Binaural fusion involves weak interaural suppression

Baker, D.H.[1,4], Vilidaite, G.[1], McClarnon, E.[1], Valkova, E.[1] & Millman, R.E.[2,3]

[1.] Department of Psychology, University of York, Heslington, York, YO10 5DD, UK
[2.] Manchester Centre for Audiology and Deafness, Division of Human Communication, Hearing and Deafness, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester M13 9PL, UK
[3.] NIHR Manchester Biomedical Research Centre, Central Manchester University Hospitals Foundation Trust, Manchester Academic Health Science Centre, M13 9WL, UK.
[4.] email: daniel.baker@york.ac.uk

Abstract

The brain combines sounds from the two ears, but what is the algorithm used to achieve this fusion of signals? Here we take a model-driven approach to interpret both psychophysical increment detection thresholds and steady-state electrophysiology (EEG) data to reveal the architecture of binaural combination for amplitude modulated tones. Increment thresholds followed a ‘dipper’ shaped function of pedestal modulation depth, and were consistently lower for binaural than monaural presentation. The EEG responses were greater for binaural than monaural presentation, and when a modulated masker was presented to one ear, it produced only weak suppression of the signal presented to the other ear. Both data sets were well-fit by a computational model originally derived for visual signal combination, but with suppression between the two channels (ears) being much weaker than in binocular vision. We suggest that the distinct ecological constraints on vision and hearing can explain this difference, if it is assumed that the brain avoids over-representing sensory signals originating from a single object. These findings position our understanding of binaural summation in a broader context of work on sensory signal combination in the brain, and delineate the similarities and differences between vision and hearing.

Keywords: binaural fusion, EEG, amplitude modulation

Introduction

The auditory system integrates information across the two ears. This operation confers several benefits, including increased sensitivity to low intensity sounds [1] and inferring location and motion direction of sound sources based on interaural time differences [2]. In some animals, such as bats and dolphins, echolocation can be precise enough to permit navigation through the environment, and there are reports of visually impaired humans using a similar strategy [3,4], which requires both ears [5]. But what precisely is the algorithm that governs the combination of sounds across the ears? The nonlinearities inherent in sensory processing mean that simple linear signal addition is unlikely. This study uses complementary techniques (psychophysics, steady-state electroencephalography (EEG) and computational modelling) to probe the neural operations that underpin binaural fusion of amplitude modulated signals.

Classical psychophysical studies demonstrated that the threshold for detecting a very faint tone is lower when the tone is presented binaurally versus monaurally. Shaw et al. [1] presented signals to the two ears that were equated for each ear's individual threshold sound level when presented binaurally. This accounted for any differences in sensitivity (or audibility), and revealed that summation (the improvement in sensitivity afforded by binaural presentation) was approximately 3.6 dB (a factor of 1.5). Subsequent studies have provided similar or slightly lower values [i.e. 6–8], and there is general agreement that two ears are better than one at detection threshold [9]. This difference persists above threshold, with loudness discrimination performance being better binaurally than monaurally [10]. Furthermore, binaural sounds are perceived as being slightly louder than monaural sounds, though typically less than twice as loud [11–14].

When a carrier stimulus (typically either a pure-tone or broadband noise) is modulated in amplitude, neural oscillations at the modulation frequency can be detected at the scalp [15–18], being typically strongest at the vertex in EEG recordings [19]. This steady-state auditory evoked potential (SSAEP) is typically greatest around 40 Hz [19,20] and increases monotonically with modulation depth [17,18]. For low signal modulation frequencies (<55Hz), brain responses are thought to reflect cortical processes [15,19–21]. The SSAEP has been used to study binaural interactions, showing evidence of interaural suppression [22,23] and increased responses from binaurally fused stimuli [17,21].

