Running head:

VISUAL ENHANCEMENT OF THE MMNm AND AUDITORY SENSITIVITY

1

2

**Poorer auditory sensitivity is related to stronger visual enhancement of the human
auditory mismatch negativity (MMNm)**

3

4

5

Cecilie Møller^{*1}, Andreas Højlund^{2,3}, Klaus B. Bærentsen⁴, Niels Chr. Hansen^{1,5},

6

Joshua C. Skewes^{3,6}, Peter Vuust¹

7

8

¹Center for Music in the Brain, Aarhus University & The Royal Academy of Music

9

Aarhus/Aalborg, Aarhus, Denmark

10

²Center of Functionally Integrative Neuroscience, Aarhus University Hospital,

11

Aarhus, Denmark

12

³Interacting Minds Centre, Aarhus University, Aarhus, Denmark

13

⁴Department of Psychology and Behavioral Sciences, Aarhus University, Aarhus,

14

Denmark

15

⁵The MARCS Institute for Brain, Behaviour, and Development, Western Sydney

16

University, Australia

17

⁶Department of Linguistics, Cognitive Science, and Semiotics; School of

18

Communication and Culture; Aarhus University, Denmark

19

20

21

*Correspondence concerning this article should be addressed to:

22

Cecilie Møller, Center for Music in the Brain, Nørrebrogade 44, 10G-4-38, 8000 Aarhus C,

23

Denmark. Phone: +45 51967927. E-mail: cecilie@clin.au.dk

Abstract

Multisensory processing facilitates perception of our everyday environment and becomes particularly important when sensory information is degraded or close to the discrimination threshold. Here, we used magnetoencephalography and an audiovisual oddball paradigm to assess the complementary role of visual information in subtle pitch discrimination at the neural level of participants with varying levels of pitch discrimination abilities, i.e., musicians and nonmusicians. The amplitude of the auditory mismatch negativity (MMNm) served as an index of sensitivity. The gain in amplitude resulting from compatible audiovisual information was larger in participants whose MMNm amplitude was smaller in the condition deviating only in the auditory dimension, in accordance with the multisensory *principle of inverse effectiveness*. These findings show that discrimination of even a sensory-specific feature as pitch is facilitated by multisensory information at a pre-attentive level, and they highlight the importance of considering inter-individual differences in uni-sensory abilities when assessing multisensory processing.

Keywords: audiovisual processing, principle of inverse effectiveness, pitch, musical expertise, magnetoencephalography

42 **Poorer auditory sensitivity is related to stronger visual enhancement of the auditory**
 43 **mismatch negativity (MMNm)**

44 Because the world is multimodal, everyday perception relies on the complementary
 45 roles of simultaneously active sensory systems. Contemporary brain science focuses
 46 increasingly on the principles underlying cortical connectivity (Friston, 2005), and the notion
 47 that multisensory processing in the brain only occurs after extensive processing in strictly
 48 sensory-specific cortical areas (Felleman & Van Essen, 1991) has been challenged in the past
 49 decades (for a classic review, see Ghazanfar & Schroeder, 2006), where an increasing
 50 number of studies have shown crossmodal influences on processing in primary sensory areas
 51 such as the auditory cortex (Atilgan et al., 2018; Brosch, Selezneva, & Scheich, 2005; Jäncke,
 52 Rogenmoser, Meyer, & Elmer, 2012; van Wassenhove & Grzeczkowski, 2015).

53 The *mismatch negativity* (MMN) and its magnetic counterpart MMNm (also termed
 54 *mismatch field* or *mismatch response*), is used to probe pre-attentive levels of unisensory
 55 auditory processing (Näätänen, Paavilainen, Rinne, & Alho, 2007). It is a robust measure
 56 derived from the event-related potential (ERP) / event-related field (ERF), which can be
 57 elicited in the auditory cortex in response to deviants in a repetitive train of auditory tones,
 58 and it is regarded as an index of various auditory functions, including neural sound
 59 discrimination and sensitivity (Näätänen, 2008). The amplitude of the MMN is positively
 60 correlated with the magnitude of the deviant (Jaramillo, Paavilainen, & Näätänen, 2000;
 61 Sams, Paavilainen, Alho, & Näätänen, 1985) as well as with behavioral performance
 62 (Amenedo & Escera, 2000; Jaramillo et al., 2000; Novitski, Tervaniemi, Huotilainen, &
 63 Näätänen, 2004; Tervaniemi, Just, Koelsch, Widmann, & Schroger, 2005), which makes the
 64 MMN paradigm a useful method to assess perceptually relevant neural sensitivity to subtle
 65 changes in an auditory stimulus material.

66 Nonetheless, MMN/MMNm elicitation is not restricted to unisensory auditory
67 processing as strong audiovisual illusions may also elicit an MMN/MMNm in the auditory
68 cortex. A long tradition of scientific studies has used crossmodal perceptual illusions to
69 elucidate the importance and the consequences of the spontaneous cooperation between the
70 sensory modalities at behavioral (McGurk & MacDonald, 1976; Sekuler, 1997; Shams, Ma,
71 & Beierholm, 2005) and neural levels (Bonath et al., 2007; Colin, Radeau, Soquet, Dachy, &
72 Deltenre, 2002; Colin, Radeau, Soquet, Demolin, et al., 2002; Möttönen, Krause, Tiippana, &
73 Sams, 2002; Saint-Amour, De Sanctis, Molholm, Ritter, & Foxe, 2007; Sams et al., 1991; for
74 review, see Alais, Newell, & Mamassian, 2010).

75 For instance, in the McGurk effect (McGurk & MacDonald, 1976), a mismatch
76 between visual (ga-ga) and auditory (ba-ba) speech phonemes most often results in a
77 perceptual experience of a syllable (da-da) which is not a veridical representation of either of
78 the sensory constituents. In a pioneering magnetoencephalography (MEG) study, Sams et al.
79 (1991) used the McGurk effect to show visual modulation of mismatch responses elicited in
80 the auditory cortex in response to deviant perception of identical sounds. Kislyuk, Möttönen,
81 and Sams (2008) complemented the study by Sams et al (1991) by showing that the MMN
82 elicited in response to deviant speech sounds ('ba' in a train consisting of 'va') may be
83 suppressed by simultaneous presentation of incongruent visual ('va') information. Notably, in
84 both of these experiments, the auditory stimuli were kept constant but generated different
85 mismatch responses across conditions. This highlights the link between MMN elicitation and
86 perceptual experience over and above physical properties of the auditory stimuli.

87 A number of studies have assessed the relevance of congruent visual information to
88 non-illusory speech processing in the auditory cortex, (Calvert et al., 1999; Möttönen,
89 Schürmann, & Sams, 2004; Okada, Venezia, Matchin, Saberi, & Hickok, 2013; Pekkola et
90 al., 2005; Smith et al., 2013). Comparatively less is known about crossmodal influences on

cortical processing of more basic auditory features such as pitch. This is noteworthy considering the longstanding and recently intensified scientific interest in crossmodal correspondences between auditory pitch and visual features such as brightness (Klapetek, Ngo, & Spence, 2012; Marks, 1987), lightness (Melara, 1989; Melara & O'Brien, 1987), angularity (Marks, 1987; Walker, 2012), size (Evans & Treisman, 2010; Gallace & Spence, 2006), and vertical position (Bernstein & Edelstein, 1971; Deroy & Spence, 2016; Melara & O'Brien, 1987; Møller et al., 2018; Orchard-Mills, Van der Burg, & Alais, 2015; Parise, Knorre, & Ernst, 2014; Pratt, 1930; see also Spence, 2011). In a series of studies, Besle and colleagues used electroencephalography (EEG) (Besle et al., 2007; Besle, Fort, & Giard, 2005; Besle, Hussain, Giard, & Bertrand, 2013) and MEG (Besle et al., 2007) to study visual influences on the MMN(m) elicited in response to salient pitch changes. These studies indicated that visual stimuli have access to the processes in the auditory cortex that underlie the generation of MMN(m) responses elicited to non-speech and non-illusory stimuli, specifically *salient* pitch changes. If that is the case, then visually induced behavioral facilitation of *subtle* pitch discrimination should also be evident at the pre-attentive stage of processing probed by the MMNm.

We recently conducted an audiovisual behavioral experiment that capitalized on the inter-individual variability that characterizes pitch discrimination. We showed that the size of visually induced gain in subtle pitch discrimination at the behavioral level is inversely related to participants' musical aptitude and auditory pitch discrimination skills as measured by unisensory tools (Møller et al., 2018). This finding is consistent with an *inter-individual* interpretation of the multisensory *principle of inverse effectiveness* (PoIE) according to which, multisensory integration is maximal when responses to the unisensory dimensions of the multisensory event are weak (Stein & Meredith, 1993). The functional relevance of this principle is obvious: when a unisensory dimension of a multisensory event is hardly

detectable, verification through other sensory modalities is essential. Notably, previous work in support of PoIE at behavioral and neural levels assessed *intra-individual* effects, i.e., by comparing group-averaged responses to varying levels of stimulus salience (Senkowski, Saint-Amour, Hofle, & Foxe, 2011). In our behavioral study, we also found a significant congruence effect, i.e., participants performed better when crossmodally matching than mismatching visual cues were presented simultaneously with target pitch deviants. The question remains whether a similar *inter-individual* interpretation of the PoIE and the congruence effect are evident at the pre-attentive stage of pitch processing probed by the MMNm or whether later levels of processing (e.g., semantic or decisional processes) are exclusively responsible for the pattern of behavioral benefits observed.

