para a compreensão de aspectos da integração multissensorial. Especificamente, respondendo às seguintes questões:

1. É possível melhorar a aprendizagem auditiva através do pareamento de um estímulo visual não consciente?
2. Qual a influência do estímulo visual não consciente no processamento auditivo?
3. Quais os efeitos da estimulação audiovisual sobre a plasticidade promovida pela aprendizagem nesta tarefa?

Além de dar um passo em direção a uma perspectiva pouco explorada nas neurociências cognitivas no Brasil, o estudo da percepção multissensorial, esta tese visa a caminhar em direção a um método que possa favorecer esta busca de muitos profissionais da música.

OBJETIVOS

2. OBJETIVOS

Considerando questões ainda abertas tanto sobre a rede neural envolvida na integração audiovisual e na consciência perceptiva, quanto sobre o desenvolvimento do ouvido musical, este projeto destina-se i) ao estudo da aprendizagem perceptual com um treinamento multissensorial para desenvolvimento do ouvido musical; ii) identificar os potenciais benefícios da estimulação audiovisual sobre a aprendizagem auditiva; e iii) à caracterização dos mecanismos e da plasticidade neural envolvida neste processo.

3. Experimento 1 - Definição dos parâmetros

3.1. Descrição

Este experimento foi realizado para estabelecer os critérios a serem utilizados para determinação do estímulo visual no Experimento 2, voltado à aprendizagem auditiva com estimulação audiovisual. Para isso, realizamos o estudo de percepção visual utilizando os estímulos preconizados para o experimento posterior. Obtivemos com isso a relação entre a curva psicométrica da percepção de contraste e a estimação dos limiares de detecção utilizando o método de *staircase* como descrito abaixo. Ainda, com este experimento investigamos evidências da existência de percepção não consciente, refletida na performance perceptual.

3.2. Materiais e Métodos

3.2.1. Sujeitos

Foram convidados 34 estudantes de graduação ou pós-graduação da Universidade de São Paulo, com visão normal ou corrigida, reportando audição normal, que não estavam tomando nenhuma medicação no momento do estudo, e que reportavam nenhum histórico de doenças neurológicas ou psiquiátricas. Dois sujeitos foram excluídos devido a limiares de detecção muito altos. Os participantes restantes foram 17 mulheres e 15 homens (idade: 24.98 ± 5.6 anos). Todos os participantes assinaram termo de consentimento livre e

esclarecido. O protocolo foi aprovado pelo comitê de ética do Hospital das Clínicas da USP de Ribeirão Preto, CAAE 34853614.0.0000.5440.

3.2.2. Estímulos e equipamentos

Todos os estímulos visuais foram apresentados em óculos de realidade virtual Oculus Rift (DK1), cuja resolução é de 1280x800 pixels. O estímulo visual era um pixel, com raio de 0.13° de ângulo visual, apresentado durante dois ciclos de *refresh* da tela ($\sim 33\text{ms}$) nos lados direito ou esquerdo, com excentricidade de 28° na linha média horizontal, em um fundo cinza (RGB = 70,70,70). No centro da tela, era apresentada uma cruz, de 2.8° de ângulo visual como ponto de fixação.

Pareado com o estímulo visual, era apresentado um estímulo auditivo, não informativo, de 1000Hz de frequência, 33ms, 70dB, com rampa de 5ms para *onset* e *offset*. Todo o programa foi escrito em linguagem Python 2.7. As respostas comportamentais foram coletadas utilizando teclado alfanumérico.

3.2.3. Protocolo experimental

3.2.3.1. *Staircase*

O termo *staircase* refere-se, em neurociência cognitiva, a um protocolo padrão dedicado a encontrar limiares perceptuais. É uma ferramenta muito usada, uma vez que elimina a necessidade de encontrar a curva psicométrica completa do indivíduo, imprimindo a agilidade muitas vezes necessária para os procedimentos. Procedimentos de *staircase* consistem em apresentar um estímulo com um nível determinado de uma variável em

particular, e.g. contraste, e variar a próxima apresentação de acordo com a performance do indivíduo. Dessa forma, quando o indivíduo faz uma resposta correta, o estímulo varia de modo a ficar mais difícil de detectá-lo. O oposto se aplica a respostas incorretas. Ao final do procedimento, os contrastes apresentados nas reversões, ou seja, quando, após uma série correta, o indivíduo não detecta o estímulo, ou após uma série incorreta, passa a detectá-lo, são utilizados para computar o limiar. O limiar indica um percentual de acerto naquele contraste, e este percentual pode variar de acordo com as características da staircase.

Nas últimas décadas vários algoritmos foram implementados, sendo os mais usados métodos Bayesianos e métodos de passos de tamanhos fixos (*fixed step sizes* - FSS). Neste trabalho foi utilizada *staircase* do tipo FSS, na qual os passos de acréscimo (*step-up*) de contraste do estímulo são sempre os mesmos, bem como de decréscimo (*step-down*).

A staircase aplicada foi de 2AFC (escolha forçada entre duas alternativas), com a regra *3-down/1-up*, na qual são necessárias três respostas certas para diminuir o contraste (*step-down*), e apenas uma para aumentar (*step-up*). Adotamos esta *staircase* seguindo as recomendações de Garcia-Perez, 1998, em cujo trabalho com uma série de simulações, indicou este ser o método mais consistente, com menor variação e com convergência mais rápida para o limiar, conferindo mais solidez do limiar estimado, em menor tempo (Garcia-Perez, 1998). O trabalho sugere algumas recomendações, dentre as quais destacamos para nossa escolha:

(1) *Usar um step-down menor que step-up, e na razão apropriada para a regra i selecionada:*

Existem diversas regras de *down/up*: *1-down/1-up*, *2-down/1-up*, *3-down/1-up*, e *4down/1-up*. De acordo com Alcalá-Quintana & Garcia-Perez (2007), *staircases* de *3-down/1-up* são a melhor escolha por ter se mostrado mais consistente em simulações computacionais. Para a regra *3-down/1-up*, a razão adequada é $\frac{\text{step-down}}{\text{step-up}} = 0,7393$, que converge em torno de 83,15% de performance. Tendo em vista que, no caso de 2AFC uma

performance ao acaso resulta em 50% de acerto, e um indivíduo capaz de identificar 50% das tentativas, teria uma performance de 75%, assumindo o limiar de 83,15%, admitimos que estimaremos um limiar acima do qual o indivíduo significativamente identificaria os estímulos mais do que 50% das tentativas .

Utilizamos neste estudo:

$$step_up = 0,11 \log units$$

$$step_down = 0,7393 \times step_up = 0,0813 \log units$$

(2) Usar o ponto de partida acima do limiar previsto.

Para nossos cálculos de contraste, utilizamos os valores de RGB, 8 bits, que poderiam variar de 0 a 255. Utilizamos como cor de fundo, 70,70,70. Como tanto o fundo da tela como o estímulo foram sempre apresentados em escala de cinza, reduzimos os valores de RGB para apenas uma variável, no caso do fundo, RGB = 70.

O contraste utilizado foi baseado no contraste de Michelson (Michelson, 1927), definido por :

$$\frac{l_{max} - l_{min}}{l_{max} + l_{min}}, \text{ em } \log units, \text{ onde}$$

l_{min} é o valor de RGB do fundo (bg), e l_{max} é o valor de RGB do estímulo, portanto, l_{max} é o valor do fundo + contraste (c).

Dessa forma, em cada iteração do procedimento da staircase, era calculado:

Valor de michelson (m):

$$m = \log\left(\frac{c}{(c+bg^2)}\right), \text{ em log units}$$

A atualização do contraste (m_{update}) ocorria de acordo com a seguinte equação:

$$m_{update} = m + step, \text{ em log units,}$$

onde *step* poderia ser *step-up* ou *step-down*, de acordo com a regra 3-down/1-up.

Por último, transformamos em RGB novamente para encontrar o contraste (c), utilizado na apresentação do estímulo :

$$c = \frac{(bg \times 2)e^{m_{update}}}{1 - e^{m_{update}}}, \text{ em RGB}$$

Como em experimentos pilotos o contraste médio de cada sujeito não ultrapassou 116 RGB, utilizamos como ponto de partida 120 RGB.

(3) *Estimar o limiar usando a média aritmética dos valores da reversão nas unidades em que os steps são constantes.*

As reversões foram computadas em *log units*.

(4) *Não descarte dados de nenhuma reversão após a segunda.*

Foram computadas as 20 reversões a partir da terceira. O bloco se encerrava após 22 reversões.

De qualquer forma, para os objetivos deste trabalho, seria importante não somente encontrar o limiar, mas garantir um contraste no qual os indivíduos não tivessem

consciência visual da apresentação do estímulo. Tendo em vista uma série de trabalhos que mostram que o cálculo dos limiares está sujeito a variações ligadas não somente à percepção do indivíduo, mas também a vieses de nível decisional (Winawer et al., 2010; Morgan et al., 2012; Garcia-Perez et al., 2013), indicando staircases serem procedimentos matemáticos susceptíveis à biologia, este experimento foi desenhado para minimizar assunções matemáticas que pudessem interferir no procedimento de apresentação de estímulos visuais durante o Experimento 2.

Por isso, o objetivo deste primeiro experimento foi encontrar a relação entre a staircase e a curva psicométrica dos indivíduos, de modo a estabelecer um protocolo para encontrar um limiar de detecção confiável, e garantir a não consciência do indivíduo quando da apresentação do estímulo.

3.2.3.2. Protocolo

O experimento foi dividido em duas partes (Figura 2). Na primeira, foi aplicada *staircase* em três blocos. O número máximo de tentativas em cada bloco era de 130, ou quando o sujeito atingisse 20 reversões após a segunda, descartando-se as duas primeiras. Cada tentativa consistia na apresentação da cruz de fixação por 600-1200ms. Em seguida o estímulo alvo era apresentado em um dos lados, direita ou esquerda. Os sujeitos eram instruídos a dar a resposta, o mais rápido e o mais correto possível, indicando com as setas do teclado a posição em que o estímulo aparecera. O contraste para cada tentativa variava de acordo com os parâmetros da *staircase*, seguindo a regra *3-down/1-up*. Ao fim do bloco era calculada a média das reversões, indicando L1, L2 e L3, respectivamente os limiares dos blocos 1 a 3. Para a determinação do limiar do sujeito (L_m), calculamos a média aritmética de L1, L2 e L3.

Na segunda parte, utilizamos L_m para criar 10 níveis de contrastes diferentes, que variavam de 0 a $2 \times L_m$ (0; 0,25; 0,5; 0,7; 0,85; 1; 1,15; 1,3; 1,5; 2). Este procedimento foi utilizado para criar uma curva psicométrica dos indivíduos e verificar a estimação do limiar

durante a *staircase*. Foram realizados cinco blocos de 80 tentativas, oito para cada contraste. Cada tentativa correspondia à apresentação do alvo, tal como na primeira parte, mas em um dos contrastes de interesse. Após a resposta do sujeito, a cruz de fixação se transformava em uma interrogação, e o sujeito devia descrever a qualidade de sua experiência visual de acordo com a escala:

- (0) nenhuma experiência visual,
- (1) uma vaga experiência visual, e
- (2) uma experiência visual clara.

Esta escala foi utilizada para termos uma resposta mais informativa que somente “*visto*” e “*não visto*” (o que poderíamos *a posteriori* fazer unindo respostas 1 e 2 como “*visto*” e considerando 0 como “*não visto*”). Desta forma, também conseguimos estimar a ocorrência de percepção não consciente, em contrastes nos quais a performance seria significativamente acima do acaso, contudo com julgamento de nenhuma experiência visual.

Figura 2. Protocolo e paradigma Experimento 1. **A.** Estrutura do experimento, dividida em 2 fases. A primeira, de *staircase*, dividida em três blocos, estimando os limiares L1, L2, e L3, cuja média aritmética (L_m) era usada na segunda fase. Na segunda fase os níveis de contraste variavam em função de L_m , e as 400 apresentações, divididas em 5 blocos, eram usadas para traçar a curva psicométrica. **B.** Apresentação e resposta do estímulo na *staircase*. A apresentação do *flash* acontecia após um intervalo variável entre 600 e 1200ms. O estímulo visual era sempre apresentado com um auditivo não informativo. Após a apresentação, o sujeito dava sua resposta o mais rápido e o mais correto possível. **C.** Apresentação e resposta utilizada nas curvas psicométricas. As tentativas eram semelhantes à fase de *staircase*, com exceção do fato de ser demandada uma resposta com relação à experiência visual, que podia variar entre 0 e 2.

3.2.4. Análise comportamental

Na primeira parte, as respostas e os tempos de reação (TR) foram registrados para cada tentativa. O TR era considerado o tempo decorrido desde a apresentação do estímulo até a resposta do indivíduo. Na segunda fase, também registramos o *score* de visibilidade.

Para estimar a curva psicométrica, utilizamos uma função de 4 parâmetros (Sandberg et al., 2011):

$$f(x) = a + \frac{b-a}{1+e^{(c-xd)}}$$

onde a indica o platô inferior, b , o platô superior, c , o ponto de inflexão, e d , a inclinação da curva.

Com a curva psicométrica determinada, analisamos as respostas do julgamento de visibilidade e a performance para cada valor de julgamento, de 0 a 2, estabelecendo um percentual de acerto e de resposta para cada nível de contraste em cada valor. Com a análise quando o julgamento foi “0”, temos evidências de performances significativamente diferentes do acaso e com julgamento de não percepção. Para isso utilizamos teste *t de student* para uma amostra contra a média populacional de 0,5. Seguindo o objetivo de encontrar a relação entre a estimação da *staircase* e a curva, foi estimada para cada indivíduo, o percentual de contraste em relação ao seu limiar no qual o indivíduo tinha a performance acima do acaso, aqui determinado como 60%. Para isso, identificou-se o contraste no qual a performance = 0,6.

3.3. Resultados

A estimação dos limiares seguiu como o esperado para os sujeitos nos três blocos de staircase, como no exemplo de um sujeito na Figura 3.

Figura 3. Blocos de staircase, exemplo de um sujeito. Cada bloco de staircase gerava uma estimação de limiar. O limiar médio (Lm) era composto pela média aritmética dos limiares L1, L2 e L3, dos três blocos respectivamente.

Os limiares médios (Lm) para os 32 sujeitos, estão listados na Tabela 1:

Tabela 01. Limiares de contraste (em RGB) estimados para cada sujeito

Sujeito	S1	S2	S3	S4	S5	S6	S7	S8
Contraste	86.0	102.0	92.0	113.0	121.0	101.0	106.0	103.0
Sujeito	S9	S10	S11	S12	S13	S14	S15	S16
Contraste	92.0	137.0	66.0	58.0	106.0	116.0	90.0	70.0
Sujeito	S17	S18	S19	S20	S21	S22	S23	S24
Contraste	112.0	72.0	110.0	86.0	64.0	94.0	119.0	84.0
Sujeito	S25	S26	S27	S28	S29	S30	S31	S32
Contraste	88.0	80.0	101.0	94.0	97.0	117.0	103.0	83.0
M (SD)							95,72 (17,96)	

A curva psicométrica média de performance, cujos parâmetros foram $a = 0,49$, $b = 0,99$, $c = 5,69$, $d = 0,14$ e de visibilidade, com parâmetros $a = 0,08$, $b = 1,69$, $c = 5,27$, $d = 0,18$, está representada na Figura 4.

Figura 4. Performance e julgamento de visibilidade. Em azul, curva psicométrica da performance de acordo com o nível de contraste. Em vermelho, curva psicométrica do julgamento de visibilidade de acordo com o nível de contraste. A faixa cinza em torno do ponto de inflexão da curva de performance indica os níveis de contrastes que foram utilizados no experimento 2, por terem potencialmente para alguns sujeitos o limiar de detecção com julgamento de visibilidade próximo a 0.

Como esperado, conforme aumentamos o contraste, aumenta-se a performance e também o julgamento de visibilidade. Contudo, nota-se que mesmo em contrastes nos quais a performance é muito próxima de 100%, o julgamento da visibilidade não se reflete em valores máximos. De maneira interessante, o oposto também ocorre, com o julgamento médio de 0,08 quando o estímulo estava ausente. A Tabela 2 ilustra a performance e o percentual de respostas para cada valor de julgamento de visibilidade.

Tabela 2. Percentual de acertos e de respostas para cada valor de experiência visual de acordo com o nível de contraste.

	“0” (nenhuma experiência visual)		“1” (uma vaga experiência visual)		“2” (uma clara experiência visual)	
Contraste	% acerto	% respostas	% acerto	% respostas	% acerto	% respostas
+100%	75%	1,25%	99,69%	25,54%	99,78%	73,20%
+50%	72,72%	2,57%	99,16%	37,34%	100%	60,07%
+30%	74,62%	5,23%	99,48%	45,07%	99,68%	49,68%
+15%	71,72%	11,32%	98,76%	50,46%	98,77%	38,20%
Lm	66,96%	26,01%	97,24%	51,09%	99,65%	22,89%
-15%	58,68%	45%	93,33%	45,70%	95,79%	9,29%
-30%	54,19%	66,27%	89,38%	29,45%	91,07%	4,3%
-50%	51,37%	85%	66,29%	13,9%	64,28%	1,09%
-75%	51,74	89,37%	50,4%	9,76%	63,63%	0,8%
-100%	47,46%	89,21%	51,21%	9,60%	66,66%	1,17%

A partir da Tabela 2 identificamos que em aproximadamente 26% das tentativas no limiar estimado de detecção, os indivíduos julgam não terem tido nenhuma experiência visual. Ainda assim a performance nesta circunstância é significativamente maior que a esperada para o acaso, de 0,5 (one sample t-test, $t = 6,57$, $p < 0,001$), revelando percepção não consciente no limiar.