The perception of amplitude-modulated stimuli shows similar properties to pure-tones in terms of binaural processing. For example, binaural sensitivity is better than monaural sensitivity [24,25], and the perceived modulation depth is approximately the average of the two monaural modulations over a wide range [26]. Presenting two different modulation frequencies to the left and right ears can produce the percept of a 'binaural beat' pattern at the difference intermodulation frequency (the highest minus the lowest frequency), suggesting that the two modulation frequencies are combined centrally [27]. Finally, increment detection of amplitude modulation [28] follows the "near miss to Weber's law" (i.e. Weber fractions for discrimination decrease as a function of pedestal level) typically reported for loudness discrimination [29]. However, despite these observations, detailed investigation and modelling of the binaural processing of amplitude-modulated tones is lacking.

Computational predictions for both psychophysical and electrophysiological results can be obtained from parallel work that considers the combination of visual signals across the left and right eyes. In several previous studies, a single model of binocular combination has been shown to successfully account for the pattern of results from psychophysical contrast discrimination and matching tasks [30,31], as well as steady-state EEG experiments [32]. The model, shown schematically in Figure 1a, takes contrast signals (sinusoidal modulations of luminance) from the left and right eyes, which mutually inhibit each other before being summed as follows:

$$resp = \frac{C_L^p}{Z^q + C_L^q + \omega C_R^q} + \frac{C_R^p}{Z^q + C_R^q + \omega C_L^q} \quad , \qquad (1)$$

where C_L and C_R are the contrast signals in the left and right eyes respectively, ω is the weight of interocular suppression, Z is a constant governing the sensitivity of the model, and p and q are exponents with the typical constraint that $p>q$. In all experiments in which the two signals have the same visual properties, the weight of interocular suppression (ω) has a value around 1.

Whereas vision studies typically modulate luminance relative to a mean background (DC) level (i.e. contrast), in hearing studies the amplitude modulation of a carrier waveform can be used to achieve the same effect. We can therefore test empirically whether binaural signal combination is governed by the same basic algorithm as binocular signal combination by replacing the C terms in equation 1 with modulation depths for amplitude modulated (AM) stimuli.

The response of the model for different combinations of inputs is shown in Figure 1b, with predictions being invariant to the sensory modality (hearing or vision). In the monaural/monocular ("mon") condition (blue), signals are presented to one channel only. In the binaural/binocular ("bin") condition (red) equal signals are presented to both channels. In the dichotic/dichoptic ("dich") condition (green) a signal is presented to one channel, with a fixed high amplitude 'masker' presented to the other channel throughout. For ω=1, the mon and bin conditions produce similar outputs, despite a doubling of the input (two channels vs one). This pattern of responses is consistent with the amplitudes recorded from steady-state visual evoked potential experiments testing binocular combination in humans [32].

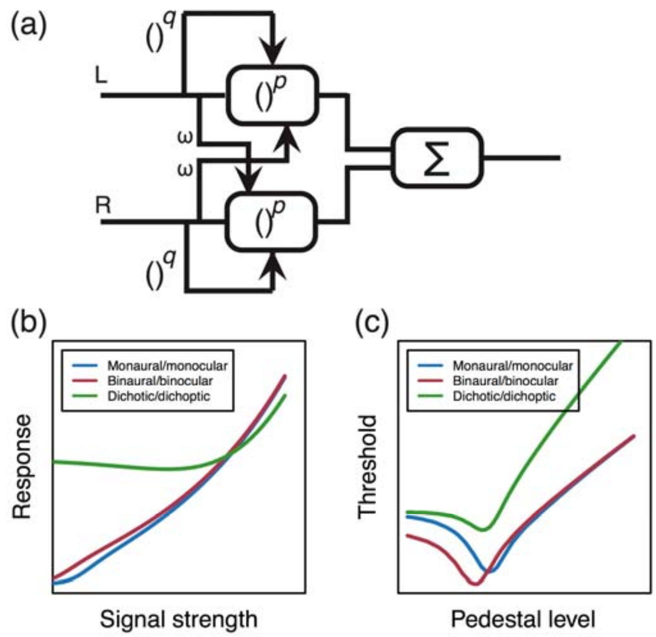

Figure 1: Schematic of signal combination model and qualitative predictions. Panel (a) shows a diagram of the signal combination model, which features weighted inhibition between left and right channels before signal combination (Σ). Panel (b) shows the predictions of this model for various combinations of inputs to the left and right channels, as described in the text. Predictions for discrimination of increases in modulation depth for similar conditions are shown in panel (c).