Consequently, in the present MEG study we asked whether the visual dimension of audiovisual stimuli could modulate the MMNm elicited in the auditory cortex in response to identical subtle pitch changes, and if so whether the size of the visual modulation depends on the degree of unisensory expertise at the level of the individual. This focus necessitated recruitment of participants spanning a wide range of pitch discrimination abilities. Motivated by previous studies that have shown enhanced MMN amplitudes in musicians compared to nonmusicians (Tervaniemi et al., 2005; Vuust, Brattico, Seppanen, Naatanen, & Tervaniemi, 2012), a subgroup of musicians were added to the participant pool.

We modified the audiovisual oddball paradigm used in the previous behavioral study, allowing us to quantify visually induced gain at the neural level by comparing the amplitude of MMNm responses elicited to deviants differing from the standards in the auditory dimension (pitch), the visual dimension (vertical position), or both dimensions. Notably, when both dimensions varied, they did so in the same or opposite directions of each other, according to the crossmodal correspondence between a high pitch and a high vertical position (Spence, 2011), as this allowed us to assess the effect of crossmodal congruence on pre-

attentive pitch processing. To testify to the applicability in the present study of the MMNm amplitude as an index of neural pitch discrimination sensitivity, we also compared individual measures of MMNm amplitude in all conditions against behavioral measures of pitch discrimination thresholds and of musical aptitude.

These steps were taken towards the main goal of the study, which was to assess inter-individual differences in audiovisual pitch processing. We tested the specific hypothesis that larger visually induced MMNm enhancements would be present in participants with smaller MMNm responses elicited to auditory-only pitch deviants, (i.e., deviants presented without any complementary visual information). This would be consistent with an inter-individual interpretation of the PoIE and with our previously reported behavioral results (Møller et al., 2018), and it would highlight the need for taking individual differences in unisensory neural sensitivity into account when assessing multisensory processing.

Results

Group-level results

Differences in MMNm amplitudes in response to the four conditions *same direction* deviants (AVsame), *opposite direction* deviants (AVopp), *auditory-only* deviants (Adev), and *visual-only* deviants (Vdev) were assessed separately in the two hemispheres using Friedman ANOVAs. Exact statistics are reported. There was a statistically significant main effect of condition in the left, $\chi^2(3) = 21.507, p < .001$, as well as in the right hemisphere, $\chi^2(3) = 20.520, p < .001$. Boxplots showing medians and the interquartile ranges of responses in the different conditions are reported in Figure 1, along with plots showing the average ERF waveforms elicited by the standard and the four deviants. The distributions of individual MMNm responses to all conditions are shown in supplementary materials, Figure S1. Group-averaged gradiometer and magnetometer topographies are also reported in supplementary materials, Figure S2, S3, S4, and S5).

Figure 1. Top: Group-averaged ERF waveforms elicited by the standard and the four deviants. Note, that statistics were performed on the amplitudes of MMNm responses identified in each individual participant (see individual responses to each deviant in supplementary materials, Figure S1). Circles show amplitudes and latencies of the P1m peaks in the four channels used for analysis in each hemisphere. Bottom: Boxplots show medians and the interquartile ranges of responses (individually identified MMNm responses) in the different conditions. * $p < .008$, ** $p < .00167$, *** $p < .000167$

Statistical significance of the subsequent six pairwise comparisons was assessed with Wilcoxon Signed-Ranks tests at a Bonferroni-corrected significance level of $\alpha = .008$. Z statistics, asymptotic p -values, and effect sizes are reported in Table 1. Importantly, significantly larger MMNm amplitudes were observed in both hemispheres in response to AVsame compared to Adev, consistent with the hypothesis. However, the present analysis

did not show evidence that congruity between the auditory and visual components of the stimuli had an effect on the size of the MMNm enhancement, as there was no significant difference between responses to AVsame and AVopp stimuli. Notably, despite no pitch change, an MMNm-like component was observed above the auditory cortex in response to Vdev stimuli. In the right hemisphere, the response was significantly smaller than that elicited by AVsame and AVopp, but the size of it was not significantly different from that elicited by Adev stimuli. In the left hemisphere, the response was significantly larger than the Adev MMNm, but it was not significantly different from that elicited by the AVsame and AVopp stimuli.

Table 1. Pairwise comparisons

Comparison	<i>Z</i>		<i>p</i>		<i>r</i>	
	Left	Right	Left	Right	Left	Right
AVsame vs. Adev	-4.442	-3.798	< .001***	< .001***	-.47	-.40
AVsame vs. AVopp	-0.141	-0.751	.888	.453	-.001	-.08
Adev vs. Vdev	-3.200	-0.528	.001 **	.600	-.34	-.06
AVopp vs. Adev	-4.295	-3.685	< .001 ***	< .001***	-.45	-.39
AVsame vs. Vdev	-1.180	-4.058	.238	< .001***	.12	-.43
AVopp vs. Vdev	-1.022	-2.749	.307	.006*	-.11	-.29

Notes. Wilcoxon Signed-Ranks tests assessed the pairwise comparisons between conditions.

Adev = auditory-only deviant, AVsame = audiovisual, same direction, AVopp = audiovisual, opposite direction, Vdev = visual-only deviant. Asymptotic *p*-values were assessed at Bonferroni-corrected significance levels. * *p* < .008, ** *p* < .00167, *** *p* < .000167.

Visually induced enhancements and unisensory skills

Visually induced MMNm-enhancement was defined as the difference in MMNm amplitude elicited by AVsame compared to Adev stimuli, henceforth denoted ΔMMNm . Behaviorally measured pitch discrimination thresholds did not predict ΔMMNm , neither in the left hemisphere, $r = -.237$, 95% CI $[-.453, .067]$, nor in the right hemisphere, $r = -.047$, 95% CI $[-.258, .364]$, as assessed with Pearson correlation analyses with bootstrapped bias-corrected and accelerated (BCa) 95% confidence intervals (5000 samples). However, the amplitude of MMNm responses elicited to Adev stimuli were statistically significantly inversely related to ΔMMNm , i.e., larger Adev amplitude was associated with smaller ΔMMNm , $r = -.397$, 95% CI $[-.595, -.194]$, in the right hemisphere. See scatterplots in Figure 2. Although no such statistically significant association was found in the left hemisphere, $r = -.183$, 95% CI $[-.399, .051]$, the correlation was significant when responses of the left and right hemisphere were averaged, (see supplementary materials, Figure S6).

Figure 2. Plots show the visually induced enhancement in MMNm amplitude, ΔMMNm , as a function of the MMNm amplitude elicited in response to auditory-only deviants, Adev, i.e., without complementary visual cues. The correlation is significant in the right hemisphere as assessed with bootstrapped bias-corrected and accelerated (BCa) 95% confidence intervals (5000 samples). Solid circles = musicians, open circles = nonmusicians

213 **Behavioral correlates to MMNm amplitude**

214 Bootstrapped Pearson's correlation analyses were run to examine the relationship between
 215 MMNm amplitude and behavioral scores obtained in the Musical Ear Test (MET) (Wallentin,
 216 Nielsen, Friis-Olivarius, Vuust, & Vuust, 2010) and the pitch discrimination threshold (PDT)
 217 estimation. Statistical significance was assessed at $\alpha = .05$. PDT data points were
 218 nonnormally distributed and hence log-transformed using the natural logarithm. Left-
 219 hemisphere MMNm amplitudes in response to all four experimental conditions were
 220 significantly correlated with PDT and with MET(melodic) scores, i.e., larger MMNm
 221 amplitudes were associated with lower (i.e., better) pitch discrimination thresholds and with
 222 better performance on the MET, testifying to the applicability in the present study of the
 223 MMNm amplitude as an index of neural pitch discrimination sensitivity. No significant
 224 associations were found between behavioral measures and MMNm amplitudes elicited in the
 225 right hemisphere. See Figure 3 for scatterplots and Table 2 for correlation coefficients and
 226 bootstrapped bias-corrected and accelerated (BCa) 95% confidence intervals. A Pearson
 227 product-moment correlation analysis showed that MET and PDT scores were also
 228 significantly correlated ($r = -.528, p < .001$), reflecting the association between musical skills
 229 and auditory sensitivity.

Table 2. Correlations between behavioral measures and MMNm amplitudes in the four conditions

	Left hemisphere		Right hemisphere	
	PDT	MET(melodic)	PDT	MET(melodic)
	<i>r</i> [lower, upper]	<i>r</i> [lower, upper]	<i>r</i> [lower, upper]	<i>r</i> [lower, upper]
Adev	-.291* [-.508, -.032]	.385* [.073, .684]	-.140 [-.389, .137]	.080 [-.164, .311]
AVsame	-.380* [-.572, -.117]	.322* [.052, .543]	-.042 [-.342, .274]	.130 [-.144, .379]
AVopp	-.471* [-.648, -.259]	.395* [.147, .613]	-.074 [-.355, .238]	.084 [-.165, .332]
Vdev	-.324* [-.524, -.059]	.345* [.070, .582]	.094 [-.235, .426]	.135 [-.167, .421]

Notes. Pearson correlation coefficients (*r*) with bootstrapped bias-corrected and accelerated (BCa) 95% confidence intervals (CI) in hard brackets, *n* = 45. * = *r* is statistically significant at α = .05, as indicated by 95% CIs not including 0. PDT= Pitch discrimination threshold; MET(melodic) = melodic part of the Musical Ear Test; Adev = auditory-only deviant, AVsame = audiovisual, same direction, AVopp = audiovisual, opposite direction, Vdev = visual-only deviant.