Contudo, mesmo demonstrada a percepção não consciente na tarefa, como para o Experimento 2 o intuito era encontrar limiares para os quais os indivíduos julgariam não ter

nenhuma experiência visual, o limiar (L_m) não seria um valor seguro de apresentação dos estímulos para uma percepção não consciente, uma vez que em 51% das tentativas, julgaram ter tido uma experiência vaga, e em 22% uma experiência clara do estímulo. De fato, em experimentos pilotos para o Experimento 2, constatamos que a apresentação neste limiar, elicitou experiência visual clara em várias apresentações.

Selecionamos, desta forma, o limiar como limite superior para os testes de detecção visual no Experimento 2, já que nenhum dos sujeitos deste experimento julgou ter visibilidade = 0 na maioria das apresentações. Tendo em vista que o nível de -50% abaixo do limiar tanto a performance quanto o julgamento já se encontravam muito próximos do platô inferior, com performance ao acaso (51%) e visibilidade em valores muito semelhantes aos encontrados quando o estímulo estava ausente, este foi definido como o limite inferior para o referido teste no Experimento 2. É notável que mesmo quando o estímulo era ausente (-100%), o julgamento de visibilidade foi em média = 0,11.

A partir destas curvas, determinamos os contrastes de referência para o experimento 2, marcados na faixa cinza na curva da Figura 4. Esta determinação também foi corroborada pela análise dos valores acima dos quais a performance do indivíduo era acima do acaso, aqui considerado 60% de performance, ($M = 0,65 \pm 0,14$, ou seja, o equivalente a -35% de contraste de $L_m \pm SD$)

4. Experimento 2 - Aprendizagem auditiva

4.1. Descrição

Este experimento visou responder às questões principais deste trabalho.

1. É possível melhorar a aprendizagem auditiva através do pareamento de um estímulo visual não consciente?
2. Qual a influência do estímulo visual não consciente no processamento auditivo?
3. Quais os efeitos da estimulação audiovisual sobre a plasticidade promovida pela aprendizagem nesta tarefa?

Para isso, foi desenvolvido um protocolo de aprendizagem auditiva baseado na identificação de notas desconhecidas pelo sujeito antes do experimento, em uma tarefa de escolha forçada dentre 6 alternativas (6AFC). A questão 1 foi respondida utilizando dados comportamentais relacionados ao processo de aprendizagem nesta tarefa. Para as questões 2 e 3, de caráter neurofisiológico, fizemos o registro de EEG durante todo o experimento, e as respostas foram baseadas nas análises das alterações ocorridas em decorrência do treinamento, bem como durante a apresentação multissensorial. A estrutura do experimento está ilustrada na Figura 5.

Figura 5. Overview do protocolo experimental. **Dia 1**, dividido em 2 fases, sendo a primeira com 3 blocos de *staircase* para estimação do limiar de detecção visual; e a segunda dividida em 5 blocos de apresentação após a estimação do limiar. **Dia 2**, com registro de EEG, dividido em 5 fases. Primeira, um bloco de aferição da performance pré-treino, segundo, três blocos de *staircase* para estimação do limiar, como no Dia 1; terceiro, cinco blocos de treinamento; quarto, aferição pós-treinamento; e quinto, apresentação de estímulos visuais sem atividade de resposta.

4.2. Materiais e métodos

4.2.1. Sujeitos

Foram recrutados 35 sujeitos, não músicos, e que não participaram do experimento anterior, dos quais dois foram excluídos devido a terem reportado claramente ter visto a informação visual, supostamente abaixo do limiar de detecção, apresentada durante o treinamento. Dos 33 sujeitos restantes, 14 eram mulheres. A idade variou de 18 a 39 anos. Todos os sujeitos reportaram ter audição normal, e visão normal ou corrigida para o normal. O protocolo foi aprovado pelo comitê de ética do Hospital das Clínicas da USP de Ribeirão Preto, CAAE 34853614.0.0000.5440.

Para controlar efeitos não somente de treinamento, pois foram recrutados indivíduos que nunca tivessem tido treinamento musical, foi aplicado um questionário para identificar o índice de sofisticação musical (MSI) de cada sujeito (Mullensiefen et al., 2014). Este índice, dentre outros fatores, mostrou-se capaz de identificar em uma escala de 1 a 7, duas variáveis principais que poderiam estar relacionadas com o desempenho nesta tarefa: percepção musical e treinamento musical. Estas variáveis foram calculadas de acordo com os cálculos propostos pelo MSI (Mullensiefen et al., 2014)

Os participantes foram divididos semialeatoriamente em três grupos: Auditivo (A), que receberam treinamento somente com estímulo auditivo, audiovisual (AV), cujo treinamento foi realizado com estímulos auditivos pareados a estímulos visuais não-conscientes congruentes temporal e espacialmente; e outro grupo controle, audiovisual incongruente (AVi), que recebeu treinamento com estímulos auditivos pareados a estímulos visuais não-conscientes congruentes temporalmente, mas incongruentes espacialmente (ver seção 4.2.3., Figura 08).

A divisão dos grupos foi realizada a princípio, aleatoriamente, contudo, o primeiro dia de experimento era determinante para identificar se o indivíduo teria a possibilidade de ser parte de um dos grupos com estimulação audiovisual. Caso o algoritmo falhasse em

identificar o limiar de detecção visual bem definido, o indivíduo era necessariamente colocado no grupo controle A.

Na **Tabela 3** encontramos a descrição dos grupos quanto às variáveis de maior relevância para o trabalho. Nenhuma das comparações entre grupos revelou-se estatisticamente significativa.

Tabela 3. Descrição dos grupos.

	n	idade	percepção musical (1 a 7)	treinamento musical (1 a 7)
A	11	27,7	4,67	2,41
AV	11	25,8	4,27	1,70
AVi	11	26,0	4,20	1,93

4.2.2. Equipamentos

Os estímulos foram gerados em programa de computador escrito na linguagem Python 2.7, sistema operacional Windows XP, placa de vídeo NVidia GTX960. Os estímulos visuais foram apresentados em óculos de realidade virtual Rift DK1, cuja resolução é de 1280x800 pixels. Dessa forma, era garantida a distância entre os olhos e a tela, e consequentemente o ângulo visual, sem nenhum desconforto com relação à postura do indivíduo. A luminância da tela quando apresentada a cor de fundo era 92 cd/m². Os estímulos auditivos eram apresentados em fone *over-the-ear* Audio-Technica modelo ATH-M10.

Para a aquisição de dados de EEG foi utilizado sistema de 128 canais (Brain AMP, Brain Products, Gilching, Alemanha), composto por 2 baterias e 4 amplificadores de sinal compatível com 32 canais cada um. Eletrodos de Ag/AgCl, com resistência interna de 5kOhm, colocados em touca num total de 127 canais + ECG, referência no eletrodo FCz.

4.2.3. Procedimento e estímulos

O experimento foi dividido em dois dias, de acordo com a Figura 5. No primeiro dia, o principal objetivo era identificar o limiar de detecção a ser utilizado no segundo dia de experimento. A determinação do Limiar foi feita utilizando-se da mesma *staircase* do Experimento 1. Entretanto, este procedimento identificava o limiar de apenas dois pontos em cada lado na tela. Uma vez que o interesse se estendia para 6 pontos, foi necessário identificar o limiar para cada um dos pontos. Os pontos eram denominados “*flash*” para os sujeitos, e consistiam de um único pixel apresentado durante dois ciclos de *refresh* da tela a 60Hz, ou seja, por aproximadamente 33ms, localizados de acordo com a Figura 6.

Figura 6. **Estímulos visuais.** Angulação e esquema de apresentação na tela. O raio r da circunferência de referência para os cálculos foi baseado em experimento de padronização, no qual, dentre 33 sujeitos, todos tiveram regulagem horizontal de acordo com a distância interpupilar acima deste valor, garantindo assim que todos os estímulos fossem apresentados no campo visual do indivíduo. A circunferência não era apresentada. As notas centrais 3 e 4, distanciavam-se 9° verticalmente da linha horizontal central. As notas 2 e 5, a 27° , e as notas 1 e 6 a 45° . A distância entre os pontos R e L era estabelecida de acordo com a distância interpupilar individual. O ângulo visual dos estímulos não está em escala. No experimento, correspondia a 1 pixel ($0,13^\circ$).

Como ilustrado na Figura 7, após o procedimento de *staircase*, para a estimação do limiar, os indivíduos passavam por cinco blocos de apresentações dos seis estímulos. Os contrastes sofriam decréscimo de 0% (sem alteração), 10%, 20%, 30%, 40%, 50% ou 100% (ausente) do contraste estimado como limiar. Os sujeitos deviam responder se os estímulos estavam presentes ou ausentes. Estes valores foram determinados com base no Experimento 1, no qual evidenciamos que a -50% do limiar, encontramos o platô inferior de detecção e visibilidade, com performance próxima a níveis de acaso, e visibilidade próxima a “0”.

Após estes cinco blocos, foram calculados os d' (Green & Swets, 1966) referentes a cada uma das seis posições. O d' é um índice de sensibilidade que remonta à Teoria da Detecção do Sinal, que usa números de acertos, alarmes falsos, erros por omissão e rejeições corretas para estabelecer a capacidade de um radar identificar a presença de um

sinal (sensibilidade), bem como sua especificidade, ou seja, não indicando presença de um sinal quando o mesmo não está presente. A neurociência cognitiva se apropriou deste conceito há algumas décadas utilizando o estímulo como sinal. O d' foi calculado como a seguir:

$$d' = Z(H) - Z(FA), \text{ onde}$$

$$H = \frac{\text{númeo de acertos}}{\text{número de tentativas com estímulo presente}},$$

$$FA = \frac{\text{número de alames falsos}}{\text{número de tentativas com estímulos ausentes}}, \text{ e}$$

$Z(p)$, $p \in [0,1]$, é o inverso da função distribuição acumulada de uma distribuição gaussiana.

Depois deste cálculo, os resultados de cada d' para cada contraste em cada posição era comparado a uma curva obtida a partir de 100 mil simulações de um observador com sensibilidade zero e especificidade zero, com respostas aleatórias para a mesma tarefa do experimento. O maior contraste cujo d' fosse menor que o valor correspondente a 0,01 de probabilidade de acordo com a curva das simulações era considerado aquele referência para a determinação do contraste a ser apresentado no treinamento (Bareither et al., 2014). Após este cálculo, o contraste era multiplicado por 0,75, resultando no contraste de apresentação durante o treinamento (Figura 7C).

Figura 7. Desenho experimental e protocolo Dia 1. **A.** Protocolo. A primeira fase era composta por três blocos de staircase para estimação do limiar. **B.** Apresentação dos estímulos em diferentes contrastes para subsidiar a curva psicométrica. Os contrastes eram níveis de contraste de -100%, ou seja, as “catch trials” para estimar a taxa de alarmes falsos, -50% do contraste, -40%, -30%, -20%, -10% e exatamente no limiar. As apresentações ocorreram em todas as 6 posições possíveis. Para cada nível de contraste, estimou-se o d' , de acordo com a Teoria de Detecção do Sinal (Green & Swets, 1966). **C.** Comparação com a distribuição traçada por 100 mil simulações de um observador com zero sensibilidade. O maior contraste cujo d' fosse menor que o valor correspondente a 0,01 de probabilidade de acordo com a curva das simulações era considerado aquele de referência para a determinação do contraste a ser apresentado no treinamento. Após este cálculo, o contraste era multiplicado por 0,75, resultando no contraste de apresentação durante o treinamento.

Alternativas foram estudadas para este procedimento. Uma delas seria fazer uma *staircase* para cada um dos três pares de estímulos, mas o tempo estimado para isso era maior que o ideal para o segundo dia do experimento.

Após o Teste de detecção visual, foi aplicado o questionário para identificar o MSI de cada sujeito.

Antes do início da segunda etapa do experimento, que ocorria no segundo dia, o sujeito era preparado para o EEG. Em seguida, recebia as instruções por escrito, e tinha a oportunidade de esclarecer quaisquer dúvidas sobre o protocolo. As instruções eram reapresentadas na tela dos óculos de realidade virtual oportunamente de acordo com a fase e o bloco ao longo do experimento. O indivíduo era colocado em posição confortável, utilizando os óculos e os fones de ouvido, e tinha total controle entre os blocos ou fases para descansar, retirar os óculos ou simplesmente se posicionar de maneira mais confortável. Por esta razão era solicitado que, após o início de um bloco, permanecesse imóvel na cadeira, evitando assim ruídos no sinal de EEG.

O segundo dia tinha como foco a aprendizagem auditiva, e o protocolo foi elaborado de acordo com a Figura 8:

Figura 8. Desenho experimental e protocolo Dia 2. **A.** Overview. **B** Descrição das apresentações pré e pós. Após um intervalo entre 600 e 1200ms com a cruz de fixação na tela, era apresentado um som durante 33ms, e aguardava-se a resposta na mesma tela inicial. **C.** Paradigma durante o treinamento. Ocorria de maneira semelhante às tentativas pré e pós, acrescido do pareamento da apresentação visual durante o estímulo auditivo, de acordo com as características do grupo. Em detalhe temos as características de cada grupo. Grupo A, estímulos somente auditivos. Grupo AV, audiovisuais congruentes, cuja correspondência está esquematizada no detalhe de AV. Grupo AVi, semelhante a AV, mas com apresentação incongruente de acordo com o detalhe da figura. Em seguida à apresentação, aguardava-se a resposta e um feedback era apresentado na tela. Eram realizadas 420 apresentações divididas em 5 blocos. **D.** Apresentação visual passiva, 30 de cada estímulo, semelhante às fases pré- e pós-treinamento, sem demanda de resposta.

Pré-treino

O teste pré-treinamento foi realizado seguindo o paradigma da FIGURA 8B. Uma tela com uma cruz de fixação visual era apresentada ao sujeito. Após um intervalo variável entre 600ms e 1200ms, randomicamente selecionado pelo computador, um som era apresentado por 33ms. O som foi produzido computacionalmente a partir de onda senoidal pura, apresentado a 70dB, com rampa ascendente inicial e descendente final de 5ms. O som apresentado era um dentre seis opções possíveis, de acordo com a Tabela 4. As frequências foram determinadas de modo a encontrarem-se, de acordo com Plack et al., 2005 em uma frequência identificável, dada a curta duração do som. As frequências, ainda que pertencentes a um espectro comum nas músicas ocidentais, não correspondiam a nenhuma nota da escala temperada, minimizando assim a possibilidade de associação de alguma das seis notas do experimento a outra referência conhecida. O intervalo entre as notas era equivalente a um quarto de tom, ou 50 cents. Como o desenho experimental implicava pouco tempo de treinamento, o objetivo destas escolhas das frequências foi tornar a tarefa bastante difícil para sujeitos não músicos, evitando assim atingirem alto desempenho mesmo após a sessão de treinamento.

Tabela 4. Frequências e seus números correspondentes

1	2	3	4	5	6
1192.1Hz	1227.0Hz	1263.0Hz	1300.0Hz	1338.1Hz	1377.3Hz

Os sujeitos foram instruídos a, após a apresentação da nota, indicar com as teclas de 1 a 6 do teclado qual acreditavam ser a nota ouvida. Portanto, a tarefa consistia de uma escolha forçada com 6 alternativas. Todos os sujeitos foram instruídos a dar a resposta o mais correto e mais rápido possível. Sobre o teclado era posicionada uma máscara de madeira com o espaço vazado precisamente sobre as teclas de interesse para as respostas (Figura 9). Desta forma, o indivíduo deveria posicionar os dedos da mão esquerda: anelar,

médio e indicador, sobre as teclas 1, 2 e 3, respectivamente; e da mão direita: indicador, médio e anelar, sobre as teclas 4, 5 e 6, respectivamente.

Figura 9. Máscara utilizada para guiar o posicionamento das teclas para os sujeitos utilizando óculos de realidade virtual.

Cada resposta do indivíduo era captada e subtraída da nota alvo. O módulo desta subtração gerava o Desvio da resposta correta. O somatório destes módulos resultava no Desvio Médio Absoluto (DMA) (Bermudez et al., 2008), de acordo com a equação abaixo, onde x é a diferença entre a resposta e a nota alvo e n o número de apresentações das notas.

$$DMA = \frac{\sum_{i=1}^n |x_i|}{n}$$

Os TR's foram coletados e correspondiam ao tempo decorrido desde a apresentação do estímulo até a resposta para a apresentação. Antes do início do teste em si, todas as notas eram apresentadas aos sujeitos, enquanto viam-se na tela seus números correspondentes. Em seguida, para familiarizarem-se com a tarefa, ouviam ao total 4 vezes cada nota, aleatoriamente, sendo demandada sua resposta, como seria feito no teste

subsequente. Após esta etapa de familiarização, o teste se iniciava. Eram apresentadas 30 vezes cada nota, num total de 180 apresentações, aleatoriamente.