Figure 1c shows the gradients of the signal vs response functions from Figure 1b. These functions predict the results for psychophysical increment detection experiments in which thresholds are measured for discriminating changes in the level of a 'pedestal' stimulus. Such experiments measure the gradient because thresholds are defined as the horizontal translation required to produce a unit increase vertically along the functions in Figure 1b. The mon and bin functions converge at higher pedestal levels, and the dich function shows strong threshold elevation owing to the suppression between the two channels (when ω=1). Again, this pattern of functions is consistent with those reported in psychophysical studies of binocular vision [31].

The present study uses two complementary methods – psychophysical amplitude modulation depth discrimination, and steady-state auditory evoked potentials – to investigate binaural signal combination in the human brain. The results are compared with the predictions of the computational model [31,32] described above (see Figure 1) and modifications to the model are discussed in the context of functional constraints on the human auditory system. This principled, model-driven approach positions our understanding of binaural summation in a broader context of work on sensory signal combination in the brain.

Methods

Apparatus & stimuli

Auditory stimuli were presented over Sennheiser (HD 280 pro) headphones (Sennheiser electronic GmbH, Wedemark, Germany), and had an overall presentation level of 80dB SPL. An AudioFile device (Cambridge Research Systems Ltd., Kent, UK) was used to generate the stimuli at a sample rate of 44100Hz. Stimuli consisted of a 1-kHz pure-tone carrier, amplitude-modulated at a modulation frequency of either 40Hz or 35Hz (see Figure 2), according to the equation:

$$w = 0.5*(1 + m*\cos(f_m*t*2\pi + \pi)) * \sin(f_c*t*2\pi) \quad (2)$$

where f_m is the modulation frequency in Hz, f_c is the carrier frequency in Hz, t is time in milliseconds, and m is the modulation depth, scaled from 0-1 (though hereafter expressed as a percentage, 100*m). We chose not to compensate for overall stimulus power [as is often done for amplitude modulated stimuli, e.g. ,33] for two reasons. First, such compensation mostly affects performance at much higher modulation frequencies than we used here [e.g. see Figure A1 of 34]. Second, it makes implicit assumptions about the cues used by the participant in the experiment. We prefer to make such cues explicit in our computational modelling. The modulation depth and the assignments of modulation frequencies delivered to the left and right ears were varied parametrically across different conditions of the experiments.

EEG data were recorded with a sample frequency of 1 kHz using a 64-electrode Waveguard cap and an ANT Neuroscan (ANT Neuro, Netherlands) amplifier. Signals were digitised and stored on the hard drive of a PC for later offline analysis. Stimulus onset was coded on the EEG trace using low latency digital triggers.

Psychophysical procedures

In the psychophysics experiment, participants heard two amplitude-modulated stimuli presented sequentially using a two-alternative-forced-choice (2AFC) design. The stimulus duration was 500ms, with a 400ms interstimulus interval (ISI) and a minimum inter-trial interval of 500ms. One stimulus was the standard interval, consisting of the pedestal modulation depth only. The other stimulus was the signal interval, which comprised the pedestal modulation depth with an additional increment.