Figure 3. Scatterplots show MMNm amplitudes (y-axes) plotted against behaviorally measured log-transformed pitch discrimination thresholds (PDT) and behavioral scores on the melodic part of the Musical Ear Test (MET). For all conditions, the correlation is significant in the left hemisphere as assessed with bootstrapped bias-corrected and accelerated (BCa) 95% confidence intervals. PDT= Pitch discrimination threshold; MET(melodic) = melodic part of the Musical Ear Test; Adev = auditory-only deviant, AVsame = audiovisual, same direction, AVopp = audiovisual, opposite direction, Vdev = visual-only deviant. Solid circles = musicians, open circles = nonmusicians

Discussion

This study was designed to assess inter-individual differences in visually induced enhancements of subtle pitch discrimination at the pre-attentive stage of perceptual processing. Across all participants (musicians and nonmusicians), the amplitude of the MMNm recorded using MEG was significantly enhanced in conditions containing audiovisual deviants (AVsame and AVopp) compared to a condition in which only the auditory dimension of the audiovisual stimuli deviated (Adev). In the right hemisphere, the size of this enhancement was significantly correlated with the amplitude measured in the Adev condition, indicating that the visual dimension of audiovisual events has a larger impact on pre-attentive pitch discrimination in the auditory cortices of people with lower auditory-only pitch discrimination sensitivity. This result is consistent with an inter-individual interpretation of the *principle of inverse effectiveness* (PoIE) and suggests not only that individual differences in multisensory processing exist, but also that they may be explained in part by individual differences in unisensory sensitivity.

Our study contributes to the literature showing that processes occurring early in sensory-specific cortices can be modulated by crossmodal information (see also Bulkin & Groh, 2006; Driver & Noesselt, 2008; Ghazanfar & Schroeder, 2006). We focused specifically on visually induced enhancements of pitch discrimination in the auditory cortex. It is important to note that perceptual enhancement is but one of several possible outcomes of multisensory integration (Driver & Noesselt, 2008) and also that the MMN(m) has been interpreted specifically as indexing the ability of the brain to detect incongruent relations between auditory and visual stimuli, i.e., as an incongruency response (Paraskevopoulos, Kuchenbuch, Herholz, & Pantev, 2012 b). Our interpretation of the MMNm as an index of perceptual sensitivity rather than an incongruency response rests on previous studies showing that the MMN is related to behavioral performance in tasks that employ similar sounds

(Jaramillo et al., 2000; Novitski et al., 2004) and in groups of participants with different levels of auditory abilities, i.e., musicians and nonmusicians (Tervaniemi et al., 2005). This is also confirmed in our data where the correlations between MMNm amplitudes and behavioral measures of pitch discrimination abilities and musical aptitude testify to the applicability of the MMNm amplitude as an index of pre-attentive neural pitch discrimination.

Group-level results

Visually induced MMNm enhancement was evident in all conditions of the experimental paradigm in which the visual dimension of the audiovisual event deviated (AVsame, AVopp, Vdev) compared to the condition in which it did not (Adev). Our primary interest was the assessment of the difference in MMNm amplitude elicited by AVsame compared to Adev stimuli, i.e., Δ MMNm. This was the most important outcome variable as the key objective of the study was to assess whether this reliance followed the PoIE at the inter-individual scale, i.e., varied with varying levels of auditory-only pitch discrimination abilities. At the group level, this difference was highly significant in both hemispheres, consistent with the notion that visual information not only induces a qualitatively different perceptual experience and hence elicits an MMNm (Möttönen et al., 2002; Saint-Amour et al., 2007; Sams et al., 1991), but that it also modulates the size of the MMNm (Besle et al., 2013) elicited to identical auditory deviants. A discussion of the individual-level results will follow after further considerations of the group-level results.

The comparison between AVsame and AVopp constitutes what we have previously referred to as the crossmodal congruence effect (Møller et al., 2018). Note, that we used the feature-based crossmodal correspondence between auditory pitch and visually perceived vertical position because we sought to maximize the likelihood that the visual stimuli would be beneficial. In this perspective, the disentanglement of two kinds of facilitatory effects (i.e., Δ MMNm and congruence effect) was not as much an aim as a by-product of the

paradigm, caused by the polarity inherent in the audiovisual stimuli, where the polar dimensions of auditory pitch and visually perceived vertical position map onto each other. From this also follows that we refer to crossmodal congruence in relative terms, in contrast to studies that conceptualize audiovisual congruence in absolute terms, i.e., violations of absolute distances in the same direction are defined as audiovisual incongruence (Nichols & Grahn, 2016) or no change in the visual dimension (comparable to our Adev) is considered an incongruent trial (Paraskevopoulos, Kuchenbuch, Herholz, & Pantev, 2012 a). To avoid confusions with the inconsistent use of the terms evident in the literature, we refrain from using the labels congruent and incongruent to denote trials, but use congruence effect merely to denote the comparison between AVsame and AVopp. We did not find a significant congruence effect in the present study, despite the fact that we did so in our previously reported behavioral study using similar stimuli. The most obvious candidate explanation for this is that the feature-based crossmodal correspondence used does not influence pre-attentive levels of perceptual processing probed by the MMNm. This interpretation of the result taps into the debate about the level at which these crossmodal interactions take place (Deroy & Spence, 2016; Parise & Spence, 2009) and could be taken as an argument that this specific correspondence only influences participants' responses at semantic and/or decisional levels of processing. Yet, first of all, one should be careful of drawing conclusions based on null-results, and secondly, there is still the possibility that less salient visual stimuli than the ones used in the present paradigm would lead to significant congruence effects, as bimodal response enhancement is most pronounced when both stimulus dimensions in isolation lead to weak responses, according to the PoIE (Stein & Meredith, 1993). As is evident in the topographies of the conditions involving changes in the visual stimuli, this is clearly not the case here (see supplementary materials, Figure S2, S3, S4, S5).

Although the visual-only deviant (Vdev) was not an explicit focus of our study, as explained in the methods section, we briefly provide two possible yet still speculative interpretations of the finding that an MMNm comparable in size to the Adev (right hemisphere) and AVsame and AV opp (left hemisphere) was elicited to Vdev, despite no change in the auditory dimension of the stimulus. The first is based on a simple idea of classical conditioning of the auditory cortex (Pavlov, 1927). Because all conditions were included in the same paradigm, it is possible that the MMNm response to Vdev was a conditioned reflex, i.e., the result of a conditioning process in which the auditory and visual dimensions of the audiovisual stimulus have been repeatedly coupled. Crossmodal conditioning and anticipatory excitation of the cortex was demonstrated already in the early 1940's (Anokhin, 1974) and more recent findings also support the notion that the perceptual system easily adapts its predictions even to unnatural stimulus events through the course of an experiment (van Laarhoven, Stekelenburg, & Vroomen, 2017). Indeed, in the visual modality it has been shown that after a visuohaptic training session, haptic shape processing activates visual cortex in the absence of visual input (Lee Masson, Bulthé, Op de Beeck, & Wallraven, 2016).

An alternative, yet most likely related account, may be rooted in the fundamental object-directed nature of perception, and the fact that spatiotemporal coherence is key to multisensory object formation processes (Atilgan et al., 2018; Lewkowicz & Ghazanfar, 2009; Parise, Spence, & Ernst, 2012). Although we presented participants with simple artificial auditory and visual stimuli, their consistent temporal synchronicity may have made them so convincing audiovisual objects, that an MMNm was elicited whether the auditory, the visual, or both dimensions of the audiovisual stimulus deviated. MMNm elicitation is not restricted to conditions in which the physical properties of the sounds deviate, but may also be elicited to violations of abstract feature patterns (Näätänen et al., 2007). Assuming the

perceptual system recognizes the stimuli as integrated audiovisual percepts, it is not surprising that a violation in any of its dimensions give rise to an abstract-feature MMNm. Validation of such an account, however, would require demonstrations that similar processes occur in the visual cortices, which is beyond the scope of the present study.

Individual-level results

This is the first study to quantify the size of visually induced MMNm enhancement across participants with varying levels of unisensory auditory abilities in order to show that the MMNm amplitude enhancement is larger in individuals with smaller amplitudes elicited to auditory-only deviants. This result matches our previously reported behavioral findings of larger gains in pitch discrimination with poorer pitch discrimination thresholds (Møller et al., 2018), and is consistent with an inter-individual interpretation of the PoIE.

Similar patterns of observation have been found at the intra-individual level in studies investigating vision as the target modality. Noesselt and colleagues (2010) employed group-level analyses of fMRI data to show sound-induced sensitivity enhancements to low-intensity but not high-intensity visual stimuli. Stevenson and colleagues (2012) recorded visual ERPs across salience levels of audiovisual speech and found greater multisensory gains as salience was reduced. The role of uni-sensory expertise was assessed by Caclin and colleagues (2011) who investigated sound-induced enhancements of visual detection thresholds and found that behavioral enhancements were only evident in a group of participants with poor performance in a visual-only condition. Similarly, in the auditory domain, visual facilitation of behavioral pitch discrimination of quite salient pitch changes (1/4 tone, i.e., 50 cents) has been found only in a group of participants with poor auditory-only abilities (Albouy et al., 2015), while conversely, only the control group exhibited visual facilitation of the subtle (1/16 tone, i.e., 12.5 cents) pitch changes. Senkowski and colleagues (2011) studied effects of stimulus intensities on both auditory and visual modalities and found evidence that early multisensory

interactions in the human brain follow the PoIE. Our results extend these group-level findings to show that the PoIE is also applicable to pre-attentive auditory processing at the inter-individual level, highlighting the value of taking individual differences in unisensory abilities into account when assessing multisensory perceptual processing.

Principle of inverse effectiveness

Though the PoIE was derived from studies of single multisensory neurons in the midbrain (Stein & Meredith, 1993), it not surprising to find inverse effectiveness as a general principle across a wide range of processing levels and functions (Albouy et al., 2015; Caclin et al., 2011; Møller et al., 2018; Noesselt et al., 2010; Senkowski et al., 2011; Stevenson et al., 2012). This is the case because multisensory processing does not take place in individual neurons or even cortical areas in isolation. The development of the multisensory processing capabilities of individual neurons in the superior colliculus depend on their embeddedness in a neural network consisting of lower-order subcortical areas as well as higher-order association cortices, and on the organism's experience in the environment (Stein, Stanford, & Rowland, 2014).