Confirmação do limiar de detecção

Após o teste pré-treinamento, os sujeitos passavam pela primeira fase do Dia 1, correspondente aos três blocos do procedimento de *staircase*. O objetivo desta etapa era confirmar limiar estimado no primeiro dia. Como esta capacidade visual combinada ao aparato era sujeita a pequenas variações, foi admitida uma variação de até 10% no contraste estimado no segundo dia comparado ao primeiro dia. Quando ocorria uma discrepância maior que 10%, o sujeito era automaticamente passado para o grupo controle A, em cujo treinamento não havia estímulo visual.

Treinamento

O treinamento foi o ponto de separação dos sujeitos em seus respectivos grupos. O grupo A foi treinado somente com estímulo auditivo, AV com estímulo auditivo + visual congruente e AVi, com estímulo auditivo + visual incongruente, de acordo com a Figura 8C. O estímulo visual era apresentado abaixo do limiar de detecção, de acordo com o procedimento descrito anteriormente. A tarefa se assemelhava àquela do teste pré-treino, com duas diferenças: i) a presença do estímulo visual em 86% das apresentações nos grupos AV e AVi, e ii) a presença do feedback, que indicava a resposta correta na tela, 600ms após a resposta do indivíduo. A presença de estímulo visual em somente 86% das tentativas garantiu haver apresentações para serem comparadas no mesmo grupo com relação à presença ou ausência do estímulo visual, possibilitando um questionamento mais claro sobre a influência do mesmo tanto no desempenho quanto no sinal de EEG.

No grupo AV, o estímulo visual era pareado de forma que a nota mais grave correspondesse ao estímulo visual mais abaixo no campo visual, e a mais aguda, ao estímulo mais elevado. No grupo AVi, o pareamento era feito invertendo-se as posições, de

forma que a nota mais grave correspondesse ao estímulo visual mais elevado. A falta de congruência espera-se que diminua a integração das informações visuais e auditivas em nível perceptual, sendo portanto esperados deste grupo uma performance e efeito de treinamento igual ou pior que o controle, e prejudicada comparada ao grupo AV, congruente. Contudo, este grupo não descarta a possibilidade de interação entre estas informações em um nível não consciente, uma vez que a escala, embora invertida, era constante.

O treinamento acontecia em 5 blocos, num total de 420 tentativas, 70 de cada nota. O sujeito tinha a possibilidade de descansar à vontade entre cada bloco. Ao final, era feita a pergunta “Você viu algum *flash* durante o treinamento?” Em caso afirmativo, era perguntada uma estimativa de percentual. de 0-10% até 90%-100%.

Quatro indivíduos reportaram ter visto *flashes*. Destes, um classificou ter visto em 70% a 80%, e após o experimento, seus relatos indicaram claramente terem sido apresentados estímulos acima de seu limiar de detecção. Por essa razão, o sujeito em questão foi excluído do experimento por pertencer ao grupo AV. Outros três indicaram ter visto 0-10%. Destes, apenas um dos que indicaram ter visto entre 0-10% pertencia ao grupo AV, e os demais todos pertenciam, curiosamente, ao grupo A. Portanto, todos foram mantidos em seus grupos.

Pós-treino

O teste pós-treinamento, foi realizado de acordo com o mesmo paradigma do teste pré-treino.

Visual passivo

Nesta última etapa, o sujeito era exposto aos mesmos estímulos visuais apresentados no treinamento. Sua tarefa era permanecer durante 3 min com a atenção voltada à cruz de fixação. Nenhum estímulo auditivo era apresentado e tampouco havia

demanda de respostas. O objetivo desta etapa foi ter uma base da resposta cerebral evocada pelo estímulo visual apresentado durante o treinamento.

Ao final de todo o experimento era perguntado ao sujeito qual o nível de confiança (NC), de 0 a 10, sendo zero, nenhuma confiança, e 10, confiança total, que ele/ela tinha de que suas respostas estavam corretas nos testes pré- e pós-treinamento.

4.2.4. Análise comportamental

As performances foram calculadas para cada um dos blocos de pré-, pós- e treinamento utilizando-se o DMA como variável dependente. Os efeitos do treinamento (ET) foram analisados através da razão entre o os DMA pré- e pós-treinamento, de modo a obtermos percentualmente o aumento da performance. Como as amostras não tinham uma distribuição normal, de acordo com o teste de normalidade de Shapiro-Wilk, os grupos foram comparados através do teste estatístico não paramétrico de Kruskal-Wallis. A análise *post hoc* foi feita através de Wilcoxon-Mann-Whitney. Também foi calculado o Cohen's d, que mede o tamanho do efeito sobre as médias. De acordo com Cohen (1988) e Sawilowsky (2009), pode-se classificar os efeitos como:

Muito pequeno	0.01	(Sawilowsky, 2009)
Pequeno	0.20	(Cohen, 1988)
Médio	0.50	(Cohen, 1988)
Grande	0.80	(Cohen, 1988)
Muito grande	1.20	(Sawilowsky, 2009)
Enorme	2.0	(Sawilowsky, 2009)

Para uma investigação mais aprofundada das performances, foram separadas as notas e calculado o DMA para cada uma delas. A distribuição da performance por nota

gerou uma curva sobre a qual foi aplicado *fit* para equação quadrática. Devido ao visível efeito de melhora nas extremidades (notas 1 e 6), as mesmas foram agrupadas, separando-as do outro grupos de notas (2 a 5). Os ET's foram calculados e comparados entre os grupos para ambas as aglutinações de notas.

Para comparar o efeito da apresentação do estímulo visual sobre a performance durante o treinamento, utilizamos os mesmos testes estatísticos para comparar os três grupos, e em seguida analisar par a par.

Os NCs foram comparados entre os grupos nas fases pré-, pós-treinamento e na razão pré/pós dos NC's. Os TRs foram comparados para cada fase. Como a demanda aos sujeitos era de responderem o mais rapidamente possível, foram descartadas tentativas que excederam 2300ms. Também foram descartadas tentativas cujas respostas foram de tempos menores que 150ms, as quais foram consideradas antecipatórias.

Porque encontramos efeitos sobre os TRs entre os grupos, utilizando Pearson's *r*, testamos se havia uma correlação entre os ETs e as diferenças entre os TRs pré- e pós-treinamento. Também testamos correlações entre os ETs e as variáveis do MSI - percepção e treinamento, e entre os TRs e os NCs.

4.2.5. Aquisição de EEG

A aquisição foi realizada utilizando os 128 canais do sistema, com referência no eletrodo FCz. Toda a preparação do sujeito envolveu a montagem da touca com colocação de gel (ABRALYT HiCl) para abrasão do couro cabeludo, limpeza e diminuição da impedância na condução do sinal para os eletrodos (Figura 10). Esta montagem durou entre 45min e 2h, variando conforme as características e quantidade de cabelo do sujeito. A amostragem do registro foi 5kHz, com filtro passa alta de 0.1Hz. A cada estímulo ou resposta, o programa gerava um trigger via porta paralela identificando a natureza do estímulo e da resposta. Estes triggers eram utilizados para o alinhamento entre a

apresentação do estímulo ou a resposta e o sinal de EEG. Utilizando um microcontrolador Arduino Uno, identificamos um delay de 70ms entre a apresentação do estímulo e o trigger gerado pelo programa. Por esta razão estes 70ms foram corrigidos durante a análise do EEG.

A **B** **C**

Figura 10. A. Setup para experimento. B. Gel em seringas utilizadas para inserir nos eletrodos para contato com o couro cabeludo. C. Sujeito com a montagem preparada da touca de 128 canais.

4.2.6. Pré-processamento do sinal de EEG

Todas as etapas do pré-processamento e análise foram realizadas em Matlab R2014b (The MathWorks) usando rotinas customizadas e complementadas com funções do pacote de rotinas eeglab (Delorme and Makeig, 2004) e chronux (<http://chronux.org/>, Mitra. & Bokil, 2007). Inicialmente os arquivos foram exportados para o formato de arquivo padrão do Matlab utilizando os dados de voltagem ao longo do tempo em todos os canais, bem como os marcadores temporais de resposta comportamental e de estímulo. Os dados foram re-amostrados para 500 Hz utilizando filtro *antialiasing* e novamente filtrados de 0,5 a 50 Hz utilizando um filtro digital FIR com janelamento de Hamming e ordem de 3 vezes a razão da frequência de amostragem pelo frequência passa alta ($3 \times \frac{F_s}{F_{pa}}$), filtrado frente e verso (usando a função *filtfilt*) dessa maneira corrigindo a distorção de fase causado pelo filtro. Após a filtragem foi realizada a inspeção visual para eliminar os canais (eletrodos) que não apresentavam sinal claro de EEG e então foi realizado o re-referenciamento usando a média de todos os eletrodos:

$$Chref_i = Ch_i - \frac{\sum_{i=1}^n Ch_i}{n}$$

onde Ch representa os canais registrados e *i* o número do canal. Posteriormente, os canais eliminados foram interpolados usando a função *eeg_interp* (Delorme & Makeig, 2004). Para a identificação de artefatos não correspondentes a atividade eletroencefalográfica, foi computada a análise de componente independente (função *run_ica*, Delorme & Makeig, 2004) e componentes claramente identificados como artefatos oculomotores, batimento cardíaco, eletromiograma e outros componentes não cerebrais foram removidos. Em média foram removidos 7,45 componentes por sujeito. Após esta análise, a rejeição de tentativas foi realizada semiautomaticamente através do software. As tentativas cujas amplitudes máxima e mínima ultrapassassem os limiares de $\pm 75\mu V$,

respectivamente, foram sinalizadas como ruidosas. Em seguida, através de inspeção visual, estas tentativas foram rejeitadas ou mantidas, de acordo com o ruído identificado. Durante o processo de inspeção visual, outras tentativas eram rejeitadas, se, mesmo com amplitudes menores que os limiares, tivessem claramente artefatos significativos ou característicos de movimento. Foram rejeitadas épocas com TRs menores que 150ms ou maiores que 2300ms. Para ERP, o sinal foi dividido em segmentos de -400ms a 1000ms em relação ao início da apresentação do estímulo. A linha de base foi corrigida pela média do sinal entre -400ms e 0. Em média foram retirados 24,6 épocas por sujeito.

4.2.7. Análise EEG

4.2.7.1. *Event Related Potentials (ERP)*

Os ERP's foram extraídos para cada indivíduo, separando-se as fases do experimento. Na fase pré, os estímulos foram agrupados todos juntos (notas 1 a 6), notas de extremo (1 e 6) e notas intermediárias (2 a 5). Durante o treinamento, nos grupos AV e AVi também foram separados os estímulos somentes auditivos dos audiovisuais. Foram calculados ERP's médios do grupo a partir dos ERP's individuais de cada fase, e, para a identificação da natureza da integração multissensorial, na qual seria feita comparação entre as apresentações multissensoriais (AV_{par}) e a soma das apresentações unissensoriais (A+V).

O estudo dos ERPs foi realizado com comparações nas principais regiões do escalpo. Foram divididas as regiões Frontais (eletrodos: Fz, F1, F2, AFF1h, AFF2h, AFP1, AFP2), centrais (Cz, FC1, FC2, FFC1h, FFC2h, FCC1h, FCC2h, CCP1h, CCP2h), temporais (hemisfério esquerdo: T7, P7, CP5, TP7, TTP7h, TPP7h; hemisfério direito: T8, P8, CP6, TP8, TTP8h, TPP8h), Parietais (Pz, P1, P2, CPP1h, CPP2h, PPO1h, PPO2h),

occipitais (O1,O2, Oz, OI1h, OI2h, O9, O10, IZ). A comparação entre grupos foi feita utilizando ANOVA de duas vias com medidas repetidas (sub-amostrado para 125Hz), e comparações post-hoc de Bonferroni. A análise da integração multissensorial foi feita por meio de teste t da comparação entre AV_{par} e $A+V$, para cada grupo separadamente (AV e AVi).

4.2.7.2. Tempo-frequência

Decomposições tempo-frequência foram realizadas através da convolução por onduletas de Morlet. Para tal foram extraídas épocas ao redor de cada estímulo (início da apresentação), de 2900 a 3500ms peri-estímulo, com janelas maiores acomodando artefatos de borda (vazamento espectral) intrínsecas a convolução por onduleta. A potência foi calculada das séries temporais cruas de EEG e também após aplicar o filtro Laplace. As séries temporais foram convoluídas com uma série de onduletas de Morlet, definidas como uma senóide complexa multiplicada por uma função Gaussiana:

$$e^{-i2\pi t f} e^{-\frac{t^2}{2\sigma^2}},$$

onde t é o tempo, f é a frequência e define a largura ou ciclos de cada frequência $\frac{4}{2\pi f}$ (Trujillo & Allen, 2007). Cinquenta frequências entre 1 e 50 Hz foram extraídas do dado contínuo, utilizando um número de ciclos espaçado logaritmicamente entre 4 (1 Hz) a 12 (50 Hz). A partir do sinal analítico resultante, obtivemos estimativas de potência instantânea (a magnitude do sinal analítico), definidas como séries temporais de potência:

$$p(t) = \{[real(z(t))]^2 + [imag(z(t))]^2\};$$

e a fase (o ângulo de fase) definido como:

$$\theta_t = \arctan\left(\frac{\text{imag}(z(t))}{\text{real}(z(t))}\right).$$

Cada época foi então cortada em comprimento (200 a 1000 ms peri-estímulo).

4.2.7.3. Conectividade

Filtro espacial Laplace

Medidas convencionais do sinal de EEG possuem pouca resolução espacial uma vez que o sinal registrado em cada eletrodo reflete uma combinação de dinâmicas eletrofisiológicas de diferentes regiões cerebrais e camadas corticais (Nunez & Srinivasan, 2005). Uma técnica conhecida para o aumento da resolução espacial consiste na estimativa da Superfície Laplaciana (Surface Laplacian - SL, também conhecida como densidade de fonte de corrente). A SL é um filtro espacial passa-banda usado para localização topográfica e para atenuar a condução por volume em registros eletroencefalográficos, reduzindo o efeito de difusão dos potenciais pelo escalpo, especialmente utilizado para análises de conectividade. O SL calcula a segunda derivada espacial da voltagem entre eletrodos vizinhos, agindo como uma representação de sinal espacialmente melhorada sem referenciamento. A transformação laplaciana destaca as atividades elétricas locais em detrimento da representação das atividades distais (condução por volume). A diminuição dos efeitos de condução de volume pela SL pode revelar dinâmicas locais sutis e também levar a uma caracterização mais precisa das atividades locais durante o cálculo da coerência de longa distância. Para as análises de conectividade, todas as épocas de EEG foram convertidas para densidade de fonte de corrente (CSD) utilizando o método de Perrin e colaboradores (Perrin et al., 1989). Os parâmetros utilizados foram a suavização

(lambda), a ordem Lagrange (número de iterações para o cálculo do polinômio de Legendre) e a ordem da *spline* esférica, ajustados em 10^{-5} , 10, e 4 (Cohen, 2015).

Sincronia de fase

A sincronia de fase entre canais (no inglês, *interchannel phase synchrony*, ICPS) mede a intensidade em que a fase das oscilações entre diferentes eletrodos são semelhantes no tempo para diferentes frequência. O ICPS, conforme descrito por Lachaux e colaboradores (1999) é dado por:

$$ICSP = \left| \frac{1}{N} \sum_{n=1}^N e^{i[\Phi_{jn} - \Phi_{kn}]} \right|,$$

onde N é o número de trials e Φ_{jt} e Φ_{kt} são os ângulos da fase dos eletrodos j e k . Dessa maneira, se a fase dos ângulos de dois eletrodos oscilam em sincronia ao longo do tempo, a diferença entre eles será constante (i.e. não uniformemente distribuída), tendo assim valores próximos de 1.

Para o cálculo da diferença entre os ângulos da fase entre eletrodos foi calculado o espectro cruzado entre dois canais de interesse utilizando o método multi-taper, no qual diferentes janelamentos ortogonais (sequências slepianas) entre si são aplicados ao sinal e é então calculada a FFT (*fast fourier transform*) para cada sinal utilizando uma janela móvel, resultando em valores com variação no tempo e na frequência. Sendo $z_k(f)$ e $z_j(f)$ o resultado da transformada de Fourier complexa das séries temporais $z'_k(t)$ e $z'_j(t)$ do canal k e j , respectivamente, o espectro cruzado é dado como: $S_{kj}(f) = z_k(f)z_j^*(f)$, onde $*$ significa o complexo conjugado. Dessa maneira, foi calculado:

$$ICSP = \left| \frac{1}{M} \sum_{t=1}^M e^{i[\arctg\left(\frac{\text{imag}[S_{kj}(f)]}{\text{real}[S_{kj}(f)]}\right)]} \right|,$$

onde M representa o número de janelas temporais utilizadas no cálculo do $S_{kj}(f)$, sendo que $S_{kj}(f)$ assume um valor diferente para cada janela temporal. Os parâmetros utilizados foram janela de 0.5s com sobreposição de 50% o que resulta num valor de M igual a 25, e produto do tempo-largura de banda de 3 e 5 janelas.