The presentation order of the standard and signal intervals was randomised, and participants were instructed to indicate the interval which they believed contained the target using a two-button mouse. A coloured square displayed on the computer screen indicated accuracy (green for correct, red for incorrect). The size of the target increment was determined by a pair of 3-down-1-up staircases, with a step size of 3 dB, which terminated after the lesser of 70 trials or 12 reversals. The percentage of correct trials at each target modulation depth was used to fit a cumulative log-Gaussian psychometric function to estimate the target modulation that yielded a performance level of 75% correct, which was defined as the threshold. Each participant completed three repetitions of the experiment, producing an average of 223 trials per condition (and an average of 7125 trials in total per participant).

Four binaural arrangements of target and pedestal were tested, at 8 pedestal modulation depths (*m* = 0, 1, 2, 4, 8, 16, 32 & 64%). The arrangements are illustrated schematically in Figure 2a, and were interleaved within a block at a single pedestal level, so that on each trial participants were not aware of the condition being tested. Note that in all conditions the carrier was presented to both ears, whether or not it was modulated by the pedestal and/or target. In the monaural condition, the pedestal and target modulations were presented to one ear, with the other ear hearing only the unmodulated carrier. The modulated stimulus was assigned randomly to an ear on each trial. In the binaural condition, the pedestal and target modulations were presented to both ears (in phase). In the dichotic condition, the pedestal modulation went to one ear and the target modulation to the other ear. Finally, in the half-binaural condition, the pedestal modulation was played to both ears, but the target modulation to only one ear. In all conditions, the modulation frequency for the pedestal and the target was 40Hz.

EEG procedure

In the EEG experiment, participants heard 11-s sequences of amplitude-modulated stimuli interspersed with silent periods of 3 seconds. There were five signal modulation depths (*m* = 6.25, 12.5, 25, 50 & 100%) and six binaural conditions, as illustrated in Figure 2b. In the first three conditions, a single modulation frequency (40 Hz, F1) was used. In the monaural condition, the modulated ‘signal’ tone was presented to one ear, and the unmodulated carrier was presented to the other ear. In the binaural condition, the signal modulation was presented to both ears. In the dichotic condition, the signal modulation was presented to one ear, and a modulated masker with a modulation depth of *m* = 50% was presented to the other ear. The remaining three conditions involved modulation at a second modulation frequency (35 Hz, F2). In the cross-monaural condition, F2 was presented to one ear as the signal, and the unmodulated carrier was presented to the other ear (F1 was not presented to either ear). In the cross-binaural condition, F1 was presented to one ear and F2 was presented to the other ear but the modulation depth of F1 and F2 was the same. In the cross-dichotic condition, F1 was presented to one ear, and F2 (*m* = 50%) was presented to the other ear. The order of conditions was randomised, and each condition was repeated ten times, counterbalancing the presentation of stimuli to the left and right ears as required.