Similarly, our findings of inverse effectiveness at the pre-attentive level of neural processing in the auditory cortex should not be taken to suggest, that other areas of the brain do not play a role in the kind of multisensory processing studied here. Rather, the interpretation put forth here is similar to previous accounts of multisensory influences on sensory specific cortices that highlight the fundamental object-directed nature of perception and how object formation processes in the brain are made possible through the spatio-temporal coherence of crossmodal information belonging to the same event (Atilgan et al., 2018; Lewkowicz & Ghazanfar, 2009; Parise et al., 2012). Specifically, it has been shown that phase resetting of ongoing neuronal oscillations is a key mechanisms underlying the amplification of neuronal responses. This phase resetting occurs when bimodal stimuli are

temporally aligned, and ensures that sensory-specific information arrives in sensory-specific cortices during a high excitability phase resulting in perceptual amplification (Atilgan et al., 2018; see also Lakatos, Chen, O’Connell, Mills, & Schroeder, 2007; Maddox, Atilgan, Bizley, & Lee, 2015). As such, it is not the visual information itself that gives rise to excitation of neurons in the auditory cortex, but rather the excitability of the auditory neurons that is enhanced, mediated by the crossmodal synchronicity. “Top-down” influences on sensory-specific cortices are thus crucial in amplifying the salience of the sensory-specific information, in this case leading to the experience of improved hearing.

The design of this study was originally motivated by the PoIE, however, the results may also be interpreted within theoretical accounts that consider multisensory percepts as a weighted average of sensory estimates (Ernst & Banks, 2002; Schwartz, 2010). Like the PoIE, these models highlight the flexibility of our perceptual systems to adapt to variations in the sensory environment (darkness, noise). It follows, that if the reliability of the auditory dimension of the audiovisual events in our experiment is considered low in some participants and high in others, then they will give more weight to the visual and auditory dimensions, respectively. This is in line with recent evidence that musicians who specialize in the auditory domain are less susceptible to audiovisual illusions (Bidelman, 2016; Proverbio, Massetti, Rizzi, & Zani, 2016). Our finding that such effects are evident in the MMNm amplitude fits well with recent research showing that reliability-based cue weighting occurs early (< 150ms) in the audio-visual integration process (Boyle, Kayser, & Kayser, 2017), i.e., well before the latencies of the MMNm responses in the present study.

Although the correlations between Δ MMNm and Adev showed similar trends in the two hemispheres, it was only statistically significant in the right hemisphere. A possible explanation for this discrepancy could be rooted in the fact that larger amplitudes were seen in the right than in

the left hemisphere. It may be the case that these larger amplitudes enabled detection of a significant correlation in the right hemisphere, whereas due to response homogeneity in Adev in the left hemisphere, there may not have been enough variance to detect a statistically significant correlation.

Analysis limitations

Explanations for functional hemispheric asymmetries are highly relevant to discuss, particularly for auditory processing in the human brain. Apart from merely statistical accounts, however, there are notable reasons why further interpretations regarding lateralization can and should not be made based on the analyses reported here. Most importantly, we ran separate analyses in the two hemispheres, leaving interpretations of differences in evidence between them unsupported. Moreover, while it has been repeatedly shown that the MMN elicited to non-speech auditory stimuli is more pronounced in the right hemisphere of the general population (Alho et al., 1996; Giard & Peronnet, 1999; Tervaniemi et al., 1999), and particularly so for change detection (Levänen, Ahonen, Hari, McEvoy, & Sams, 1996), it has also been shown that expert musicians exhibit the reverse pattern, i.e., larger MMNm amplitudes in the left hemisphere (Herholz, Lappe, & Pantev, 2009; Kuchenbuch, Paraskevopoulos, Herholz, & Pantev, 2012; Vuust et al., 2005). Furthermore, Foxton and colleagues (2009) provided evidence that people with difficulty in pitch processing show greater recruitment of the left than the right hemisphere when performing active determination of pitch glide direction, a finding which may explain their poorer behavioral performance, which in turn was also evident at the pre-attentive level as assessed with the MMNm. Clearly, effects of expertise are likely to interact with any lateralization patterns evident in the present data, considering the fact that musicians and nonmusicians were not separated in the present analysis. Again, the reason for running the analyses across all participants was to ensure sufficient variance in the data to investigate individual

differences, a focus that was prioritized over that of lateralization. For transparency, we have reported one group-level analysis for each hemisphere. Note, however, that when the responses were averaged across hemispheres within participants, a procedure justified by the particular focus on individual differences, the correlation between Δ MMNm and Adev was statistically significant, consistent with the PoIE (see supplementary materials, Figure S6).

We acknowledge that a sensor-space analysis limits the conclusions we can make regarding the sources of the neural activity measured. However, there is little doubt that the MMNm is elicited in the auditory cortex, and maximally detected with gradiometers right above the local cortical source (Levänen et al., 1996; Luck, 2005; Näätänen et al., 2007), which was also confirmed in the present data at the individual as well as the group level. Furthermore, a recent study showed that MMNm detectability at the individual level is similar whether sensor-space or source-space analyses are employed. Hence, a sensor-space analysis constrained to the four gradiometer pairs of each hemisphere that picked up the largest MMNm response was considered appropriate in the present context, considering the focus on the MMNm and the fact that the MMN is very well described in the literature as a component with primary sources in the auditory cortices (Alho, 1995; Näätänen et al., 2007).

Conclusion

Most research on perceptual processing involves stimulation of a single sense. Yet, even perceptual experiences that we consider sensory-specific, such as hearing, are influenced by information in other sensory modalities. We have shown visually induced enhancement of subtle pitch discrimination already at the pre-attentive stage of auditory processing probed by the MMNm. Importantly, participants with lower sensitivity to auditory-only deviants gain more from complementary visual information, in accordance with the multisensory *principle of inverse effectiveness*. In the quest towards understanding hearing as it unfolds outside the

lab, we may gain substantially by paying attention to the possibility that hearing is related in systematic ways to other factors than sound waves.

Materials and methods

The present magnetoencephalography (MEG) experiment forms a separate part of a larger study. In the larger study, the same participants subsequently completed a behavioral experiment (reported in Møller et al., 2018) and participated in a study employing magnetic resonance imaging (MRI) and diffusion tensor imaging (DTI) (manuscript in preparation). Behavioral measures included pitch discrimination threshold (PDT) estimations, a musical aptitude test, i.e., the Musical Ear Test (MET) (Wallentin et al., 2010), two questionnaires, i.e., the Goldsmiths Musical Sophistication Index (Gold-MSI) (Müllensiefen, Gingras, Musil, & Stewart, 2014), and the Autism Spectrum Quotient (AQ) (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001), as well as a custom-made test of pitch direction sensitivity. We report results based on the MEG data and correlate these to the behavioral measures from the MET and PDT tests. MET and PDT data were also analyzed in relation to our previously reported behavioral experiment (Møller et al., 2018), however using a smaller sample of the participants. The PDT data were also analyzed in conjunction with the DTI data (manuscript in preparation).

No explicit power analysis was used to compute appropriate sample size due to the novelty of the specific topic under study. However, sample size was established by comparisons with sample sizes used in behavioral studies of crossmodal gains and inverse effectiveness (Albouy et al., 2015; Caclin et al., 2011; Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004; Senkowski et al., 2011), where reasonable effect sizes were found with sample sizes between 11 and 34. Accordingly, we considered recruitment of ~50 participants appropriate as we expected a somewhat high exclusion rate in the subsequent behavioral study due to chance and ceiling level performance.

Participants

MEG data were recorded from 48 participants. Data from one participant were discarded from all analyses due to stimulus presentation failure. Because of the present focus on unisensory abilities, an *a priori* decision was made to exclude data from participants from whom no pitch threshold estimation was made. This was the case for two individuals. Hence, data from 45 participants (mean age = 24.2 years, $SD = 3.6$, 23 female) were included in the analyses reported here. As described in the introduction, a sub-group of these were professional musicians. Specifically, participants included 29 nonmusicians (mean age = 24.1 years, $SD = 3.3$, 16 female) and 16 musicians (mean age = 24.3 years, $SD = 4.1$, 7 female). All were right-handed, all reported normal or corrected-to-normal visual acuity and no hearing impairments.

Participants gave their written consent before participation and received a taxable compensation of DKK 400 for participating in the larger study which took place on two separate days. The study protocol was approved by The Central Denmark Regional Committee on Health Research Ethics (Project-ID: M-2014-52-14).

506

507 *Figure 4. Stimuli and paradigm. The experiment consisted of a semi-attended oddball paradigm with*
 508 *80% standards and eight different deviants, categorized into four deviant categories: same direction*
 509 *deviants (AVsame), opposite direction deviants (AVopp), auditory-only deviants (Adev), and visual-*
 510 *only deviants (Vdev). 12.1% of the standards (always the first in a train) were colored control task*
 511 *stimuli. In order for participants to stay focused on the visual stimuli, they were instructed to count one*
 512 *of four colors per block and report the count orally to the experimenter after each block.*

Stimuli and paradigm

The auditory stimuli consisted of sinusoidal tones of 100 ms duration, including 5 ms fade in/out followed by a 316 ms inter-stimulus interval (ISI). Standard tones were 523.25 Hz (corresponding to C5 on a piano), deviant tones were 20 cents higher (i.e., 529.33 Hz) and 20 cents lower (i.e., 517.24 Hz) with respect to the standard. The visual stimuli consisted of an image of a rectangle containing a light grey disc (see Figure 4), which subtended a visual angle of 3 degrees watched from a distance of 135 cm. The rectangle was static, yet the disc was positioned in three vertical positions: either above, below, or in the center of the rectangle, which was clearly marked by the crossing of two lines serving as a fixation cross. The displacement of the disc from the center was 0.5 degrees visual angle, i.e., small enough to be perceived as an excursion of the disc rather than as the sudden pop up of a new disc, but large enough to be clearly visible. The duration of the visual stimuli was 416 ms, i.e., there was no ISI. This was preferred because the perceptual salience of a flickering visual stimulus (i.e., one with a 316 ms ISI) much exceeded the perceptual salience of the excursion of the disc.