Estatística

Para os valores de potência e ICPS foi realizada a média entre os trials que foi então normalizada pela conversão para uma escala de decibéis (dB):

$$(10 * \log_{10}[p(t)/p(\text{linha de base})]),$$

permitindo uma comparação direta de efeitos através de bandas de frequência. Foi então realizada a média entre os sujeitos de cada grupo e calculada a diferença do período pós-treino com o pré-treino. Para fins de comparação entre grupos foi calculado a potência ou ICPS para banda específicas theta (3-8 Hz). Para testar se os dados calculados possuem uma distribuição normal foi realizado o teste de Kolmogorov-Smirnov e para comparação entre grupos foi utilizada ANOVA de uma via (3 grupos) e teste *t-Student* não pareado (2 grupos).

4.3. Resultados

4.3.1. Comportamental

Para responder à primeira pergunta desta tese:

1. *É possível melhorar a aprendizagem auditiva através do pareamento de um estímulo visual não consciente?* ,

foram analisadas as performances em todas as fases do experimento: pré-treino, treino, e pós-treino (Figura 11A). Quanto menor o DMA, melhor a performance, uma vez que, em média, as respostas se aproximaram mais das respostas corretas. Dessa forma, pode-se observar que todos os grupos melhoraram após o treinamento, ao compararmos a performance pré- e pós-treinamento. Ao compararmos os efeitos do treinamento (ET), mostrado na Figura 11B, não encontramos efeito de grupo significativo (Kruskal Wallis, $H = 3,65$ $p = 0,16$), contudo encontramos diferença significativas nas comparações de AV vs A (Wilcoxon-Mann Whitney, $U = 34,0$, $p < 0,05$). Na comparação entre AV vs AVi, obtivemos valor $p = 0,06$ ($U = 37,0$). Não foram encontradas diferenças significativas entre AVi vs A ($U = 60$, $p = 0,50$).

Figura 11. Performance no experimento. A. Performance separada por blocos. Em destaque a performance durante o treinamento. **B.** Efeito do treinamento. Aumento percentual da performance pré/pós. * $p < 0.05$.

Outra medida relevante para comparar os grupos é o tamanho de efeito, aqui analisado através do *Cohen's d*. As comparações estão mostradas na Tabela 5.

Tabela 5. Tamanho de efeito nas comparações do Efeito do Treinamento entre grupos

		<i>Cohen's d</i>	Interpretação
Intragrupo	$A_{\text{pré}} \times A_{\text{pós}}$	0,41	entre Pequeno e Médio
	$AV_{\text{pré}} \times AV_{\text{pós}}$	1,39	Muito Grande
	$AVi_{\text{pré}} \times AVi_{\text{pós}}$	0,57	entre Médio e Grande
Entre grupos Efeito de pré/pós entre os grupos	$AV \times A$	0,62	entre Médio e Grande
	$AV \times AVi$	0,60	entre Médio e Grande
	$A \times AVi$	0,008	Muito pequeno

Nas conversas com os sujeitos realizadas após o experimento, foi frequente a percepção de melhora no desempenho nas notas de extremidade (1 e 6), em relação às notas centrais (3 e 4). Por esta razão, julgamos pertinente separar as performances de acordo com as notas. O resultado está ilustrado na Figura 12.

Figura 12. Performance nas fases pré- e pós-treinamento. **A.** Curva de performance em função das notas. A linha cinza representa uma performance aleatória. A linha pontilhada representa a fase pré-treinamento, e a linha grossa representa a fase pós-treinamento. A faixa cinza clara marca as notas de extremo 1 e 6. A faixa escura indica as notas de 2 a 5. **B.** Efeito do treinamento para cada agrupamento de notas. Faixa cinza claro: notas 1 e 6. Faixa cinza escuro: notas 2 a 5.

Na Figura 12A, podemos observar a performance pré- e pós-treino para todos os grupos e compará-las a uma performance aleatória. Na Figura 12B, ilustramos o ET para os

grupos de notas intermediárias (2 a 5) e dos extremos (1 e 6). Na comparação das notas intermediárias, nota-se uma melhora muito pequena, e sem diferença significativa (entre os grupos: $p = 0,70$; AVn vs A: $U = 59,0$, $p = 0,47$; AVi vs A: $U = 51,0$, $p = 0,28$; AVn vs AVni: $U = 48,0$, $p = 0,21$). Contudo, a performance para as notas dos extremos é próxima da significância em efeito de grupo ($H = 4,69$, $p = 0,09$), com efeito significativo ao compararmos *post-hoc* A vs AV ($U = 30$, $p = 0,02$), próximo da significância AV vs AVi ($U = 37,0$, $p = 0,06$) e não significativo A vs AVi ($U = 49,0$, $p = 0,23$).

Os TR's coletados ao longo do experimento podem ser observados na Figura 13.

Figura 13. Tempos de Reação ao longo das fases do experimento.

Os TR's não diferiram entre os grupos na fase Pré-treino ($p = 0,23$). Contudo, foi observado um efeito de grupo ao compararmos o pós-treino, sendo o AV significativamente diferente dos demais ($p < 0,01$). É sabido que o tempo de reação está relacionado com o nível de confiança dos sujeitos em suas respostas. A tabela abaixo mostra o julgamento das confianças nas respostas de cada grupo, antes e depois do treinamento.

Tabela 6. Níveis de confiança dos grupos nas fases de teste.

	A	AV	AVi
pré-treino	5,4 ± 1,50	4,1 ± 0,99	5,0 ± 2,42
pós-treino	6,8 ± 0,91	5,9 ± 1,45	7,2 ± 1,16

Quando comparados, não há efeito de grupos entre os NC's pré-treino ($H = 3,36$, $p = 0,18$), tampouco no pós-treino ($H = 4,628$, $p = 0,11$). Em média, todos os grupos aumentaram sua confiança, e não houve diferença significativa neste aumento (pós-pré) ($H = 2,30$, $p = 0,31$).

Contudo, a tendência de menores CRs para o grupo experimental AV poderia explicar os TRs maiores, mesmo não havendo efeito de grupo. Por isso, traçamos a correlação entre TRs e o CRs, e encontramos uma correlação desprezível (*Pearson's $r = -0,22$*), embora próxima da significância ($p = 0,07$).

Com estas informações, pode-se levantar a questão do tempo de reação ter influência nos resultados comportamentais, favorecendo o grupo que teoricamente poderia ter se beneficiado do tempo para resposta. Para investigar esta hipótese, traçamos a correlação de Pearson entre o ET e o RT. Encontramos também uma correlação desprezível e não significativa (*Pearson's $r = 0,29$* , $p = 0,18$).

Outras correlações foram realizadas utilizando dados do MSI (FIGURA 14). Nenhuma correlação mostrou-se significativa. A mais próxima da significância foi entre a performance Pré-treino e o índice de percepção ($-0,29$, $p = 0,09$).

Correlações experiência musical e performance

Figura 14. Correlações entre índice de sofisticação e performance. Nas colunas estão subgrupos do índice (MSI), nas linhas são representadas as fases. Não houve efeito significativo ($p < 0,05$) em nenhuma das correlações.

Para responder à segunda questão,

2. *Qual a influência do estímulo visual não consciente no processamento auditivo?*

comparamos as performances durante os cinco blocos de treinamentos juntos (destaque na Figura 11 A), e encontramos diferenças significativas entre A e AV ($t = 2,09$, $p = 0,03$), e não

encontramos diferenças significativas nas demais comparações (A vs AVi, $t = 1,21$, $p = 0,22$; AV vs AVi, $t = -0,67$, $p = 0,50$). Já em uma análise intragrupo, separando as tentativas com o estímulo visual presente daquelas sem estímulo visual, a comparação não revelou diferenças significativas nos grupos AV ($U = 55,0$, $p = 0,37$), e AVi ($U=59$, $p = 0,47$).

4.3.2. Eletrofisiologia

A primeira análise dos dados de EEG, ainda com relação à nossa questão 2), foi voltada para a confirmação da literatura sobre a adição não linear do sinal evocado pela apresentação multissensorial, ou seja, o par audiovisual (AV_{par}), quando comparada à soma das apresentações unissensoriais ($A+V$). Os resultados mostram efeitos marginais em grupos de eletrodos parietais e centro-frontais para o grupo experimental (Figura 15).

Figura 15. Integração multissensorial subaditiva. A. Eletrodos centrais. À esquerda, localização dos eletrodos no escalpo. À coluna do centro, ERP do grupo AV para apresentação audiovisual ($AV_{(par)}$) e para soma dos ERP's da apresentação auditiva e visual ($A + V_{(soma)}$). À direita, ERP's do grupo AVi ($AVi_{(par)}$) e $A + Vi_{(soma)}$). As barras nos gráficos inseridos representam a média dos valores nas faixas temporais marcadas em cinza no gráfico. **B. Eletrodos parietais.**

Em seguida, analisamos os potenciais evocados durante o treinamento nos diferentes grupos. Foram encontradas diferenças significativas na região parietal entre 190ms e 270ms ($F = 1,195$, $p < 0,01$), sendo a amplitude do sinal em AVi maior que os demais nesta faixa temporal (Figura 16). Demais regiões do escalpo não apresentaram diferenças significativas.

Figura 16. Distribuição dos ERPs pelo escalpo durante o treinamento nos três grupos. Os canais utilizados para computar o respectivo ERP estão marcados em vermelho à esquerda de cada gráfico. A linha pontilhada nos gráficos marca a apresentação do estímulo. As diferenças significativas ($p < 0,01$) são representadas por marcações em linhas pretas acima do eixo x.

As comparações para verificação dos efeitos do treinamento foram realizadas com a análise pré- e pós- treinamento, bem como as diferenças entre estas fases. A Figura 17 mostra estas fases para os eletrodos centrais e occipitais, a Figura 18, os eletrodos frontais e parietais, e a Figura 19, temporais de ambos os hemisférios. Para todas as regiões do escalpo, com exceção da parietal, encontramos uma atenuação no início do processamento, como o complexo N1-P2 evidenciado nos eletrodos centrais (Figura 17A). Nesta atenuação, a diferença relativa ao componente negativo (N1) resulta em valores

positivos, e o no componente P2, a diferença assume valores positivos. Este resultado pode ser observado nos gráficos de diferenças pós-pré (Figuras 17, 18 e 19).

Na análise da plasticidade, em que comparamos estas diferenças entre os grupos, foram encontrados efeitos de grupo nos eletrodos parietais ($F = 1,175$, $p = 0,01$), no início do processamento, em N1 e P2, evidenciando a atenuação (Figura 18B). As maiores diferenças encontradas foram sobretudo nas notas de extremo. Nos eletrodos frontais, observamos efeitos de grupo entre 640ms e 770ms ($F = 1,239$, $p < 0,01$) nas diferenças pós-pré (Figura 18A) Na região frontal também encontramos diferenças durante as fases pré- ($F = 1,244$, $p < 0,01$) e pós-treinamento ($F = 1,21$, $p < 0,01$). Nos eletrodos temporais, não encontramos efeitos de grupo nas diferenças em nenhum dos hemisférios, mas é digno de nota a diferença encontrada somente na fase pós-treinamento no hemisfério esquerdo, em um processamento tardio, entre 530 e 660ms ($F = 1,244$, $p < 0,01$) (Figura 19A).

Figura 17. ERP's durante as fases pré e pós-treino. A. Eletrodos centrais. Os canais utilizados para computar o respectivo ERP estão marcados em vermelho à esquerda de cada gráfico. A linha pontilhada nos gráficos marca a apresentação do estímulo. Coluna da esquerda: todas as notas (1 a 6), coluna central: notas intermediárias (2 a 5), coluna da direita: notas dos extremos (1 e 6). Linha superior: fase pré, linha do meio: fase pós, linha inferior: diferença entre as fases. A atenuação é observada no gráfico de diferença pós-pré, onde valores acima de zero indicam maior sinal na fase pós- A atenuação de valores negativos se reflete em valores positivos no gráfico das diferenças. Não houve diferença significativa em nenhuma das comparações. **B. Eletrodos occipitais.** Não houve diferenças significativas em nenhuma das comparações.

Figura 18. ERP's durante as fases pré e pós-treinamento. A Eletrodos frontais. Os canais utilizados para computar o respectivo ERP estão marcados em vermelho à esquerda de cada gráfico. A linha pontilhada nos gráficos marca a apresentação do estímulo. Coluna da esquerda: todas as notas (1 a 6), coluna central: notas intermediárias (2 a 5), coluna da direita: notas dos extremos (1 e 6). Linha superior: fase pré, linha do meio: fase pós, linha inferior: diferença entre as fases. A atenuação é observada no gráfico de diferença pós-pré, onde valores acima de zero indicam maior sinal na fase pós. A atenuação de valores negativos se reflete em valores positivos no gráfico das diferenças. As diferenças significativas ($p < 0,01$) são representadas por marcações em linhas pretas acima do eixo x. **B. Eletrodos parietais.** As diferenças significativas ($p < 0,01$) são representadas por marcações em linhas pretas acima do eixo x.

Figura 19. ERP's durante as fases pré e pós-treinamento. A Eletrodos temporais hemisfério esquerdo. Os canais utilizados para computar o respectivo ERP estão marcados em vermelho à esquerda de cada gráfico. A linha pontilhada nos gráficos marca a apresentação do estímulo. Coluna da esquerda: todas as notas (1 a 6), coluna central: notas intermediárias (2 a 5), coluna da direita: notas dos extremos (1 e 6). Linha superior: fase pré, linha do meio: fase pós, linha inferior: diferença entre as fases. A atenuação é observada no gráfico de diferença pós-pré, onde valores acima de zero indicam maior sinal na fase pós- A atenuação de valores negativos se reflete em valores positivos no gráfico das diferenças. As diferenças significativas ($p < 0,01$) são representadas por marcações em linhas pretas acima do eixo x. **B. Eletrodos temporais hemisfério direito.** Não houve diferenças significativa em nenhuma das comparações.

A primeira análise de potência das frequências ao longo do tempo mostra um efeito comum a todos dos grupos: sincronização em theta e dessincronização em beta, como mostrado no eletrodo Cz (Figura 20).

Figura 20. Espectrograma dos eletrodos centrais. Nas linhas estão representados os grupos e nas colunas as fases pré- e pós-treinamento. Para todos os grupos, em todas as fases é possível observar uma sincronização em theta, com maior intensidade em torno de 200ms, mas distribuída ao longo de toda a janela temporal; e uma dessincronização em beta, a partir de 250ms.

Após aplicar a transformação laplaciana, buscando identificar com maior precisão os eletrodos de origem do sinal, separamos as frequências theta (3-8Hz), com destaque para 3-5Hz, e beta e analisamos a potência ao longo do tempo. A dessincronização em beta ocorre tanto na fase pré-treino como na pós treino, sem diferenças significativas entre as fases e não está representada nas figuras subsequentes. Contudo, em Cz, observa-se uma diminuição da sincronização em theta na fase pós nos grupos controles. O grupo A apresenta uma sincronização significativamente menor ($p < 0,05$) ao considerarmos todas as notas (Figura 21). Ao separarmos as notas intermediárias e de extremo, ambos os grupos controles apresentam uma dessincronização nas notas intermediárias ($p < 0,05$).

Figura 21. Potência das faixas de frequência 3-5Hz e 4-8Hz (Theta) em pré- e pós-treinamento no eletrodo Cz. Nas colunas estão representados os agrupamentos de notas e nas linhas, os grupos. Linhas vermelhas representam 3-5Hz, e verde 4-8 Hz. Linhas claras são fase pré- e linhas escuras são fase pós-treinamento. As faixas cinzas representam o intervalo de análise de 100ms a 250ms, onde foram encontradas diferenças significativas ($p < 0,05$).

Comparamos em seguida, o efeito do treinamento nestas frequências nos diferentes grupos através da diferença da potência entre pré- e pós-treinamento (Figura 22). Ao compararmos estas diferenças para todas as notas, temos efeitos significativos para grupos nos eletrodos Cz, em 3-5Hz ($p < 0,05$), e T7 ($p < 0,05$), 4-8Hz sendo AV diferente dos demais (Figura 22, coluna da esquerda). A diferença na potência pré vs pós encontrada em AV mostra que a frequência theta não diminuiu ao longo do experimento neste grupo. Novamente testamos para as notas separadamente. Nas notas intermediárias (Figura 22, coluna central), em AV, esta diferença somente se comporta de maneira estatisticamente significativa em T7, na janela temporal de 100-200ms, 4-8Hz. Embora esta queda seja aparentemente mais pronunciada em A, não encontramos diferenças significativas nas notas intermediárias nos eletrodos Cz ou POz. Já nas notas de extremo (Figura 22, coluna

da direita), o efeito de grupo fica evidente e com diferença estatisticamente significativa em Cz ($p < 0,05$), na janela de 200-400ms.

Figura 22. Variação das potências da frequência theta da fase pré para a fase pós-treinamento.

Coluna da esquerda, todas as notas (1 a 6), coluna central notas intermediárias (2 a 5), coluna da direita notas dos extremos (1 e 6). Linha superior Cz, 3-5Hz, linha central POz, 4-8Hz, e linha inferior, T7, 4-8Hz. Cada gráfico exibe dados da potência pré- subtraída de pós-treinamento. A linha pontilhada horizontal demarca o 0, ou seja, acima desta linha temos potências maiores na fase pós. Diferenças significativas ($p < 0,05$) estão marcadas em linha preta acima do eixo x.