The audiovisual (AV) stimuli comprised all possible combinations of the three auditory and three visual stimulus types, resulting in nine different stimulus pairs. An AV standard consisted of a standard tone coupled with a disc in center position. With higher pitch corresponding to higher vertical position, the eight AV deviants were categorized into *same direction* deviants (AV_{same}) (i.e., high pitch with disc above the center, or low pitch with disc below the center), *opposite direction* deviants (AV_{opp}) (i.e., high pitch with disc below the center, or low pitch with disc above the center), and *auditory-only* deviants (A_{dev}) (high pitch or low pitch with no excursion of the disc from the center). A *visual-only* deviant (V_{dev}) (no pitch change with disc below or above the center) also appeared in the paradigm. However, because of the focus on putative visual modulation of the auditory MMNm, the

present paradigm was not designed to assess visual MMNm. Indeed, despite the visual only deviant included, it is not possible to extract proper vMMNm from it that can be likened to the vMMNm reported in the literature (Stefanics, Astikainen, & Czigler, 2015; Winkler & Czigler, 2012). This is the case because no actual visual standard *events* are present in a visual stream with no ISI, except on the standard immediately following the deviants, which in turn is contaminated by the response to the preceding deviant. The purpose of including the visual deviant nonetheless was to reduce predictability by ensuring that a change in the visual stream was not always accompanied by an auditory deviant.

Presentation software (Neurobehavioral systems Inc., Albany, CA, USA) was used to present the stimuli. The timing of the onset of auditory and visual stimuli was adjusted to mimic sound source perception from the perspective of the participant. This included ensuring that the sounds arrived at participants' ears with the same delay (4 ms) as if they had been emitted from the position of the visual images themselves. The AV stimuli were presented with a stimulus onset asynchrony (SOA) of 416 ms in an oddball paradigm with 20% deviants (2.5% per deviant, 5% per deviant category). 12.1% of the remaining stimuli were control task stimuli. They were similar to the standard stimuli with respect to auditory frequency and position of the disc, yet the disc was colored: green, red, blue, yellow, all matched with respect to the luminance of the grey discs. These control task stimuli were always presented as the first stimulus after a deviant. Each of the eight deviants was presented 100 times, yet pairs of deviants that fell into the same category (Adev, AVsame, AVopp, Vdev, respectively) were concatenated in the analysis to increase the signal-to-noise ratio, yielding 200 trials per deviant category. The deviants were pseudo-randomly presented with a minimum of three and a maximum of seven standards between two deviants. As such, the distribution of the number of standards was centered on four, however, a right skew was enforced in the distribution to give relatively few instances of six and seven standards. This

approached was preferred in order to reduce the predictability arising from equiprobable stimulus presentation, which may attenuate the MMNm amplitude. The total number of audiovisual stimuli was thus 4000, 800 of which were deviants. They were presented in four identical blocks preceded by 5 instances of AV standards. The duration of each block was 7 min. In addition to the four AV blocks, the experiment included two unimodal control blocks, presented as the third and fifth of the six blocks; the presentation order was counterbalanced across participants within each group. The unimodal visual control block was identical to the AV block except that the sound was turned off. In the unimodal auditory control block, which also contained 200 deviant trials (100 high, 100 low pitch trials), the visual stimuli were replaced by a silent Charlie Chaplin movie.

Procedure

During MEG recordings, participants were seated in a magnetically shielded and sound-attenuated room with ambient lighting. Visual stimuli were projected onto a screen with a refresh rate of 60 Hz. The screen was positioned 135 cm in front of the participant. Auditory stimuli were presented via in-ear headphones (Etymotic ER•30) at a comfortable hearing level of approximately 70 dB sound pressure level (SPL). Before the experiment, a short experimental procedure varying interaural-level-difference (ILD) was administered to ensure that participants perceived the sounds as originating straight in front of them (i.e., SPL was perceived as equal in the two ears). Participants were informed about the presence of sounds and images in the experiment, and instructed to focus on the cross at the center of the screen, pay attention to the color changes of the disc and silently count the number of occurrences. There were 28 green, 32 red, 37 blue and 24 yellow discs in each block. Participants counted each color separately, and reported the count orally after each block. This task applied to the four AV blocks as well as to the visual only block in which the targets were green discs. In the auditory only block, participants were encouraged to take a break from the task and enjoy

the movie. In all encounters with the participants, we referred to the auditory stimuli as “sounds” rather than “notes” or “tones” in order not to induce a semantic confound into the experiment by using a vocabulary that is often associated with the relative terms “high” and “low”. This was particularly important as all instructions were in Danish and the Danish word for “pitch” literally translates into “tone height” (*tonehøjde*). The duration of the experiment was approximately 50 minutes, including breaks between the blocks.

Pitch threshold estimation

In a separate session, two to four weeks after the MEG session, individual pitch discrimination thresholds (PDTs) were estimated using a *two-down, one-up* adaptive staircase procedure that converges on the 70.7% performance level on the psychometric function (Levitt 1971). Using an AXB forced-choice task, the reference (X) was always a 523.25 Hz sinusoidal tone, and the task was to state whether the first (A) or the last (B) tone differed from the two other tones. The staircase terminated after 14 reversals and the threshold was calculated on the basis of an average of the last 6 reversals. Depending on participants’ responses, the duration was approximately three to five minutes. The staircase procedure was similar to the one used by Williamson, Liu, Peryer, Grierson, and Stewart (2012), yet adapted to match the stimuli and participants of the present study. The pitch threshold estimation data were also included in the analysis of a behavioral experiment reported in Møller et al., (2018).

Musical abilities assessment

To assess the musical abilities of the participants, we used the melodic part of the Musical Ear Test (MET) (Wallentin et al., 2010). In the melodic subtest, participant were asked to judge whether pairs of melodic piano phrases were identical or not. MET scores have been shown to correlate with results of musical imitation tests typically used in auditions for music conservatories.

MEG recording

Continuous MEG data were sampled at 1000 Hz using Elekta Neuromag Triux™ whole-head 306-channel MEG system, with 204 planar gradiometers and 102 magnetometers. This system is part of the MINDLab Core Experimental Facility at Aarhus University Hospital, Denmark. Participants' head positions were recorded using four continuous Head Position Indicator (cHPI) coils, and bipolar surface electrodes positioned above and below the right eye (vertical EOG), at the outer canthi of both eyes (horizontal EOG), and on the right clavicle and left rib (ECG) recorded eye blinks and cardiac activity, respectively.

MEG preprocessing

Initial preprocessing was performed using the temporal extension of the signal space separation (tSSS) technique (Taulu, Kajola, & Simola, 2004; Taulu & Simola, 2006)) implemented in Elekta's MaxFilter software (Version 2.2.15). This involved removal of noise from electromagnetic sources outside the head and compensation for within- and between-block differences in head position as well as continuous head movement compensation using cHPI. EOG and ECG artifacts were detected and removed from the maxfiltered data with independent component analysis (ICA) using the *find_bads_eog* and *find_bads_ecg* algorithms in MNE Python (Gramfort et al., 2013, 2014). The topographies and averaged epochs of all identified EOG- and ECG-related components were visually inspected for all participants to ensure the validity of rejected components. The remaining data were band-pass filtered, using an infinite impulse response (IIR) filter with cutoff frequencies of 1 and 40 Hz.

The subsequent preprocessing was conducted using FieldTrip (Oostenveld, Fries, Maris, & Schoffelen, 2011). The MEG data were downsampled to 250 Hz and separated in epochs of 516 ms starting 100 ms before the onset of the trial and ending at the onset of the following trial. Baseline correction was applied based on the 100 ms pre-stimulus window.

Trials containing SQUID jumps were discarded using automatic artefact rejection with a z-value cutoff of 30.

Planar gradiometers pick up the largest signal just above a local source where the gradient has its maximum (Levänen et al., 1996; Luck, 2005). In contrast, magnetometers are more likely to also pick up activity from deeper sources and elsewhere in the cortex (Vrba & Robinson, 2002). Hence, because our focused interest was in the modulation of the neural signals emanating specifically from the auditory cortex, the sensor space analysis focused on combined planar gradiometer recordings from immediately above auditory cortex. Planar gradiometer recordings were calculated as vector sums by taking the root-mean-square (RMS) of the two gradients at each gradiometer pair resulting in a single positive RMS value for each pair.

Channel selection was accomplished through the use of the MMNm response obtained in the unimodal auditory control block. As such, the well-described MMNm response obtained in a classic oddball paradigm as used in our control block served as a functional localizer at the sensor level (Luck & Gaspelin, 2017). Initially, group-averaged difference waveforms for each channel were obtained by first subtracting each participant's average response to the third standard after a deviant from their average response to the deviants, and then averaging across participants. Peak amplitudes and latencies of the P1m, i.e., the early, sensory-driven component which was prominent for all sounds in the present data, were identified in the standard waveforms within a 40-80 ms time window from stimulus onset. On the difference waveforms, local MMNm peaks were identified within a 100-200 ms time window following the timepoint of the P1m peak as it appeared in the standard waveform. This was accomplished using Matlab's *findpeaks* function, which identified data samples with a greater voltage than each of its neighboring data samples. A polarity switch in the magnetometer topographies between the P1m and the MMNm

components was confirmed through visual inspection of the group-averaged topographies (see Figure 5).

Figure 5. Top: Channel selection was accomplished using combined gradiometer data recorded in the unisensory auditory control block. The MMNm response obtained in this classic oddball paradigm served as a functional localizer in the spatial domain. Middle: The four channels of each hemisphere that picked up the largest group-averaged MMNm response defined the sensors-of-interest in the subsequent analyses of data obtained in the audiovisual experimental blocks. Circles in the group-averaged ERF waveforms show amplitudes and latencies of the P1m peaks in the four channels. Bottom: A polarity switch in the magnetometer data confirmed that the magnetic field of the identified MMNm peak was indeed of opposite polarity to the P1m.

MMNm amplitudes were defined as the average within a 20 ms time window centered at the MMNm peak latencies. Having identified the MMNm amplitudes for each channel in the unimodal auditory control block, all subsequent analyses were constrained to the four sensors of each hemisphere picking up the largest group-averaged MMNm response (see Figure 5) in the unimodal auditory control block. Rather than using just the peak channel, the average of four channels was used to account for individual variations in head position, size, etc. The four selected channels were also the channels with the largest P1m amplitude in the group-average standard waveform. To confirm the relevance of these four channels at the level of the individual participants, they were compared to the P1m peak channel of each participant for each hemisphere. Individual P1m amplitudes and latencies were identified within the 40-80 ms time window from stimulus onset and the P1m peak channel was defined as the channel containing the largest peak. Except for two participants whose P1m peak channel in the left hemisphere was a more posterior neighboring channel, the P1m peak channel of all participants was identical to one of the four selected channels of each hemisphere.

Due to the large inter-individual variability of the ERF difference waveforms in the experimental conditions, individual local MMNm peaks and latencies of each hemisphere and each condition were identified for each participant individually using the same procedure as described for the functional localizer, i.e., within a 100-200 ms time window following the individual participant's P1m peak in the auditory-only block. A summary value of the MMNm amplitude was extracted for each participant, each hemisphere and each condition (see individual participants' responses in supplementary materials, Figure S1). In cases where the MMNm peak was exceeded in amplitude by a peak which did not show opposite polarity to the P1m on the magnetometer topography, the window was narrowed further to avoid the spurious peak being picked up by the fully automated procedure. Similarly, plots of

699 waveforms and topographies were compared across conditions within each hemisphere for
 700 each participant. Inconsistencies were typically found when MMNms were less prominent,
 701 and the search criteria were adapted when applicable to ensure consistency between the
 702 automatically selected peaks. Altogether, further constraints were necessary in one or both
 703 hemispheres for 10 participants in the unimodal auditory control block and in 1-5 out of 8
 704 instances (4 conditions x 2 hemispheres) for 11 participants in the audiovisual experimental
 705 blocks, accounting for 9.3 % of the total number of peaks.

706 All statistical analyses were performed in IBM SPSS Statistics 24. Bootstrapping with
 707 5000 samples was used to estimate 95% bias-corrected and accelerated (BCa) confidence
 708 intervals (CIs), when this more robust method was required due to skewness in the
 709 distribution of one or more variable outcomes.

710 Funding

711 This work was supported by seed funding from the Interacting Minds Centre, Aarhus
 712 University, Denmark; Center for Music in the Brain is funded by the Danish National
 713 Research Foundation (DNRF117).

715 Acknowledgements

716 We wish to thank Signe Hagner for assistance in data collection, and Christopher Bailey
 717 for valuable technical and scientific advice at various stages of MEG recording and analyses.

References

- Alais, D., Newell, F. N., & Mamassian, P. (2010). Multisensory processing in review: from physiology to behaviour. *Seeing and Perceiving*, 23(1), 3–38.
<https://doi.org/10.1163/187847510X488603>
- Albouy, P., Leveque, Y., Hyde, K. L., Bouchet, P., Tillmann, B., & Caclin, A. (2015). Boosting pitch encoding with audiovisual interactions in congenital amusia. *Neuropsychologia*, 67(Journal Article), 111–120.
<https://doi.org/10.1016/j.neuropsychologia.2014.12.006>
- Alho, K. (1995). Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear and Hearing*, 16(1), 38–51.
- Alho, K., Tervaniemi, M., Huottilainen, M., Lavikainen, J., Tiitinen, H., Ilmoniemi, R., ... Näätänen, R. (1996). Processing of complex sounds in the human auditory cortex as revealed by magnetic brain responses. *Psychophysiology*, 33(4), 369–375.
- Amenedo, E., & Escera, C. (2000). The accuracy of sound duration representation in the human brain determines the accuracy of behavioural perception. *European Journal of Neuroscience*, 12(7), 2570–2574.
- Anokhin, P. K. (1974). *Biology and Neurophysiology of the Conditioned Reflex and Its Role in Adaptive Behavior* (Vol. 3). Oxford: Pergamon.
- Atilgan, H., Town, S., Wood, K., Jones, G., Maddox, R., Lee, A., & Bizley, J. K. (2018). Integration of visual information in auditory cortex promotes auditory scene analysis through multisensory binding. *Neuron*, 97(3), 640–655.e4.
<https://doi.org/10.1016/j.neuron.2017.12.034>
- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism,

- 742 males and females, scientists and mathematicians. *Journal of Autism and*
 743 *Developmental Disorders*, 31(1), 5.
- 744 Bernstein, I. H., & Edelstein, B. A. (1971). Effects of some variations in auditory input upon
 745 visual choice reaction time. *Journal of Experimental Psychology*, 87(2), 241–247.
- 746 Besle, J., Caclin, A., Mayet, R., Delpuech, C., Lecaigard, F., Giard, M.-H., & Morlet, D.
 747 (2007). Audiovisual Events in Sensory Memory. *Journal of Psychophysiology*, 21(3–
 748 4), 231–238. <https://doi.org/10.1027/0269-8803.21.34.231>
- 749 Besle, J., Fort, A., & Giard, M.-H. (2005). Is the auditory sensory memory sensitive to visual
 750 information? *Experimental Brain Research*, 166(3), 337–344.
 751 <https://doi.org/10.1007/s00221-005-2375-x>
- 752 Besle, J., Hussain, Z., Giard, M.-H., & Bertrand, O. (2013). The representation of audiovisual
 753 regularities in the human brain. *Journal of Cognitive Neuroscience*, 25(3), 365–373.
 754 https://doi.org/10.1162/jocn_a_00334
- 755 Bidelman, G. M. (2016). Musicians have enhanced audiovisual multisensory binding:
 756 experience-dependent effects in the double-flash illusion. *Experimental Brain*
 757 *Research*, 234(10), 3037–3047. <https://doi.org/10.1007/s00221-016-4705-6>
- 758 Bonath, B., Noesselt, T., Martinez, A., Mishra, J., Schwiecker, K., Heinze, H.-J., & Hillyard,
 759 S. A. (2007). Neural Basis of the Ventriloquist Illusion. *Current Biology*, 17(19),
 760 1697–1703. <https://doi.org/10.1016/j.cub.2007.08.050>
- 761 Boyle, S. C., Kayser, S. J., & Kayser, C. (2017). Neural correlates of multisensory reliability
 762 and perceptual weights emerge at early latencies during audio-visual integration.
 763 *European Journal of Neuroscience*, 46(10), 2565–2577.
- 764 Brosch, M., Selezneva, E., & Scheich, H. (2005). Nonauditory events of a behavioral
 765 procedure activate auditory cortex of highly trained monkeys. *The Journal of*

- 766 *Neuroscience : The Official Journal of the Society for Neuroscience*, 25(29), 6797–
767 6806.
- 768 Bulkin, D. A., & Groh, J. M. (2006). Seeing sounds: visual and auditory interactions in the
769 brain. *Current Opinion in Neurobiology*, 16(4), 415–419.
- 770 Caclin, A., Bouchet, P., Djoulah, F., Pirat, E., Pernier, J., & Giard, M. H. (2011). Auditory
771 enhancement of visual perception at threshold depends on visual abilities. *Brain*
772 *Research*, 1396(Journal Article), 35–44.
773 <https://doi.org/10.1016/j.brainres.2011.04.016>
- 774 Calvert, G. A., Brammer, M. J., Bullmore, E. T., Campbell, R., Iversen, S. D., & David, A. S.
775 (1999). Response amplification in sensory-specific cortices during crossmodal
776 binding. *Neuroreport*, 10(12), 2619–2623.
- 777 Colin, C., Radeau, M., Soquet, A., Dachy, B., & Deltenre, P. (2002). Electrophysiology of
778 spatial scene analysis: the mismatch negativity (MMN) is sensitive to the
779 ventriloquism illusion. *Clinical Neurophysiology : Official Journal of the*
780 *International Federation of Clinical Neurophysiology*, 113(4), 507–518.
- 781 Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F., & Deltenre, P. (2002). Mismatch
782 negativity evoked by the McGurk-MacDonald effect: a phonetic representation within
783 short-term memory. *Clinical Neurophysiology : Official Journal of the International*
784 *Federation of Clinical Neurophysiology*, 113(4), 495–506.
- 785 Deroy, O., & Spence, C. (2016). Crossmodal Correspondences: Four Challenges.
786 *Multisensory Research*, 29(1–3), 29–48.
- 787 Driver, J., & Noesselt, T. (2008). Multisensory Interplay Reveals Crossmodal Influences on
788 Sensory-Specific Brain Regions, Neural Responses, and Judgments. *Neuron*, 57(1),
789 11–23. <https://doi.org/10.1016/j.neuron.2007.12.013>

- 790 Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a
791 statistically optimal fashion. *Nature*, 415(6870), 429–433.
792 <https://doi.org/10.1038/415429a>
- 793 Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and
794 auditory features. *Journal of Vision*, 10(1), 6.1-6. <https://doi.org/10.1167/10.1.6>
- 795 Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the
796 primate cerebral cortex. *Cerebral Cortex (New York, N.Y.: 1991)*, 1(1), 1–47.
- 797 Foxton, J. M., Weisz, N., Bauchet-Lecaigard, F., Delpuech, C., & Bertrand, O. (2009). The
798 neural bases underlying pitch processing difficulties. *NeuroImage*, 45(4), 1305–1313.
799 <https://doi.org/10.1016/j.neuroimage.2008.10.068>
- 800 Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal*
801 *Society of London. Series B, Biological Sciences*, 360(1456), 815–836.
- 802 Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded
803 classification of visual size. *Perception & Psychophysics*, 68(7), 1191–1203.
- 804 Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends*
805 *in Cognitive Sciences*, 10(6), 278–285. <https://doi.org/10.1016/j.tics.2006.04.008>
- 806 Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object
807 recognition in humans: a behavioral and electrophysiological study. *Journal of*
808 *Cognitive Neuroscience*, 11(5), 473–490.
- 809 Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., ...
810 Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in*
811 *Neuroscience*, 7(Journal Article), 267. <https://doi.org/10.3389/fnins.2013.00267>
- 812 Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., ...
813 Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data.