Para explorar a relação entre os córtices auditivos e visuais entre si, realizamos a análise de conectividade entre eletrodos posicionados sobre o córtex auditivo (T7 e T8) com demais regiões corticais. Como os resultados anteriores mostraram efeitos claramente distintos entre notas intermediárias e de extremo, fizemos a análise separadamente para cada uma destas condições. Os resultados, mostrados na Figura 23, revelam que após o

treinamento, nas notas de extremo, a conectividade na banda theta aumenta entre o córtex auditivo no hemisfério esquerdo e eletrodos posicionados sobre o córtex visual, nos grupos audiovisuais no início do processamento (janela de 140ms). Este aumento deve-se sobretudo à diminuição da latência do aumento desta conectividade, iniciando-se nesta janela. No teste ANOVA se observou efeito de grupo somente marginal ($F = 3,10$, $p = 0,06$), contudo, ao se comparar os grupos entre si, observa-se efeito significativo entre AV e A ($t = 2,5938$, $p = 0,0178$) e efeito marginal entre AV e AVi ($t = 1,7943$, $p = 0,0879$) (Figura 23A). No hemisfério direito, esta conectividade tem um aumento menor, sem efeitos significativos (ANOVA: $F = 2,079$, $p = 0,14$; AV x A: $t = 1,6844$, $p = 0,1$; AV x AVi: $t = 1,7793$, $p = 0,09$) (Figura 23B). Estes efeitos não são observados nas notas intermediárias.

Figura 23. Conectividade Funcional entre córtices sensoriais. A. Conectividade referente a T7. *Topoplots* representam a conectividade pós-pré: Coluna da esquerda, notas intermediárias (2 a 5), coluna da direita notas dos extremos (1 e 6). Linha superior, Grupo A; linha intermediária Grupo AV; linha inferior, Grupo AVi. Em destaque nos *topoplots*, estão os eletrodos utilizados para o correlograma à direita. Correlograma: linha superior, fase pré; intermediária, fase pós; e inferior, diferença pós-pré. Em destaque no correlograma, os dados de tempo e frequência utilizados para *topoplots*. **B. Conectividade referente a T8.**

DISCUSSÃO

5. Discussão

Esta tese tem como objetivos principais responder e acrescentar elementos às seguintes perguntas:

1. É possível melhorar a aprendizagem auditiva através do pareamento de um estímulo visual não consciente?
2. Qual a influência do estímulo visual não consciente no processamento auditivo?
3. Quais os efeitos da estimulação audiovisual sobre a plasticidade promovida pela aprendizagem nesta tarefa?

Para isso, desenvolvemos um novo protocolo de aprendizagem auditiva, com estimulação audiovisual, com estímulo visual abaixo do limiar de detecção. A aprendizagem ocorria acerca da identificação de notas musicais através de uma escolha forçada dentre seis alternativas.

A principal hipótese levantada pelo trabalho é a de que a estimulação audiovisual, feita de maneira congruente entre os estímulos auditivos e visuais, favoreceria a aprendizagem na modalidade auditiva.

De fato, nossos resultados mostraram que o protocolo favoreceu a aprendizagem auditiva, levando a um aumento na performance do grupo experimental (AV), e este efeito foi maior que os grupos controles, com estimulação somente auditiva (A) e audiovisual incongruente (AVi). Ainda, uma análise mais detalhada destes resultados mostrou que a maior contribuição para a melhora da performance foi derivada da performance nas notas da extremidade da escala definida para o experimento.

Em termos neurais, nossos resultados mostram que houve uma atenuação dos ERPs ao compararmos os registros pré- e pós-treino para todos os grupos, com destaque para os grupos controles. A comparação entre os grupos das diferenças entre os ERPs pré-

e pós treino revelou diferenças significativas entre grupos em eletrodos parietais. Ao separarmos as análises entre as notas de extremos e intermediárias, as diferenças se sobressaem nas notas de extremo, e surgem diferenças em regiões frontais e na fase pós treinamento em eletrodos temporais no hemisfério esquerdo.

A plasticidade neural é refletida também nas oscilações theta, de 3Hz a 8Hz. Semelhante ao estudo de ERP, a potência desta faixa de frequência diminui entre as fases pré e pós, sobretudo nos grupos controles. O grupo experimental ou mantém ou aumenta a potência no pós-treinamento. Estas diferenças são encontradas em eletrodos centrais e temporais, em janelas temporais que evidenciam diferenças tanto no aspecto perceptual, quanto cognitivo, na memória e tomada de decisão dos indivíduos.

5.1 Os efeitos da estimulação multissensorial na aprendizagem auditiva

Para responder à questão:

É possível melhorar a aprendizagem auditiva através do pareamento de um estímulo visual não consciente? ,

aplicamos o protocolo referido em 33 voluntários sem histórico de estudo musical formal, como auferido pela índice Gold-MSI (Mullensiefen et al., 2014). Os resultados indicam que o pareamento do estímulo visual, mesmo apresentado abaixo do limiar de detecção favorece a aprendizagem auditiva. Os três grupos tiveram uma melhora significativa na performance pós- quando comparada com a pré-treinamento, sinalizando que o protocolo por si só gera aprendizagem perceptual. Embora ao se comparar os três grupos não tenha sido revelado efeito significativo, a comparação direta do efeito do treinamento entre pares dos grupos mostrou que este efeito foi maior no grupo experimental, com estímulo congruente. Este aumento foi de 53,8% no grupo AV, significativamente maior que grupo A (24,3%), e com

efeito marginal ($p = 0,06$) quando comparado com o grupo AVi (24,9%) (Figura 11). De um outro ponto de vista, verificamos o tamanho do efeito do treinamento de acordo com o *Cohen's d* (Cohen, 1988; Sawilowsky, 2009). Na interpretação de *Cohen's d*, o valor é considerado pequeno quando 0,2, médio quando 0,5, e muito grande quando acima de 1,2. O tamanho do efeito do treinamento para o grupo experimental foi de 1,39. Para o grupo A, o valor foi de 0,41, e para o grupo AVi, 0,57 (Tabela 5). Desta forma, ao levarmos em conta a definição da interpretação desta medida de tamanho do efeito, revela-se uma diferença substancial entre estes grupos. Estas duas abordagens estatísticas nos dão uma clara indicação de efeito diferente entre os grupos, favorecendo o grupo experimental, treinado com estímulo audiovisual congruente.

O estudo sobre a influência de um estímulo visual sobre o auditivo data de décadas (Pick et al., 1969; Warren et al., 1981). Contudo, somente nos anos 2000, surgiram evidências da influência de uma modalidade sobre a outra tanto na aprendizagem perceptual visual (e.g. Seitz et al., 2006, Sham & Seitz, 2008, Deveau et al 2014) quanto na auditiva (e.g. Lappe et al., 2008; Paraskevopoulos et al., 2012). Em geral, estes protocolos de treinamento se estendem por vários dias, mesmo que, por vezes, desde as primeiras sessões já evidenciam um efeito diferente entre os grupos treinados com protocolos uni- e multissensoriais (Seitz et al., 2006). De maneira semelhante, a maioria dos trabalhos de aprendizagem auditiva, e/ou musical, ocorre em protocolos de dias de treinamento (e.g. Cuddy, 1970; Menning et al., 2000; Foxton & Brown, 2004; Russo et al., 2005; Seppanen et al., 2012; Paraskevopoulos et al., 2012), embora também haja trabalhos evidenciando efeitos comportamentais e plasticidade neural após uma única sessão de treinamento com estimulação auditiva e visual (Souza et al., 2013) ou audiovisual (Zilber et al., 2014). Portanto, é digno de nota que nossos resultados são frutos de um protocolo de minutos de aprendizagem, e indicam, sobretudo, uma avenida a se explorar.

De maneira curiosa, outro aspecto importante da tarefa, o tempo de reação, foi diferente para os grupos, de modo que após o treinamento, o grupo experimental manteve-se com TRs semelhantes à fase pré-treinamento, diferentemente dos demais, que ficaram

significativamente mais rápidos. Como todos os sujeitos foram colocados sob uma mesma condição e instruídos com material escrito para serem o mais rápido e o mais correto possível, algumas hipóteses surgem para explicar este comportamento. A primeira delas relaciona-se com o nível de confiança nas respostas, que tende a favorecer um aumento na velocidade das respostas em tarefas *speeded*. Entretanto, quando comparados, não foram encontradas diferenças entre os grupos no pré-treino, no pós-treino, tampouco no efeito de treinamento (ET). Contudo, existe uma tendência de menores níveis de confiança no grupo experimental, e isto poderia explicar os TRs maiores, mesmo não havendo efeito de grupo. Há relatos de melhora na performance significativamente maior em grupo experimental sem aumento no nível de confiança quando comparado ao controle (Zilber et al., 2014). No entanto, em nosso trabalho, a correlação entre os níveis de confiança e os TRs foi desprezível (Hinkle et al., 2003). Outra hipótese levantada é a de que o grupo experimental poderia ter se beneficiado do tempo para resposta. Se este fosse o caso, encontraríamos uma correlação entre o ET e os TRs, de modo que quanto maior o tempo para resposta, maior o efeito do treinamento. No entanto, ao agrupar todos os indivíduos, a correlação também se mostrou desprezível e não significativa. As reais causas deste aumento ainda restam por serem investigadas.

Ao aprofundarmos a análise da performance nesta tarefa, revela-se que a principal fonte da diferença entre os grupos se encontra nas notas de extremo (Figura 12). O efeito do treinamento nas notas centrais mostra que pouco houve de melhora nos três grupos. Já as notas de extremo, embora representem somente 33,3% do total, foram as grandes responsáveis pela melhora observada ao se comparar as fases pré- e pós-treinamento. A música ocidental é baseada em escalas, nas quais um conjunto de notas desenha uma estrutura de intervalos entre si, e em geral, atribui-se um status especial à nota tônica (Herzog, 1949). Ainda na música ocidental, a probabilidade de se encontrar a nota tônica em uma escala é maior que as demais notas, e indivíduos são sensíveis a estas probabilidades (Krumhansl et al., 2004) fazendo com que sejam geradas expectativas, essenciais para o trabalho de tensão e alívio nas melodias (Bigand et al., 1996). Mesmo que

a escala desenvolvida para este trabalho não tivesse nenhuma diferença de probabilidade, ou seja, todas as notas eram tocadas o mesmo número de vezes e de maneira aleatória, estas duas notas de extremo adquiriram status diferente das demais. Assumindo que havia uma constante avaliação por parte dos sujeitos sobre a nota ouvida com relação à anterior, avaliação esta inevitável pela proximidade temporal da apresentação dos estímulos, podemos concluir que as expectativas criadas ao longo do experimento nunca eram satisfeitas na condição de uma nota mais grave que a nota 1 (extremo baixo), ou mais aguda que a nota 6 (extremo alto). Dessa forma, ambas as notas de extremo compartilhavam o fato de que a expectativa de encontrar uma nota além delas nunca era satisfeita, e portanto ganhavam um status especial, tal qual a tônica (por vezes as notas de extremo da escala, vide **dó**, ré, mi, fá, sol, lá, si, **dó**) em uma escala ocidental usual. Vale ressaltar que temos ciência que a comparação com a escala de dó a dó não representa fielmente o que se passou no experimento, uma vez que as oitavas guardam características muito diferentes das notas 1 e 6 (equivalentes a um intervalo entre 2ª e terça menor). De qualquer forma, a escala parece iniciar sua consolidação, para todos os grupos, a partir das notas de extremo.

5.2 A integração multissensorial nesta tarefa

Estudos anteriores mostraram que o processamento de estímulos visuais não percebidos conscientemente pode ser influenciado pela presença de um estímulo auditivo (Faivre et al., 2014; Alsius & Munhall, 2013; Chen & Spence, 2010, 2011; Ngo & Spence, 2010; Palmer & Ramsey, 2012). Nesta tese, utilizamos um estímulo visual potencialmente influenciando o processamento auditivo.

Para uma melhor compreensão dos substratos neurais envolvidos na tarefa e a plasticidade promovida, conduzimos o experimento com registro de EEG. Pelo fato do estudo envolver integração multissensorial, o suporte neural para estas influências estendem-se da participação de áreas de associação multissensoriais, temporais, como o

STS (Beauchamp et al., 2004), regiões do córtex pré-frontal (Calvert, 2001), parietais, como SIP (para cérebro multissensorial e música, ver Zimmerman & Lahav, 2012), até regiões “unissensoriais”, através da comunicação de córtices no início do processamento auditivo e visual. Neste caso, restava averiguar se a apresentação de estímulo visual abaixo do limiar de detecção poderia promover esta integração e em que nível de processamento. De maneira mais assertiva, responder à questão 2:

Qual a influência do estímulo visual não consciente no processamento auditivo?

Comportamentalmente, observamos diferenças significativas na performance entre os grupos durante o treino, quando há apresentação de estímulos visuais pareados a auditivos. Contudo, nestes grupos, apresentamos em 1/7 das vezes estímulos unissensoriais, ou seja, somente estímulo auditivo, e não observamos diferenças significativas entre a performance para estímulos unissensoriais e audiovisuais dentro de um mesmo grupo. Isso ocorreu tanto para o grupo AV quanto para AVi.

Em termos neurais, trabalhos anteriores mostraram efeitos tanto *superaditivos* quanto *subaditivos* ao comparar a resposta do potencial relacionado à apresentação do estímulo audiovisual (AV, par), com a soma daquela relacionada ao estímulo auditivo e ao estímulo visual (A+V, soma) (Giard & Perronet, 1999; Fort et al., 2002; Molholm et al., 2002; Vidal et al., 2008; Cappe et al., 2010). Como o estímulo visual é abaixo do limiar de detecção, é esperada uma resposta cortical sutil (Dehaene et al., 2006) quando captada por potenciais evocados de escalpo. De acordo com a teoria postulada por Dehaene e seus colaboradores (2006), baseada no espaço de trabalho global (Baars, 1988), esta resposta sutil, se restringe a regiões do início do processamento visual, e de maneira previsível, com um impacto também bastante tênue nas regiões de projeção tanto dentro do sistema visual, como de associação, e eventuais comunicações com outras regiões “unissensoriais”.

Esta resposta cortical relaciona-se diretamente com o desafio perceptual para o sistema visual que é a apresentação de um estímulo abaixo ou em torno do limiar de

detecção. O estímulo visual utilizado neste trabalho segue exatamente as características deste desafio. Por outro lado, o estímulo auditivo, embora desenhado para não ser facilmente identificado, foi selecionado dentro de uma faixa de frequência, volume e duração de modo a permitir a percepção, mesmo que apresentado durante um curto tempo (33ms) (Plack et al., 2005). Com esta combinação de fatores, de um estímulo degradado (visual) e outro apresentado claramente (auditivo), é oportuno lançar mão do princípio da Efetividade Inversa (*inverse effectiveness*) (Stanford et al., 2005). De acordo com este princípio, quando a saliência de um estímulo é pequena, a apresentação de estímulos de outra modalidade sensorial evoca uma resposta aumentada, tanto de neurônios individuais, quanto de regiões corticais. No caso de estímulos com grande saliência, a apresentação de estímulos pareados, resulta em uma simples adição, ou até na supressão da resposta neural, ou seja, da subadição. De um ponto de vista neurofisiológico, existem evidências de subadição tanto de registros de neurônios (Perrault et al., 2005) quanto de LFPs (Kayser et al., 2008).

Werner et al., (2009) mostraram que quando estímulos visuais e auditivos são apresentados de maneira degradada, ou seja, gerando uma dificuldade na percepção, ocorrem interações superaditiva no STS, enquanto que quando os estímulos eram claramente apresentados, a interação multissensorial se dava de maneira subaditiva.

Em nosso caso, mesmo com o estímulo visual degradado, a atenção era destinada ao auditivo, sendo este o objetivo central da tarefa. Portanto, poderíamos esperar uma resposta subaditiva. Na região parietal, encontramos diferenças marginais indicando subadição. Esse resultado está em acordo com o demonstrado por Cappe et al., (2010), em que uma das regiões do escalpo afetada foi grupos de eletrodos parieto-occipitais. Nesse trabalho, eles investigaram as potenciais fontes geradoras deste sinal, e encontraram, coerentemente com outros trabalhos, regiões auditivas, visuais e pSTS. Não estendemos nossa análises de maneira semelhante, mas a ocorrência em grupos de eletrodos em regiões semelhantes no escalpo, nos permite sugerir interações de regiões auditivas, visuais e de integração, tal qual STS e IPS.

Além destas regiões, nossos dados revelam influência marginal também em regiões centro-frontais. De fato, trabalhos recentes mostram que o pareamento escolhido para a tarefa, de elevação e frequência da nota, também pode refletir influências top-down, mais especificamente relacionada à atenção multissensorial (McCormick et al., 2018), e o envolvimento de regiões centrais e frontais estaria em linha com estes achados recentes.