- 814 *NeuroImage*, 86(Journal Article), 446–460.
- 815 <https://doi.org/10.1016/j.neuroimage.2013.10.027>
- 816 Herholz, S. C., Lappe, C., & Pantev, C. (2009). Looking for a pattern: an MEG study on the
- 817 abstract mismatch negativity in musicians and nonmusicians. *BMC Neuroscience*,
- 818 10(1), 42.
- 819 Jäncke, L., Rogenmoser, L., Meyer, M., & Elmer, S. (2012). Pre-attentive modulation of
- 820 brain responses to tones in coloured-hearing synesthetes. *BMC Neuroscience*, 13(1),
- 821 151–151. <https://doi.org/10.1186/1471-2202-13-151>
- 822 Jaramillo, M., Paavilainen, P., & Näätänen, R. (2000). Mismatch negativity and behavioural
- 823 discrimination in humans as a function of the magnitude of change in sound duration.
- 824 *Neuroscience Letters*, 290(2), 101–104. [https://doi.org/10.1016/S0304-](https://doi.org/10.1016/S0304-3940(00)01344-6)
- 825 3940(00)01344-6
- 826 Kislyuk, D. S., Möttönen, R., & Sams, M. (2008). Visual processing affects the neural basis
- 827 of auditory discrimination. *Journal of Cognitive Neuroscience*, 20(12), 2175–2184.
- 828 <https://doi.org/10.1162/jocn.2008.20152>
- 829 Klapetek, A., Ngo, M. K., & Spence, C. (2012). Does crossmodal correspondence modulate
- 830 the facilitatory effect of auditory cues on visual search? *Attention, Perception, &*
- 831 *Psychophysics*, 74(6), 1154–1167. <https://doi.org/10.3758/s13414-012-0317-9>
- 832 Kuchenbuch, A., Paraskevopoulos, E., Herholz, S. C., & Pantev, C. (2012). Electromagnetic
- 833 correlates of musical expertise in processing of tone patterns. *PloS One*, 7(1), e30171.
- 834 <https://doi.org/10.1371/journal.pone.0030171>
- 835 Lakatos, P., Chen, C. M., O’Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal
- 836 oscillations and multisensory interaction in primary auditory cortex. *Neuron*, 53(2),
- 837 279–292.

- 838 Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004).
839 Semantic congruence is a critical factor in multisensory behavioral performance.
840 *Experimental Brain Research*, 158(4), 405–414. <https://doi.org/10.1007/s00221-004->
841 1913-2
- 842 Lee Masson, H., Bulthé, J., Op de Beeck, H. P., & Wallraven, C. (2016). Visual and Haptic
843 Shape Processing in the Human Brain: Unisensory Processing, Multisensory
844 Convergence, and Top-Down Influences. *Cerebral Cortex (New York, N.Y.: 1991)*,
845 26(8), 3402–3412. <https://doi.org/10.1093/cercor/bhv170>
- 846 Levänen, S., Ahonen, A., Hari, R., McEvoy, L., & Sams, M. (1996). Deviant auditory stimuli
847 activate human left and right auditory cortex differently. *Cerebral Cortex*, 6(2), 288–
848 296. <https://doi.org/10.1093/cercor/6.2.288>
- 849 Lewkowicz, D. J., & Ghazanfar, A. A. (2009). The emergence of multisensory systems
850 through perceptual narrowing. *Trends in Cognitive Sciences*, 13(11), 470–478.
851 <https://doi.org/10.1016/j.tics.2009.08.004>
- 852 Luck, S. J. (2005). *An introduction to the event-related potential technique*. In *Cognitive*
853 *Neuroscience: Vol. Book, Whole*. Cambridge, Mass.: MIT Press.
- 854 Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP
855 experiment (and why you shouldn't). *Psychophysiology*, 54(1), 146–157.
856 <https://doi.org/10.1111/psyp.12639>
- 857 Maddox, R. K., Atilgan, H., Bizley, J. K., & Lee, A. K. C. (2015). Auditory selective
858 attention is enhanced by a task-irrelevant temporally coherent visual stimulus in
859 human listeners. *ELife*, 4(Journal Article). <https://doi.org/10.7554/eLife.04995>
- 860 Marks, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded
861 discrimination. *Journal of Experimental Psychology: Human Perception and*
862 *Performance*, 13(3), 384.

- 863 McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588),
864 746.
- 865 Melara, R. D. (1989). Dimensional interaction between color and pitch. *Journal of*
866 *Experimental Psychology: Human Perception and Performance*, 15(1), 69.
- 867 Melara, R. D., & O'Brien, T. P. (1987). Interaction between synesthetically corresponding
868 dimensions. *Journal of Experimental Psychology: General*, 116(4), 323–336.
869 <https://doi.org/10.1037/0096-3445.116.4.323>
- 870 Møller, C., Højlund, A., Bærentsen, K. B., Hansen, N. C., Skewes, J. C., & Vuust, P. (2018).
871 Visually induced gains in pitch discrimination: Linking audio-visual processing with
872 auditory abilities. *Attention, Perception, & Psychophysics*, 80(4), 999–1010.
873 <https://doi.org/10.3758/s13414-017-1481-8>
- 874 Möttönen, R., Krause, C. M., Tiippana, K., & Sams, M. (2002). Processing of changes in
875 visual speech in the human auditory cortex. *Cognitive Brain Research*, 13(3), 417–
876 425. [https://doi.org/10.1016/S0926-6410\(02\)00053-8](https://doi.org/10.1016/S0926-6410(02)00053-8)
- 877 Möttönen, R., Schürmann, M., & Sams, M. (2004). Time course of multisensory interactions
878 during audiovisual speech perception in humans: a magnetoencephalographic study.
879 *Neuroscience Letters*, 363(2), 112–115. <https://doi.org/10.1016/j.neulet.2004.03.076>
- 880 Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The Musicality of Non-
881 Musicians: An Index for Assessing Musical Sophistication in the General Population.
882 *PLoS ONE*, 9(2), e89642. <https://doi.org/10.1371/journal.pone.0089642>
- 883 Näätänen, R. (2008). Mismatch negativity (MMN) as an index of central auditory system
884 plasticity. *International Journal of Audiology*, 47(sup2), S16–S20.
- 885 Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity
886 (MMN) in basic research of central auditory processing: A review. *Clinical*
887 *Neurophysiology*, 118(12), 2544–2590. <https://doi.org/10.1016/j.clinph.2007.04.026>

- 888 Nichols, E. S., & Grahn, J. A. (2016). Neural correlates of audiovisual integration in music
889 reading. *Neuropsychologia*, (Journal Article).
890 <https://doi.org/10.1016/j.neuropsychologia.2016.08.011>
- 891 Noesselt, T., Tyll, S., Boehler, C. N., Budinger, E., Heinze, H.-J., & Driver, J. (2010). Sound-
892 induced enhancement of low-intensity vision: multisensory influences on human
893 sensory-specific cortices and thalamic bodies relate to perceptual enhancement of
894 visual detection sensitivity. *The Journal of Neuroscience: The Official Journal of the*
895 *Society for Neuroscience*, 30(41), 13609–13623.
896 <https://doi.org/10.1523/JNEUROSCI.4524-09.2010>
- 897 Novitski, N., Tervaniemi, M., Huotilainen, M., & Näätänen, R. (2004). Frequency
898 discrimination at different frequency levels as indexed by electrophysiological and
899 behavioral measures. *Cognitive Brain Research*, 20(1), 26–36.
- 900 Okada, K., Venezia, J. H., Matchin, W., Saberi, K., & Hickok, G. (2013). An fMRI Study of
901 Audiovisual Speech Perception Reveals Multisensory Interactions in Auditory Cortex.
902 *PloS One*, 8(6), e68959. <https://doi.org/10.1371/journal.pone.0068959>
- 903 Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source
904 software for advanced analysis of MEG, EEG, and invasive electrophysiological data.
905 *Computational Intelligence and Neuroscience*, 2011(Journal Article), 156869–
906 156869.
- 907 Orchard-Mills, E., Van der Burg, E., & Alais, D. (2015). Crossmodal Correspondence
908 Between Auditory Pitch and Visual Elevation Affects Temporal Ventriloquism.
909 *Perception*, 45(4), 409–424. <https://doi.org/10.1177/0301006615622320>
- 910 Paraskevopoulos, E., Kuchenbuch, A., Herholz, S. C., & Pantev, C. (2012a). Evidence for
911 Training-Induced Plasticity in Multisensory Brain Structures: An MEG Study. *PLoS*
912 *ONE*, 7(5), e36534. <https://doi.org/10.1371/journal.pone.0036534>