Os resultados deste trabalho se aproximam do encontrado na literatura e sugerem uma informação importante: ser possível encontrarmos efeitos da integração audiovisual do ponto de vista neural mesmo com a apresentação de estímulo abaixo do limiar de detecção. Naturalmente, dadas as características, além de degradadas, de não consciente do estímulo visual, é razoável esperar que o efeito subaditivo seja marginal. De qualquer forma, nossos resultados mostram que a subadição parece ocorrer em regiões associativas. Esta questão poderá ser endereçada estendendo a análise dos dados apresentados neste trabalho, para a estimação de fonte elétrica. Outra abordagem para futuros trabalhos seria um desenho experimental específico para testar esta interação, aumentando o número de tentativas somente auditivas, e distribuindo as visuais intercaladas com as auditivas e audiovisuais.

Embora não tenhamos encontrado efeito comportamental da estimulação multissensorial ao compararmos as tentativas unissensoriais com aquelas multissensoriais dentro do mesmo grupo, este efeito foi observado comparando-se os grupos, efeito este sobretudo devido ao bloco final de treinamento. Mesmo que não tenha havido efeito comportamental observável intragrupo durante a apresentação, este efeito impactou no resultado do treinamento, promovendo inclusive a plasticidade observada na fase pós-treinamento

5.3. A plasticidade nesta tarefa

Quais os efeitos da estimulação audiovisual sobre a plasticidade promovida pela aprendizagem nesta tarefa?

Os resultados eletrofisiológicos mostraram que houve uma atenuação de marcadores eletrofisiológicos nos três grupos, quando comparamos a fase pré- e pós-treino. Contudo um destaque deve ser dado ao grupo experimental AV, que teve uma menor atenuação ou não mostrou diferença significativa entre estas fases na análise de tempo-frequência.

Por um lado, temos elementos relacionados à percepção do estímulo, como a atenuação da resposta dos componentes do ERP N1-P2, e a banda theta entre 0 e 200ms. A plasticidade do complexo N1-P2 relacionada à experiência musical tem sido documentada frequentemente, seja ao se considerar a experiência musical de longo prazo (Baumann et al., 2008; Kuriki et al., 2006; Shahin et al., 2003, 2005; Jongsma et al., 2005), seja no curto prazo (Seppanen et al., 2012). Após numerosas evidências de alterações no P2 seguindo aprendizagem, como as supracitadas, Tremblay et al., (2014) levantam a hipótese do componente P2 ser um biomarcador de aprendizagem. Em seu trabalho, acompanham a evolução deste componente em quatro diferentes momentos ao longo de diversas sessões de treinamento, e posteriormente em um período de retenção. Seus resultados indicam que o aumento deste componente não parece estar relacionado diretamente com a aprendizagem, e os autores interpretam as alterações como uma atividade neural associada ao processo de aquisição e não com o resultado da aprendizagem em si (Tremblay et al., 2014). No entanto o que observamos em nossos resultados foi uma diminuição deste componente, em linha com achados anteriores (Atienza et al., 2002), refletindo mais um efeito de habituação neural que de aprendizagem.

No que tange as oscilações, encontramos diferenças relacionadas ao processo perceptual, em regiões temporais no hemisfério cerebral esquerdo, entre 100-250ms, para frequências 4-8Hz (Figura 22). Diferentemente dos controles, o grupo AV não apresentou queda de potência da fase pré- para pós-treinamento, e as diferenças estatísticas entre grupos são significativas nas notas intermediárias, ainda que não tenhamos encontrado impacto na performance. Este fato é interessante, uma vez que esta região cortical e esta janela temporal refletem sobretudo o processamento *bottom-up* do estímulo. Trabalhos com tarefas auditivas, como Tesche e Karhu (2000) sugerem que oscilações theta observadas relacionadas à apresentação do estímulo para uma tarefa de memória de trabalho, estariam relacionadas a redes neurais envolvendo o hipocampo no processamento de *inputs* sensoriais. Contudo, a tarefa deste trabalho envolve a co-ativação de inúmeras áreas cerebrais, e tem sido mostrado que a frequência theta tem um papel integrador em demandas cognitivas nas quais vários recursos são requisitados (Dai et al., 2017; Sarnthein et al., 1998; Sauseng et al., 2010).

Por outro lado, janelas temporais características de elementos ligados à atenção, ao processamento da informação e elaboração da resposta também foram observadas exibindo diferenças entre os grupos. Nossos resultados sugerem que as frequências em theta (3-8Hz) tendem a diminuir a sincronização no pós- quando comparado ao pré-treinamento nos grupos controles, mas não no grupo experimental, especialmente em 3-5Hz, na janela de 300 - 600ms. Diferentemente das regiões temporais, estas diferenças foram observadas no eletrodo Cz, sobretudo em notas de extremo (Figura 22), nas quais a performance comportamental foi melhor. Estes resultados foram condizentes com as observações feitas pelos participantes que relatavam identificar estas notas de maneira mais eficaz, uma vez que os componentes eletroencefalográficos nesta janela temporal, tal qual P300 relacionam-se diretamente com processo endógenos, influenciados por exemplo com a atenção do indivíduo (para revisão, ver Joos, 2014). Alguns trabalhos mostram que o componente P300 está relacionado a uma maior capacidade de memória de trabalho (e.g., Nittono et al., 1999; Gevins & Smith, 2000), ativamente envolvida nesta tarefa. Dong et al.

(2015) mostram que indivíduos classificados com alta performance em tarefas de memória de trabalho, exibem um maior P300 para uma tarefa com diferentes níveis de dificuldade (1 a 4 *n-back*), uma maior sincronização em theta (3-7Hz), e menor desincronização de alfa (8-12Hz) (Dong et al., 2015). Gevins e Smith (2000) encontraram uma correlação positiva entre o desempenho da memória de trabalho e sincronização em theta em regiões frontais (3 - 7Hz). E Lee et al. (2005) sugeriram em seu trabalho que theta pode estar relacionada com a manutenção e regulação de informações relevantes durante a memória de trabalho.

Esta informação, somada ao conjunto de evidências em favor de uma maior sincronização theta e o desempenho durante tarefas de memória de trabalho, sugerem que o fato do grupo experimental não exibir uma diminuição nesta sincronização pode estar relacionado ao melhor efeito de treinamento. Outro elemento importante que corrobora esta relação é o fato de que, assim como pudemos observar uma melhora no desempenho concentrada nas notas de extremo, isto se reflete na análise de tempo-frequência. Ao compararmos as diferenças pré- vs pós-treinamento, notamos que o grupo A mostra uma diminuição significativa, AVi uma menor diminuição e AV não mostra diminuição, ao contrário, mostra um aumento da potência de theta em regiões centrais, parietais e temporais. Esta distribuição ao longo do escalpo deixa evidente o envolvimento de uma rede distribuída pelo córtex, e que o efeito do treinamento sobre esta rede é diferente de acordo com a estimulação durante o treinamento. Esta diferença de processamento pode estar relacionada à não diminuição do tempo de reação da fase pré- para pós-treinamento. E é natural encarmos uma rede distribuída uma vez que esta tarefa demandou diversas capacidades cognitivas, dentre elas, a memória de trabalho.

A música e a memória de trabalho têm uma relação bastante íntima. Diversos estudos mostram que musicistas performam melhor que não musicistas em tarefas de memória de trabalho (e.g., Franklin et al., 2008; Parbery-Clark et al., 2009; George & Coch, 2011). Entretanto, estudos sobre os processos neurais envolvidos na memória de trabalho auditiva têm sido até recentemente bastante controversos. Trabalhos utilizando Imagens por Ressonância Magnética funcional e Tomografia por Emissão de Pósitron mostraram tanto

supressão (Linke et al., 2011, em tarefa de manutenção de duas notas na memória; e Zatorre et al., 1994, ao comparar duas notas com interferências de outras notas entre elas), quanto aumento da atividade no córtex auditivo (Linke & Cusack, 2015, ao se manter um único som complexo; e Gaab et al., 2003 ao comparar duas notas com interferências de outras notas entre elas). Recentemente, Kumar et al. (2016) se propuseram a separar o processo de codificação, retenção e recuperação da informação de uma dentre duas notas apresentadas. Os autores sugerem que a manutenção de uma nota na memória, conta com o envolvimento do córtex auditivo, e que há uma conectividade funcional entre córtex auditivo, hipocampo e córtex pré-frontal (giro frontal inferior) durante o processo de manutenção de representação dos sons no córtex auditivo. Outros trabalhos recentes corroboram esta conectividade e a relação com oscilações theta (Gruber et al., 2018; Kaplan et al., 2017; Herweg et al., 2016; Fuentemilla et al., 2014). Ao considerarmos que a tarefa demandava constantemente a manutenção da informação tanto da escala quanto da nota anterior, a conservação da sincronização theta ocorrida no grupo experimental ao nas fases pré- e pós-treinamento, sugere que as oscilações theta favoreceram este processo. Isso pode ser notado tanto durante o início do processamento sensorial, quanto em processamentos mais tardios, sugerindo que esta faixa de frequência reflete um melhor efeito do treinamento quando comparado aos grupos controles.

Estudos recentes têm mostrado que o envolvimento dos córtices unissensoriais, bem como a comunicação entre eles, tem papel importante na integração multissensorial (Falchier et al., 2002, 2009; Rockland & Ojima, 2003; Cappe & Barone, 2005; Clarke & Innocenti, 1990; Clemo et al., 2008; Vaudano et al., 1991; Laramée et al., 2011, 2013). Neste trabalho, encontramos evidências desta relação através da conectividade funcional avaliada pela sincronia de fase. Encontramos efeitos robustos de aumento na conectividade após o treinamento entre eletrodos de regiões auditivas e visuais, mesmo na fase do experimento sem nenhum estímulo visual (pós-treinamento) na janela temporal característica de processamento *bottom-up*. Os resultados mostram ainda que este aumento deve-se sobretudo à diminuição da latência da conectividade entre estas regiões.

Dessa forma, demonstramos que o treinamento com a estimulação audiovisual, mesmo que nãoconsciente, aumenta a conectividade entre as regiões sensoriais, em especial no grupo com estimulação congruente, no qual foram observados os resultados de performance mais expressivos. Este aumento de conectividade é notável, uma vez que áreas visuais passam a ser recrutadas na tarefa, mesmo quando não há estímulo visual presente.

Se por um lado, durante tarefas de memória, a potência de theta aumenta, por outro lado a diminuição da potência de alfa está relacionada a altas demandas de atenção (Klimesch 1999; Ward, 2003), e frequentemente são observadas em uma mesma tarefa (Cohen & Donner 2013; Bastos et al., 2015; Michalareas et al., 2016). Em nosso trabalho pudemos observar uma dessincronização da frequência beta após a apresentação do estímulo entre todos os grupos, corroborando dados da literatura.

Esta demanda cognitiva vem sendo aos poucos melhor descrita, levando a uma nova linha de pensamento de que as oscilações alfa estão associadas a inibição de ativações de regiões cerebrais irrelevantes para a tarefa (Klimesch et al., 2007; Mazaheri & Jensen, 2010; Scheeringa et al., 2012; Uusberg et al., 2013), incluindo tarefas de memória de trabalho (Roux et al., 2012; León-Domínguez et al., 2015). Embora diversos trabalhos mostrem uma clara diferença funcional entre ritmos alfa e beta (Wang, 2010; Haegens et al., 2014; Bressler & Richter, 2015; Gregoriou et al., 2015), outros revelam o envolvimento de ambas as bandas, com um pico de atividade em uma banda que Michalareas e colaboradores (2016) denominaram de alfa-beta (Buffalo et al., 2011; Michalareas et al., 2016). Em linha com outros trabalhos (e.g. Bastos, 2015), os autores mostram que bandas de frequência alfa e beta, exercem um papel de controle top-down, com projeções de feedback para os córtices sensoriais, no caso, visuais.

De fato, o envolvimento da banda beta mostrou-se alinhado com a apresentação do estímulo após em torno de 300ms até a execução da resposta. Em um trabalho desenhado para compreender a correspondência entre pitch e elevação, Cormick et al., (2018) mostraram que o par frequência-elevação parece estar relacionado mais com a atenção multisensorial do que com elementos sensoriais, ou seja, em favor de envolvimento *top-*

down sobre *bottom-up*. Contudo, não observamos diferenças significativas entre as fases do pré- e pós-treinamento em nenhum dos grupos.

Observamos também efeitos mais tardios no processamento entre 600ms e 850ms em eletrodos frontais, com maior amplitude no grupo AV na fase pós-treinamento, e também um maior efeito pós - pré (Figura 19 A). Dentro desta rede distribuída é esperado o envolvimento de regiões como o córtex pré-frontal dorso-lateral, córtex cingulado, córtex pré-motor e área motora suplementar (Cole & Schneider, 2007). Alterações nestas regiões poderiam estar relacionadas aos efeitos observados nos ERPs. Na literatura, efeito semelhante foi encontrado em Tremblay et al., (2014), contudo sob a forma de uma negatividade.

Uma grande parte dos trabalhos relacionando sinais de EEG com a performance em tarefas de memória de trabalho, indicam uma correlação com a dificuldade da tarefa e/ou diferenças inter individuais, tal como capacidade da memória de trabalho (Nittono et al., 1999; Dong et al., 2015; Dai et al., 2017). Dessa forma, a heterogeneidade dos grupos, com indivíduos com alta e outros com mais baixa capacidade da memória de trabalho, não controladas, pode amenizar os efeitos. Uma possível análise futura é justamente encontrar dentro de cada grupo os “good learners” e “bad learners” (Luft, 2014), e buscar pelas correlações nesta tarefa de aprendizagem auditiva com estimulação multissensorial.

De qualquer forma, mesmo assumindo esta heterogeneidade, este trabalho abre caminhos para uma maior compreensão e aplicação de métodos multissensoriais sobre a aprendizagem auditiva.

LIMITAÇÕES E DIREÇÕES FUTURAS

6. Limitações do trabalho e direções futuras

A primeira limitação deste trabalho está relacionada ao desenho experimental, que, embora tenha revelado elementos suficientes para identificarmos efeitos da estimulação multissensorial e do treinamento, um protocolo mais longo, com dias de treinamento certamente potencializaria estes efeitos e trariam uma acentuação da plasticidade promovida na tarefa. Ainda dentro do protocolo, a maneira que os sujeitos efetuavam a resposta, favoreceu a criação de um mapa espacial horizontal das notas de grave a agudo. Isso pode ter contribuído negativamente para o pareamento dos estímulos, uma vez que a congruência se baseava na verticalidade do estímulo visual. Outra limitação se relaciona com o grupo AVi, cuja apresentação visual foi incongruente, mas constante, ou seja, as notas eram pareadas sempre com o mesmo estímulo visual. Isso pode permitir que, embora não seja esperado com um estímulo abaixo do limiar, ocorra de integração destas informações. Também reconhecemos a limitação imposta na interpretação dos níveis de confiança reportados da maneira que foi feito, já que o ideal é ter a resposta sobre a confiança imediatamente após cada resposta dada na identificação das notas. Contudo, esta abordagem mostrou-se impraticável pelo tempo de experimento aumentar significativamente.

Futuras direções envolvem encontrar regiões cerebrais geradoras dos resultados encontrados através de análise da estimação de fonte elétrica; expandir a análise para identificar *good learners* e *bad learners* e os processos cerebrais envolvidos nestas diferenças (Luft, 2014); estender a análise a partir de correlações entre o sinal do EEG e performance, como por exemplo, independente de grupo, a correlação entre potência de theta e o Desvio Absoluto na tentativa; modificar o protocolo de modo a se tornar viável para aplicação prática fora da ciência.

CONCLUSÃO

7. Conclusão

A primeira conclusão do trabalho é que a estimulação multissensorial contribuiu significativamente para a melhora da aprendizagem auditiva relacionada ao desenvolvimento do ouvido musical. A banda de frequência theta parece ter um papel fundamental, desde o início do processamento em córtices “unissensoriais” quanto em processos de alta ordem do processamento.

Tendo em vista o grande desafio para músicos no desenvolvimento tanto do ouvido absoluto quanto relativo, especialmente quando o início do treinamento se dá após a infância, este método abre uma nova avenida para protocolos mais eficientes de aprendizagem auditiva, tanto no caráter experimental e investigativo, quanto na exploração potencialmente de mercado.

E por fim, se em uma peça musical apreciada em absoluto silêncio por 4’33” reina a informação visual, embarcar no caráter multissensorial da aprendizagem, é mais que um desafio, é um passo determinante para ouvidos melhores. De olhos abertos.

REFERÊNCIAS BIBLIOGRÁFICAS

8. Referências Bibliográficas

- Alcalá-Quintana R, García-Pérez MA. A comparison of fixed-step-size and Bayesian staircases for sensory threshold estimation. *Spatial Vision* ; 20(3): 197-218, 2007.
- Alsius, A., & Munhall, K. G. Detection of audiovisual speech correspondences without visual awareness. *Psychological Science*, 24, 423–431 (2013).
- Atienza, M., Cantero, J.L. Dominguez-Martin, E. The Time Course of Neural Changes Underlying Auditory Perceptual Learning *Learn. Mem.* . 9: 138-150, 2002.
- Bartlett, J. C., & Dowling, W. J. . Recognition of transposed melodies: A key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 501–515, 1980.
- Baars, B.J.. A Cognitive Theory of Consciousness, Cambridge University Press, 1988.
- Baggaley, J. Measurement of absolute pitch. *Psychology of Music*, 2(2), 11-17, 1974.
- Bareither, I., Chaumon, M., Bernasconi, F., Villringer, A., Busch, N.A.. Invisible visual stimuli elicit increases in alpha-band power. *J Neurophysiol* 112:1082–1090, 2014.
- Baumann, S., Meyer, M., Jäncke, L. Enhancement of auditory-evoked potentials in musicians reflects an influence of expertise but not selective attention. *J. Cogn. Neurosci.* 20, 2238–2249, 2008.
- Beauchamp M. S., Lee K. E., Argall B. D., Martin A. Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 41, 809–823, 2004.
- Ben-Artzi, E., & Marks, L. E. Visual–auditory interaction in speeded classification: Role of stimulus difference. *Perception & Psychophysics*, 57, 1151–1162, 1995.

- Bermudez, P. & Zatorre, R. J. A distribution of absolute pitch ability as revealed by computerized testing. *Music Perception*. 27 (2): 89-101, 2008.
- Bertelson, P., Vroomen, J., Wiegeraad, G., & de Gelder, B. Exploring the relation between McGurk interference and ventriloquism. *Proceedings of the 1994 International Conference on Spoken Language Processing*, 2, 559–562, 1994.
- Bigand, E., Parncutt, R, Lerdahl, F. Perception of musical tension in short chord sequences: the influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Percept Psychophys.*; 58(1):124-41,1996.
- Bressler, S.L., Richter, C.G. Interareal oscillatory synchronization in top-down neocortical processing. *Curr. Opin. Neurobiol.* 31, 62–66, 2015.
- Calvert, G.A. Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb Cortex*; 11:1110-23, 2001.
- Cappe, C., Barone, P. Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. *European Journal of Neuroscience*, 22, 2886– 2902, 2005.
- Chen, Y.-C., & Spence, C. When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures. *Cognition*, 114, 389–404, 2010.
- Chen, Y.-C., & Spence, C. The crossmodal facilitation of visual object representations by sound: Evidence from the backward masking paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, 37,1784–1802, 2011.
- Chen, Y.-C., Huang, P-C., Woods, A., & Spence, C. When “Bouba” equals “Kiki”: Cultural commonalities and cultural differences in sound-shape correspondences. *Scientific Reports*, 6, 26681, 2016.

- Clarke, S., Innocenti, G.M. Auditory neurons with transitory axons to visual areas form short permanent projections. *European Journal of Neuroscience*, 2(3), 227-242, 1990.
- Clemon, H.R., Sharma, G.K., Allman, B.L., Meredith, M.A. Auditory projections to extrastriate visual cortex: connectional basis for multisensory processing in 'unimodal' visual neurons. *Experimental Brain Research*, 191(1), 37-47, 2008.
- Cohen, J. *Statistical power analysis for the behavioral sciences*. Hillsdale, N.J: L. Erlbaum Associates, 1988.
- Cohen, M., Donner, T. Midfrontal conflict-related theta-band power reflects neural oscillations that predict behavior *J Neurophysiol* 110: 2752–2763, 2013.
- Cohen, M.X. Comparison of different spatial transformations applied to EEG data: A case study of error processing. *International Journal of Psychophysiology*. Volume 97, Issue 3, Pages 245-257, 2015.
- Cohen, M.A. & Dennett, D.C. Consciousness cannot be separated by function. *Trends in Cognitive Sciences*. 15(8), 358-364, 2011.
- Cole, M. W., & Schneider, W. The cognitive control network: integrated cortical regions with dissociable functions. *Neuroimage* 37, 343–360, 2007.
- Connor, S. *Dumbstruck: A Cultural history of ventriloquism effect* (Oxford: Oxford University Press), 2000.
- Crick, F. & Koch C. A framework for consciousness. *Nature Neuroscience*. 6(2),119-126, 2003.
- Cuddy, L. L. Training the absolute identification of pitch. *Perception & Psychophysics*, 8, 265-269, 1970.
- Cuddy, L. L., & Cohen, A. J.. Recognition of transposed melodic sequences. *Quarterly Journal of Experimental Psychology*, 28, 255–270, 1976.

- Dai, Z., de Souza, J., Lim, J., Ho, P.M., Chen, Y., Li, J., Thakor, N., Bezerianos, A., Sun, Y. EEG Cortical Connectivity Analysis of Working Memory Reveals Topological Reorganization in Theta and Alpha Bands. *Front. Hum. Neurosci.* 11:237. doi: 10.3389/fnhum.2017.00237, 2017
- Dehaene, S., Changeaux, J-P., Naccache, L., Sackur, J., Sergent, C. Conscious, preconscious and subliminal processing: a testable taxonomy. *Trends in Cognitive Sciences.* 10(5), 204- 211, 2006.
- Dehaene, S. & Naccache, L. Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* 79, 1–37, 2001.
- Delorme A., Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* 134:9 – 21, 2004.
- Deveau, J., Ozer, D., Seitz, A. (2014). Improved vision and on-field performance in baseball through perceptual learning. *Current Biology.* Vol 24 (4), pR146–R147
- Deutsch, D. Tones and numbers: specificity of interference in immediate memory. *Science* 168: 1604-1605, 1970.
- Dong, S., Reder, L.M., Yao, Y., Liu, Y., Chen, F. Individual differences in working memory capacity are reflected in different ERP and EEG patterns to task difficulty. *Brain Res.* 2015;1616:146–56. pmid:25976774, 2015.
- Dowling, W. J. Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, 85, 341–354, 1978.
- Evans, K.K., Treisman, A. Natural cross-modal mappings between visual and auditory features. *Journal of vision.* 10(1):6, 1-12, 2010.

- Fahrenfort, J.J. & Lamme, V.A.F. A true science of consciousness explains phenomenology: comment on Cohen and Dennet. *Trends in Cognitive Sciences*. 16(3), 138-139, 2012.
- Faivre, N., Mudrik, L., Schwartz, N., Koch, C. Multisensory integration in complete unawareness: evidence from audiovisual congruency priming. *Psychol. Sci.* 25, 2006–2016, 2014.
- Falchier, A., Clavagnier, S., Barone, P., Kennedy, H. Anatomical evidence of multimodal integration in primate striate cortex. *Journal of Neuroscience*, 22, 5749–5759, 2002.
- Falchier, A., Schroeder, C. E., Hackett, T. A., Lakatos, P., Nascimento-Silva, S., Ulbert, I., Karmos, G, Smiley, J.F. Projection from visual areas V2 and prostriata to caudal auditory cortex in the monkey. *Cerebral Cortex*, 20, 1529–1538, 2009.
- Fort, A., Delpuech, C., Pernier, J., Giard, M.H. Dynamics of corticosubcortical cross-modal operations involved in audio-visual object detection in humans. *Cereb Cortex* 12:1031–1039, 2002.
- Foxe, J.J., Schroeder, C.E. The case for feedforward multisensory convergence during early cortical processing. *Neuroreport* 16, 419–423, 2005.
- Foxton, J. M., Brown, A. C. B. Training improves acoustic pattern perception. *Current Biology*, 14, 322-325, 2004.
- Franklin, M., Moore, K., Yip, C., Jonides, J. The effects of musical training on verbal memory. *Psychology of Music*, 36(3), 353–365, 2008.
- Frens, M. A., Van Opstal, A. J., & Van der Willigen, R. F. Spatial and temporal factors determine audio-visual interactions in human saccadic eye movements. *Perception & Psychophysics*, 57, 802–816, 1995.

- Fuentemilla, L., Barnes, G. R., Duzel, E., & Levine, B. Theta oscillations orchestrate medial temporal lobe and neocortex in remembering autobiographical memories. *Neuroimage*, 85, 730–737, 2014
- Fujioka, T., Trainor, L.J., Ross, B., Kakigi, R., Pantev, C. Musical training enhances automatic encoding of melodic contour and interval structure. *Journal of Cognitive Neuroscience* 16: 1010-1021, 2004.
- Gaab, N., Gaser, C., Zaehle, T., Jancke, L., Schlaug, G. Functional anatomy of pitch memory: an fMRI study with sparse temporal sampling. *Neuroimage* 19:1417–1426, 2003.
- García-Pérez, M.A. Forced-choice staircases with fixed-step sizes: Asymptotic and small sample properties, *Vision Research*; 38:1861– 1881, 1998.
- García-Pérez, M.A., Alcalá-Quintana, R. Shifts of the psychometric function: distinguishing bias from perceptual effects. *Q. J. Exp. Psychol.*; 66, 319–337, 2013.
- Gaser, C., Schlaug, G. Brain structures differ between musicians and non-musicians. *J. Neurosci.* 23:9240–45, 2003.
- Gevins, A., Smith, E. Neurophysiological Measures of Working Memory and Individual Differences in Cognitive Ability and Cognitive Style, *Cerebral Cortex*, 10(9), 829–839, 2000.
- George, E. M., Coch, D. Music training and working memory: An ERP study. *Neuropsychologia*, 49, 1083–1094. doi:10.1016/ j.neuropsychologia.2011.02.001, 2011.
- Ghazanfar, A.A., Schoeder, C.E. Is neocortex essentially multisensory? *Trends in Cognitive Sciences*. 10(6), 278-285, 2006.

- Giard, M.H., Peronnet, F. Auditory–visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J Cogn Neurosci* 11:473–490, 1999.
- Gold, J.I., Watanabe, T. Perceptual learning. *Curr Biol* 20:R46 –R48. 2010.
- Green, D.M., Swets, J.A. Signal detection theory and psychophysics: Wiley New York, 1996.
- Greenwald, A.G., Draine, S.C., Abrams, R.L. Three cognitive markers of unconscious semantic activation. *Science* 273, 1699 – 1702, 1996 .
- Grefkes, C., Fink, G.R. The functional organization of the intraparietal sulcus in humans and monkeys. *J Anat.* 207(1), 3-17. 2005
- Gregoriou, G.G., Paneri, S., Sapountzis, P. Oscillatory synchrony as a mechanism of attentional processing. *Brain Res.* 1626, 165–182, 2015.
- Gruber, M. J., Hsieh, L.-T., Staresina, B. P., Elger, C. E., Fell, J., Axmacher, N., & Ranganath, C. Theta Phase Synchronization between the Human Hippocampus and Prefrontal Cortex Increases during Encoding of Unexpected Information: A Case Study. *Journal of Cognitive Neuroscience*, 1–11. doi:10.1162/jocn_a_01302, 2018.
- Haegens, S., Cousijn, H., Wallis, G., Harrison, P.J., Nobre, A.C. Inter- and intra-individual variability in alpha peak frequency. *Neuroimage* 92, 46–55, 2014.
- Herweg, N. A., Apitz, T., Leicht, G., Mulert, C., Fuentemilla, L., & Bunzeck, N. Theta–alpha oscillations bind the hippocampus, prefrontal cortex, and striatum during recollection: Evidence from simultaneous EEG–fMRI. *Journal of Neuroscience*, 36, 3579–3587, 2016.

- Herzog G. Song: Folk song and the music of folk song. In: Leach M, editor. Funk & Wagnall's Standard Dictionary of Folklore, Mythology, and Legend. Vol. 2 Funk & Wagnall's Co; . pp. 1032–1050, 1949.
- Hinkle, D.E., Wiersma, W., Jurs, S.G. Applied Statistics for the Behavioral Sciences. 5th ed. Boston: Houghton Mifflin; 2003.
- Holender, D., Duscherer, K. Unconscious perception: the need for a paradigm shift. *Attention, Perception, & Psychophysics* 66: 872–881, 2004.
- Ibrahim, L.A., Mesik, L., Ji, X-Y., Fang, Q., Li, H-F., Li, Y.T., Zingg, B., Zhang, L., Tao, H.W. Cross-modality sharpening of visual cortical processing through layer-1-mediated inhibition and disinhibition. *Neuron* 89:1031–45, 2016
- Innes-Brown, H., Crewther, D. The impact of spatial incongruence on an auditory–visual illusion. *PLoS ONE*, 4, e6450, 2009.
- Jongsma, M., Eichele, T., Quiñero Quiroga, R., Jenks, K., Desain, P., Honing, H., Van Rijn, C. Expectancy effects on omission evoked potentials in musicians and non-musicians. *Psychophysiology* 42, 191–201, 2005.
- Jones, J. A., & Jarick, M. Multisensory integration of speech signals: The relationship between space and time. *Experimental Brain Research*, 174, 588–594, 2006.
- Jones, J. A., Munhall, K. G. The effects of separating auditory and visual sources on audiovisual integration of speech. *Canadian Acoustics*, 25, 13–19, 1997.
- Joos, K., Gilles, A., Van de Heyning, P., De Ridder, D., Vanneste, S. From sensation to percept: the neural signature of auditory event-related potentials. *Neurosci Biobehav Rev* 42:148 –156, 2014.
- Kanai, R., Tsuchiya, N., Verstraten, F.A.J. The scope and limits of top-down attention in unconscious visual processing. *Current Biology* 16: 2332–2336, 2006.

- Kaplan, R., Bush, D., Bisby, J. A., Horner, A. J., Meyer, S. S., & Burgess, N. Medial prefrontal–medial temporal theta phase coupling in dynamic spatial imagery. *Journal of Cognitive Neuroscience*, 29, 507–519, 2017.
- Kayser, C., Logothetis, N.K. Do early sensory cortex integrate crossmodal information? *Brain Struct Funct* 212:121–132, 2007.
- Keenan, J.P., Thangaraj, V., Halpern, A.R., Schlaug G., Absolute pitch and planum temporale. *Neuroimage* 14 (6), 1402–1408, 2001.
- Kim, J., Kroos, C., Davis, C. Hearing a point-light talker: an auditory influence on a visual motion detection task. *Perception* 39 (3), 407–416, 2010.
- King A.J. Visual influences on auditory spatial learning. *Phil. Trans. R. Soc. B* 364, 331–339, 2009.
- Klimesch, W. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res. Brain*, 1999.
- Klimesch, W., Sauseng, P., Hanslmayr, S. EEG alpha oscillations: the inhibition-timing hypothesis. *Brain Res. Rev.* 53, 63–88, 2007.
- Köhler, W. *Gestalt Psychology*. New York: Liveright; 1929.
- Köhler, W., *Gestalt Psychology: An Introduction to New Concepts in Modern Psychology*. New York: Liveright; 1947.
- Kouider, S., Dehaene, S. Levels of processing during non-conscious perception: a critical review of visual masking. *Phil. Trans. R. Soc. B* 362, 857– 875, 2007.
- Krumhansl, C.L. The cognition of tonality - as we know it today. *Journal of New Music Research*; 33:253–268, 2004.

- Kuriki, A., Kanda, S., Hirata, Y. Effects of musical experience on different components of MEG responses elicited by sequential piano-tones and chords. *J. Neurosci.* 26, 4046–4053, 2006.
- Kumar, S., Joseph, S., Gander, P.E., Barascud, N., Halpern, A.R., Griffiths, T.D. A brain system for auditory working memory. *Journal of Neuroscience* 36:4492–4505, 2016.
- Lacey, S., Martinez, M., McCormick, K., & Sathian, K. Synesthesia strengthens sound-symbolic cross-modal correspondences. *The European Journal of Neuroscience*, 44(9), 2716–2721, 2016.
- Lachaux, J.P., Rodriguez, E., Martinerie, J., Varela, F.J. Measuring phase synchrony in brain signals. *Hum Brain Mapp*; 8:194–208, 1999.
- Lamme, V.A. Why visual attention and awareness are different. *Trends Cogn. Sci.* 7, 12–18, 2003.
- Lamme, V.A. Towards a true neural stance on consciousness. *Trends in Cognitive Sciences*. 10(11), 494-501, 2006.
- Lappe, C., Herholz, S. C., Trainor, L. J., & Pantev, C. Cortical plasticity induced by short-term unimodal and multimodal musical training. *Journal of Neuroscience*, 28, 9632–9639, 2008.
- Laramée, M.E., Kurotani, T., Rockland, K.S., Bronchti, G., Boire, D. Indirect pathway between the primary auditory and visual cortices through layer V pyramidal neurons in V2L in mouse and the effects of bilateral enucleation. *European Journal Neuroscience*, 34(1), 65- 78, 2011.
- Lee, H., Simpson, G.V., Logothetis, N.K., Rainer, G. Phase locking of single neuron activity to theta oscillations during working memory in monkey extrastriate visual cortex. *Neuron* 45, 147–156, 2005.

- León-Domínguez, U., Martín-Rodríguez, J. F., León-Carrión, J. Executive n-back tasks for the neuropsychological assessment of working memory. *Behav. Brain Res.* 292, 167–173, 2015.
- Lewis, R., Noppeney, U. Audiovisual synchrony improves motions discrimination via enhanced connectivity between early visual and auditory areas. *The Journal of Neuroscience.* 30(37)12329-12339, 2010.
- Liang, M., Mouraux, A., Hu, L., Iannetti, G.D. Primary sensory cortices contain distinguishable spatial patterns of activity for each sense. *Nature Communications* 4,1979, 2013.
- Lieberman, A. M., Harris, K. S., Kinney, J., Lane, H. The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *J. Exptl. Psychol.*, 61, 379-88, 1961.
- Lin, Z., He, S. Seeing the invisible: The scope and limits of unconscious processing in binocular rivalry. *Progress in neurobiology* 87: 195–211, 2009.
- Linke, A.C., Vicente-Grabovetsky, A., Cusack, R. Stimulus-specific suppression preserves information in auditory short-term memory. *Proc Natl Acad Sci U S A* 108:12961–12966, doi:10.1073/pnas.1102118108,pmid:21768383, 2011.
- Linke, A.C., Cusack, R. Flexible information coding in human auditory cortex during perception, imagery, and STM of complex sounds. *J Cogn Neurosci* 27:1322–1333, doi:10.1162/jocn_a_00780, pmid:25603030, 2015.
- Loui, P., Zamm, A., Schlaug, G. Enhanced functional networks in absolute pitch, *NeuroImage.* 63(2), 632-640, 2012.