- 913 Paraskevopoulos, E., Kuchenbuch, A., Herholz, S. C., & Pantev, C. (2012b). Musical
914 expertise induces audiovisual integration of abstract congruency rules. *The Journal of*
915 *Neuroscience: The Official Journal of the Society for Neuroscience*, 32(50), 18196–
916 18203. <https://doi.org/10.1523/JNEUROSCI.1947-12.2012>
- 917 Parise, C. V., Knorre, K., & Ernst, M. O. (2014). Natural auditory scene statistics shapes
918 human spatial hearing. *Proceedings of the National Academy of Sciences*, 201322705.
919 <https://doi.org/10.1073/pnas.1322705111>
- 920 Parise, C. V., & Spence, C. (2009). “When birds of a feather flock together”: synesthetic
921 correspondences modulate audiovisual integration in non-synesthetes. *PloS One*, 4(5),
922 e5664. <https://doi.org/10.1371/journal.pone.0005664>
- 923 Parise, C. V., Spence, C., & Ernst, M. O. (2012). When Correlation Implies Causation in
924 Multisensory Integration. *Current Biology*, 22(1), 46–49.
925 <https://doi.org/10.1016/j.cub.2011.11.039>
- 926 Pavlov, I. P. (1927). *Conditioned reflexes*. Oxford: Oxford University Press.
- 927 Pekkola, J., Ojanen, V., Autti, T., Jääskeläinen, I. P., Möttönen, R., Tarkiainen, A., & Sams,
928 M. (2005). Primary auditory cortex activation by visual speech: an fMRI study at 3 T.
929 *Neuroreport*, 16(2), 125–128. <https://doi.org/10.1097/00001756-200502080-00010>
- 930 Pratt, C. C. (1930). The Spatial Character of High and Low Tones. *Journal of Experimental*
931 *Psychology*, 13(3), 278.
- 932 Proverbio, A. M., Massetti, G., Rizzi, E., & Zani, A. (2016). Skilled musicians are not subject
933 to the McGurk effect. *Scientific Reports*, 6(Journal Article), 30423.
934 <https://doi.org/10.1038/srep30423>
- 935 Saint-Amour, D., De Sanctis, P., Molholm, S., Ritter, W., & Foxe, J. J. (2007). Seeing voices:
936 High-density electrical mapping and source-analysis of the multisensory mismatch

- 937 negativity evoked during the McGurk illusion. *Neuropsychologia*, 45(3), 587–597.
- 938 <https://doi.org/10.1016/j.neuropsychologia.2006.03.036>
- 939 Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O. V., Lu, S., & Simola, J.
- 940 (1991). Seeing speech: visual information from lip movements modifies activity in the
- 941 human auditory cortex. *Neuroscience Letters*, 127(1), 141–145.
- 942 [https://doi.org/10.1016/0304-3940\(91\)90914-F](https://doi.org/10.1016/0304-3940(91)90914-F)
- 943 Sams, M., Paavilainen, P., Alho, K., & Näätänen, R. (1985). Auditory frequency
- 944 discrimination and event-related potentials. *Electroencephalography and Clinical*
- 945 *Neurophysiology/ Evoked Potentials Section*, 62(6), 437–448.
- 946 [https://doi.org/10.1016/0168-5597\(85\)90054-1](https://doi.org/10.1016/0168-5597(85)90054-1)
- 947 Schwartz, J.-L. (2010). A reanalysis of McGurk data suggests that audiovisual fusion in
- 948 speech perception is subject-dependent. *The Journal of the Acoustical Society of*
- 949 *America*, 127(3), 1584–1594.
- 950 Sekuler, R. (1997). Sound alters visual motion perception. *Nature*, 385(Journal Article), 308.
- 951 Senkowski, D., Saint-Amour, D., Hofle, M., & Foxe, J. J. (2011). Multisensory interactions
- 952 in early evoked brain activity follow the principle of inverse effectiveness.
- 953 *NeuroImage*, 56(4), 2200–2208. <https://doi.org/10.1016/j.neuroimage.2011.03.075>
- 954 Shams, L., Ma, W. J., & Beierholm, U. (2005). Sound-induced flash illusion as an optimal
- 955 percept. *Neuroreport*, 16(17), 1923–1927.
- 956 <https://doi.org/10.1097/01.wnr.0000187634.68504.bb>
- 957 Smith, E., Duede, S., Hanrahan, S., Davis, T., House, P., & Greger, B. (2013). Seeing is
- 958 believing: neural representations of visual stimuli in human auditory cortex correlate
- 959 with illusory auditory perceptions. *PloS One*, 8(9), e73148.
- 960 <https://doi.org/10.1371/journal.pone.0073148>

- 961 Spence, C. (2011). Crossmodal correspondences: a tutorial review. *Attention, Perception &*
962 *Psychophysics*, 73(4), 971–995. <https://doi.org/10.3758/s13414-010-0073-7>
- 963 Stefanics, G., Astikainen, P., & Czigler, I. (2015). Visual mismatch negativity (vMMN): a
964 prediction error signal in the visual modality. *Frontiers in Human Neuroscience*,
965 8(Journal Article), 1074. <https://doi.org/10.3389/fnhum.2014.01074>
- 966 Stein, B. E., & Meredith, M. A. (1993). *The Merging of the Senses*. Cambridge, Mass: MIT
967 Press.
- 968 Stein, Barry E., Stanford, T. R., & Rowland, B. A. (2014). Development of multisensory
969 integration from the perspective of the individual neuron. *Nature Reviews*
970 *Neuroscience*, 15(8), 520.
- 971 Stevenson, R. A., Bushmakina, M., Kim, S., Wallace, M. T., Puce, A., & James, T. W. (2012).
972 Inverse effectiveness and multisensory interactions in visual event-related potentials
973 with audiovisual speech. *Brain Topography*, 25(3), 308–326.
974 <https://doi.org/10.1007/s10548-012-0220-7>
- 975 Taulu, S., Kajola, M., & Simola, J. (2004). Suppression of Interference and Artifacts by the
976 Signal Space Separation Method. *Brain Topography*, 16(4), 269–275.
977 <https://doi.org/10.1023/B:BRAT.0000032864.93890.f9>
- 978 Taulu, S., & Simola, J. (2006). Spatiotemporal signal space separation method for rejecting
979 nearby interference in MEG measurements. *Physics in Medicine and Biology*, 51(7),
980 1759–1768. <https://doi.org/10.1088/0031-9155/51/7/008>
- 981 Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schroger, E. (2005). Pitch
982 discrimination accuracy in musicians vs nonmusicians: an event-related potential and
983 behavioral study. *Experimental Brain Research*, 161(1), 1–10.
984 <https://doi.org/10.1007/s00221-004-2044-5>

- 985 Tervaniemi, M., Kujala, A., Alho, K., Virtanen, J., Ilmoniemi, R. J., & Näätänen, R. (1999).
986 Functional specialization of the human auditory cortex in processing phonetic and
987 musical sounds: a magnetoencephalographic (MEG) study. *NeuroImage*, 9(3), 330–
988 336.
- 989 van Laarhoven, T., Stekelenburg, J. J., & Vroomen, J. (2017). Temporal and identity
990 prediction in visual-auditory events: Electrophysiological evidence from stimulus
991 omissions. *Brain Research*, 1661(Journal Article), 79–87.
992 <https://doi.org/10.1016/j.brainres.2017.02.014>
- 993 van Wassenhove, V., & Grzeczkowski, L. (2015). Visual-induced expectations modulate
994 auditory cortical responses. *Frontiers in Neuroscience*, 9(Journal Article), 11.
995 <https://doi.org/10.3389/fnins.2015.00011>
- 996 Vrba, J., & Robinson, S. (2002). SQUID sensor array configurations for
997 magnetoencephalography applications. *Superconductor Science and Technology*,
998 15(9), R51.
- 999 Vuust, P., Brattico, E., Seppanen, M., Naatanen, R., & Tervaniemi, M. (2012). The sound of
1000 music: differentiating musicians using a fast, musical multi-feature mismatch
1001 negativity paradigm. *Neuropsychologia*, 50(7), 1432–1443.
1002 <https://doi.org/10.1016/j.neuropsychologia.2012.02.028>
- 1003 Vuust, P., Pallesen, K. J., Bailey, C., van Zuijen, T. L., Gjedde, A., Roepstorff, A., &
1004 Østergaard, L. (2005). To musicians, the message is in the meter pre-attentive
1005 neuronal responses to incongruent rhythm are left-lateralized in musicians.
1006 *NeuroImage*, 24(2), 560. <https://doi.org/10.1016/j.neuroimage.2004.08.039>
- 1007 Walker, P. (2012). Cross-sensory correspondences and cross talk between dimensions of
1008 connotative meaning: Visual angularity is hard, high-pitched, and bright. *Attention*,
1009 *Perception*, & *Psychophysics*, 74(8), 1792–1809.

- 1010 Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C., & Vuust, P. (2010). The
 1011 Musical Ear Test, a new reliable test for measuring musical competence. *Learning*
 1012 *and Individual Differences*, 20(3), 188–196.
 1013 <https://doi.org/10.1016/j.lindif.2010.02.004>
- 1014 Williamson, V. J., Liu, F., Peryer, G., Grierson, M., & Stewart, L. (2012). Perception and
 1015 action de-coupling in congenital amusia: sensitivity to task demands.
 1016 *Neuropsychologia*, 50(1), 172–180.
 1017 <https://doi.org/10.1016/j.neuropsychologia.2011.11.015>
- 1018 Winkler, I., & Czigler, I. (2012). Evidence from auditory and visual event-related potential
 1019 (ERP) studies of deviance detection (MMN and vMMN) linking predictive coding
 1020 theories and perceptual object representations. *International Journal of*
 1021 *Psychophysiology: Official Journal of the International Organization of*
 1022 *Psychophysiology*, 83(2), 132–143. <https://doi.org/10.1016/j.ijpsycho.2011.10.001>
 1023