- Lovett, M., Reder, L., Lebiere, C. An ACT-R perspective on working memory. In: A. Miyake, & P. Shah (Eds.), *Models of working memory: mechanisms of active maintenance and executive control*. 135 – 182. New York: Cambridge University Press, 1999.
- Luft, C.D.B. Learning from feedback: the neural mechanisms of feedback processing facilitating better performance. *Behav. Brain Res.* 261, 356–368, 2014.
- Maeda, F., Kanai, R., Shimojo, S. Changing pitch induced visual motion illusion. *Current Biology*, 14, R990–R991, 2004.
- Mazaheri, A., Jensen, O. Rhythmic pulsing: linking ongoing brain activity with evoked responses. *Front. Hum. Neurosci.* 4:177, 2010.
- McCormick, K., Lacey, S., Stilla, R., Nygaard, L.C., Sathian, K. Neural basis of the crossmodal correspondence between auditory pitch and visuospatial elevation. *Neuropsychologia*, 112:19-30, 2018.
- McGurk, H., MacDonald, J. Hearing lips and seeing voices. *Nature* 264, 746–748, 1976.
- Menning, H., Roberts, L.E., Pantev, C. Plastic changes in the auditory cortex induced by intensive frequency discrimination training. *NeuroReport* ;11:817–22, 2000.
- Melara, R. D., O'Brien, T. P. Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General*, 116, 323–336, 1987.
- Mesulam, M.M. From sensation to cognition. *Brain*, 121, 1013–1052, 1998.
- Michalareas, G., Vezoli, J., van Pelt, S., Schoffelen, J.-M., Kennedy, H., Fries, P. Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. *Neuron* 89, 384–397, 2016.
- Michelson, A. *Studies in optics*. U. of Chicago Press. 1927.
- Mitra, P. & Bokil, H. *Observed Brain Dynamics* (Oxford University Press, USA, 2007).

- Molholm, S., Ritter, W., Murray, M.M., Javitt, D.C., Schroeder, C.E., Foxe, J.J. Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res Cogn Brain Res* 14:115–128, 2002.
- Morein-Zamir, S., Soto-Faraco, S., Kingstone, A. Auditory capture of vision: examining temporal ventriloquism. *Brain Res. Cogn. Brain Res.* 17, 154–163, 2003.
- Morgan, M., Dillenburger, B., Raphael, S., Solomon, J.A. Observers can voluntarily shift their psychometric functions without losing sensitivity. *Attention, Perception & Psychophysics*; 74 (1): 185–193, 2012.
- Müllensiefen, D., Gingras, B., Musil, J., Stewart, L. The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population. *PLoS ONE* 9(2): e89642. <https://doi.org/10.1371/journal.pone.0089642> , 2014.
- Murray, M. M., Thelen, A., Thut, G., Romei, V., Martuzzi, R., & Matusz, P. J. The multisensory function of the human primary visual cortex. *Neuropsychologia*, 83, 161–169. doi:10.1016/j.neuropsychologia.2015.08, 2016.
- Nittono, H., Nageishi, Y., Nakajima, Y., & Ullsperger, P. Event-related potential correlates of individual differences in working memory capacity. *Psychophysiology*, 36, 745–754, 1999.
- Ngo, M. K., & Spence, C. Auditory, tactile, and multisensory cues facilitate search for dynamic visual stimuli. *Attention, Perception, & Psychophysics*, 72, 1654–1665, 2010.
- Opstal, F.V., Gevers, W., Osman, M., Verguts, T. Unconscious task application. *Consciousness and Cognition* 19: 999–1006, 2010.

- Ozturk, O., Krehm, M., Vouloumanos, A. Sound symbolism in infancy: evidence for sound-shape cross-modal correspondences in 4-month-olds. *J. Exp. Child Psychol.* 114, 173–186, 2013.
- Palmer, T. D., & Ramsey, A. K. The function of consciousness in multisensory integration. *Cognition*, 3, 353–364, 2012.
- Patching, G. R., Quinlan, P. T. Garner and congruence effects in the speeded classification of bimodal signals. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 755–775, 2002.
- Paraskevopoulos, E., Kuchenbuch, A., Herholz, S.C., Pantev, C. Evidence for training-induced plasticity in multisensory brain structures: an MEG stud. *PLoS One* 7, e36534, 2012.
- Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. Musician enhancement for speech-in-noise. *Ear & Hearing*, 30(6), 653–661, 2009.
- Parise, C. V. Crossmodal correspondences: Standing issues and experimental guidelines. *Multisensory Research*, 29, 7–28, 2016.
- Paus, T. Functional anatomy of arousal and attention systems in the human brain. *Prog. Brain Res.* 126, 65–77, 2000.
- Pechmann, T., Mohr, G. Interference in memory for tonal pitch: implications for a working-memory model. *Mem. cognit.* 20:314-420, 1992.
- Perrin, F., Pernier, J., Bertrand, O. Echallier, J.F.. Spherical splines for scalp potential and current density mapping. *Electroencephalogr. Clin. Neurophysiol.*, 72 , pp. 184-187, 1989.
- Pick, H.L., Warren, D.H., Hay, J.C. Sensory conflict extensively, over about 60, enough to blur most scenes in judgements of spatial direction. *Percept. Psychophys.* 6,203–205, 1969.

- Plack, C. J., Oxenham, A. J., & Fay, R. R. (Eds.). Pitch: Neural coding and perception (Vol. 24). New York: Springer. 2005
- Powers, A.R., Hevey, M.A., Wallace, M.T. Neural correlates of multisensory perceptual learning. *The Journal of Neuroscience* 32 (18), 6263–6274, 2012.
- Proulx, M., Brown, D., Pasqualotto, A., Meijer, P. Multisensory perceptual learning and sensory substitution. *Neurosci. Biobehav. Rev.* 41:16-25. doi: 10.1016/j.neubiorev.2012.11.017, 2012.
- Ramachandran, V. S., Hubbard, E. M. Synaesthesia - A window into perception, thought and language. *Journal of Consciousness Studies*, 8, 3–34, 2001.
- Ramachandran, V. S., Hubbard, E. M. Hearing colors, tasting shapes. *Scientific American*, 288, 43–49, 2003.
- Recanzone, G. H. Auditory influences on visual temporal rate perception. *Journal of Neurophysiology*, 89, 1078–1093, 2003.
- Rees G., Kreiman, G., Koch C. Neural correlates of consciousness in humans. *Nat. Rev. Neurosci.* 3, 261–270, 2002.
- Reybrouck, M., Vuust, P., Brattico, E. Brain Connectivity Networks and the Aesthetic Experience of Music. *Brain Sci.* 8(6) pii: E107, 2018.
- Rockland, K. S., Ojima, H. Multisensory convergence in calcarine visual areas in macaque monkey. *International Journal of Psychophysiology*, 50, 19–26, 2003.
- Ross, D., Olson, I.R., Gore, J.C. Absolute pitch does not depend on early musical training *Ann. N.Y. Acad. Sci.* 999: 522-526, 2003.
- Roux, F., Wibral, M., Mohr, H. M., Singer, W., Uhlhaas, P. J. Gamma-band activity in human prefrontal cortex codes for the number of relevant items maintained in working memory. *J. Neurosci.* 32, 12411–12420, 2012.

- Rusconi, E., Priftis, K., Rusconi, M. L., Umiltà, C. Arithmetic priming from neglected numbers. *Cogn. Psychol.* 23, 227–239, 2006.
- Russo, F.A., Windell, D.L. & Cuddy, L. L. Learning the “Special Note”: evidence for a critical period for absolute pitch acquisition. *Music Perception* 21(1), 119-127, 2003.
- Salgado-Montejo, A., Marmolejo-Ramos, F., Alvarado, J. A., Arboleda, J. C., Suarez, D. R., & Spence, C. Drawing sounds: representing tones and chords spatially. *Experimental Brain Research*, 234(12), 3509–3522, 2016.
- Sandberg, K., Bibby, B.M., Timmermans, B., Cleeremans, A., Overgaard, M. Measuring consciousness: task accuracy and awareness as sigmoid functions of stimulus duration. *Conscious. Cogn.* 2011; 20:1659–1675.
- Sarnthein, J., Petsche, H., Rappelsberger, P., Shaw, G. L., Von Stein, A. Synchronization between prefrontal and posterior association cortex during human working memory. *Proc. Natl. Acad. Sci. U.S.A.* 95, 7092–7096. doi: 10.1073/pnas.95.12.7092, 1998.
- Sauseng, P., Griesmayr, B., Freunberger, R., Klimesch, W. Control mechanisms in working memory: a possible function of EEG theta oscillations. *Neurosci. Biobehav. Rev.* 34, 1015–1022. doi: 10.1016/j.neubiorev.2009.12.006, 2010.
- Sawilowsky, S. New effect size rules of thumb. *Journal of Modern Applied Statistical Methods.* 8 (2): 467–474, 2009.
- Schlaug, G. Musicians and music making as a model for the study of brain plasticity. *Prog. Brain Res.* 217, 37–55, 2015.
- Schlaug, G., Altenmüller, E., Thaut, M. Music listening and music making in the treatment of neurological disorders and impairments. *Music Percept.*; 27(249–250), 2010.

- Scheeringa, R., Petersson, K. M., Kleinschmidt, A., Jensen, O., Bastiaansen, M. C. EEG alpha power modulation of fMRI resting-state connectivity. *Brain Connect.* 2, 254–264, 2012.
- Schutz, M., Lipscomb, S. Hearing gestures, seeing music: visual influences perceived tone duration. *Perception.* 36: 888-897, 2007.
- Schroeder, C.E., Molholm, S., Lakatos, P., Ritter, W., Foxe, J.J. Human–simian correspondence in the early cortical processing of multisensory cues. *Cogn. Process* 5, 140–151, 2004.
- Seppanen, M., Hamalainen, J., Pesonen, A.K., Tervaniemi, M., Music training enhances rapid neural plasticity of N1 and P2 source activation for unattended sounds. *Front Hum Neurosci* 6:43, 2012.
- Seitz, A.R., Kim, R., Shams, L. Sound facilitates visual learning. *Current Biology.* 16. 1422, 2006.
- Shahin, A., Roberts, L., Pantev, C., Trainor, L., Ross, B. Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. *Neuroreport* 16, 1781–1785, 2005.
- Shams, L., Seitz, A.R. Benefits of multisensory learning. *Trends in Cognitive Sciences* 12 (11), 411–417, 2008,
- Shore, D. I., Barnes, M. E., Spence, C. The temporal evolution of the crossmodal congruency effect. *Neuroscience Letters*, 392, 96–100, 2006.
- Slutsky, D. A., & Recanzone, G. H. Temporal and spatial dependency of the ventriloquism effect. *NeuroReport*, 12, 7–10, 2001.

- Souza, A. C. S., Yehia, H. C., Sato, M., Callan, D. Brain activity underlying auditory perceptual learning during short period training: simultaneous fMRI and EEG recording. *BMC Neuroscience* 14:8, 2013.
- Spence, C. Crossmodal correspondences: a tutorial review. *Attention, Perception and Psychophysics*. 73:1-25, 2011.
- Stanford, T. R., Quessy, S., Stein, B. E. Evaluating the operations underlying multisensory integration in the cat superior colliculus, *J. Neurosci.* 25, 6499–6508, 2005.
- Takeuchi, A. H., Hulse, S. H. Absolute pitch. *Psychol. Bull.* 113: 345–361, 1993.
- Tamietto, M. De Gelder, B. Neural basis of the non-conscious perception of emotional signals. *Nat. Rev. Neurosci.* 11, 697–709, 2011.
- Tesche, C.D., Karhu, J. Theta oscillations index human hippocampal activation during a working memory task. *Proceedings of the National Academy of Sciences*, 97 (2) 919-924; DOI:10.1073/pnas.97.2.919, 2000.
- Trehub, S. E., Trainor, L. J., Unyk, A. M. Music and speech perception in the first year of life. In H. W. Reese & L. P. Lipsitt (Eds.), *Advances in child development and behavior* (Vol. 24, pp. 1–35). New York: Academic Press, 1993.
- Tsay, C. J. Sight over sound in the judgment of music performance. *Proc. Natl. Acad. Sci. USA*. 110 (36), 14580-14585, 2013.
- Tyll, S., Budinger, E., Noesselt, T. Thalamic influences on multisensory processing. *Communicative & Integrative Biology* 4:4, 378-381, 2011.
- Uusberg, A., Uibo, H., Kreegipuu, K., Allik, J. EEG alpha and cortical inhibition in affective attention. *Int. J. Psychophysiol.* 89, 26–36, 2013.
- van Wassenhove, V., Grant, K. W., Poeppel, D. Temporal window of integration in auditory–visual speech perception. *Neuropsychologia*, 45, 598–607, 2007.

- Vaudano, E., Legg, C.R., Glickstein, M. Afferent and efferent connections of temporal association cortex in the rat: a horseradish peroxidase study. *European Journal of Neuroscience*, 3, 317–330, 1991.
- Vidal, J., Giard, M.H., Roux S, Barthélémy, C., Bruneau, N. Crossmodal processing of auditory-visual stimuli in a no-task paradigm: a topographic event-related potential study. *Clin Neurophysiol* 119:763–771, 2008.
- Vroomen, J., Keetels, M. The spatial constraint in intersensory pairing: No role in temporal ventriloquism. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 1063–1071, 2006.
- Vuoskoski, J. K., Gatti, E., Spence, C., Clarke, E. F. Do Visual Cues Intensify the Emotional Responses Evoked by Musical Performance? A Psychophysiological Investigation. *Psychomusicology: Music, Mind, and Brain*, 26(2), 179-188, 2016.
- Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., Johnson, S.P. Preverbal infants' sensitivity to synesthetic cross-modality correspondences. *Psychological Science*, 21, 21–25, 2010.
- Wang, X.-J. Neurophysiological and computational principles of cortical rhythms in cognition. *Physiol. Rev.* 90, 1195–1268, 2010.
- Ward, W. D., Burns, E. M. Absolute pitch. in *The Psychology of Music*, ed. Deutsch, D. (Academic, New York), pp. 431–451. 1982
- Ward, L. M. Synchronous neural oscillations and cognitive processes. *Trends Cogn. Sci.* 7, 553–559, 2003.
- Warren, D.H., Welch, R. B., McCarthy, T.J. The role of visual-auditory “compellingness” in the ventriloquism effect: implications for transitivity among the spatial senses. *Percept Psychophys.* 1981 Dec;30(6):557-64. 1981.

- Watanabe, T., Nañez, J.E., Sasaki Y. Perceptual learning without perception. *Nature*. 413, 844-848, 2001.
- Welch, R. B. Warren, D. H. Immediate perceptual response to intersensory discrepancy, *Psycholog. Bull.* 88, 638–667, 1980.
- Wengenroth, M., Blatow, M., Heinecke, A., Reinhardt, J., Stippich, C., Hofmann, E., Schneider, P. Increased volume and function of right auditory cortex as a marker for absolute pitch. *Cereb Cortex* 24:1127–1137, 2014.
- Widmann, A., Kujala, T., Tervaniemi, M., Kujala, A., Schröger, E. From symbols to sounds: Visual symbolic information activates sound representations. *Psychophysiology*, 41, 709–715, 2004.
- Winawer, J., Huk, A.C., Boroditsky, L. A motion aftereffect from visual imagery of motion. *Cognition* 114: 276–284, 2010.
- Zatorre, R.J., Evans, A.C., Meyer, E. Neural mechanisms underlying melodic perception and memory for pitch. *J Neurosci* 14:1908 –1919, 1994.
- Zatorre, R.J., Perry, D.W., Beckett, C.S., Westbury, C.F., Evans, A.C. Functional anatomy of musical processing in the listeners with absolute pitch and relative pitch. *Proc. Natl. Acad. Sci. USA* 95, 3172-3177, 1998.
- Zatorre, R.J. Absolute pitch: a model for understanding the influence of genes and development on neural and cognitive function. *Nature Neuroscience* 6 (7), 692-695, 2003.
- Zatorre, R.J., Chen, J.L., Penhune, V.B. When the brain plays music: auditory-motor interactions in music perception and production. *Nat Rev Neurosci.*; 8(7):547–558, 2007.
- Zeki S. The disunity of consciousness. *Trends Cogn. Sci.* 7,214–218, 2003.

Zilber, N., Ciuciu, P., Gramfort, A., Azizi, L., van Wassenhove, V. Supramodal processing optimizes visual perceptual learning and plasticity, *NeuroImage*, 93(1)32-46, 2014.

Zimmerman, E., Lahav, A. The multisensory brain and its ability to learn music. *Ann NY Acad Sci*;1252:179–84, 2012.