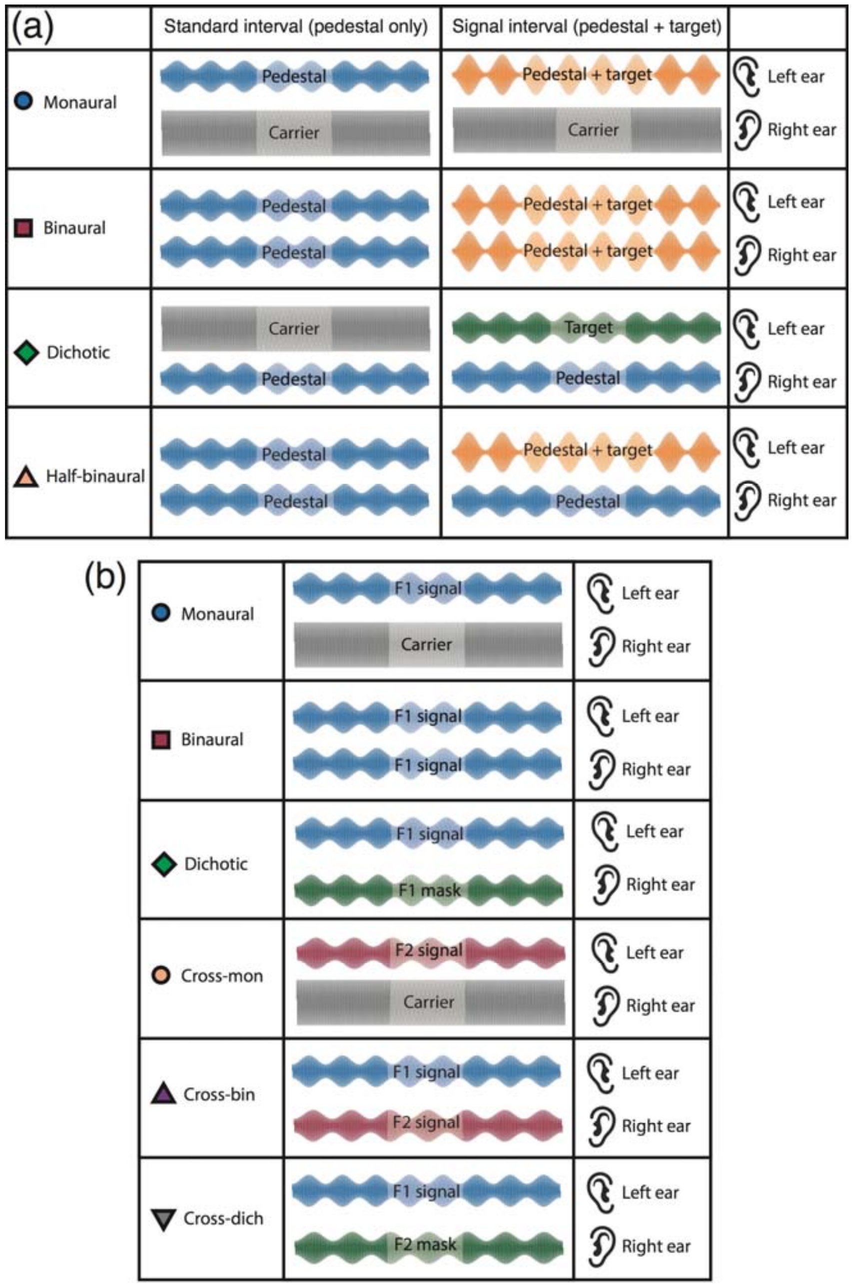

Figure 2: Summary of conditions and stimuli. Panel (a) illustrates the arrangement of pedestal and target modulations for the psychophysics experiment in the standard (pedestal only) interval (left) and signal (pedestal + target) interval (right) for four different interaural arrangements (rows). In all cases, the modulation frequency was 40 Hz and participants were asked to indicate the interval containing the target. A range of pedestal levels were tested, with target modulation depths determined by a staircase algorithm. Panel (b) shows stimulus arrangements for six conditions in the EEG experiment. Stimuli designated 'signal' had different modulation depths in different conditions, whereas stimuli designated 'masker' had a fixed modulation depth (m = 50%) at all signal levels. In all experiments stimulation was counterbalanced across the two ears, so the left ear/right ear assignments here are nominal.

EEG data for each trial at each electrode were then analysed offline. The first second following stimulus presentation was discarded to eliminate onset transients, and the remaining ten seconds were Fourier transformed using the fast Fourier transform function in Matlab (version 8.5, The MathWorks, Inc., Natick, MA). The dependent variables were the signal-to-noise ratios (SNR) at the Fourier components corresponding to the two modulation frequencies used in the experiment (40 Hz, F1 and 35 Hz, F2). This was calculated by dividing the amplitude at the frequency of interest (35 or 40 Hz) by the average amplitude in the surrounding 10 bins (±0.5 Hz in steps of 0.1 Hz). The SNRs for each of the ten repetitions of each condition were then averaged coherently (taking into account the phase angle). This coherent averaging procedure minimises noise contributions (which have random phase across repetitions), and previous studies [e.g. 32] have indicated that this renders artifact rejection procedures unnecessary. The absolute SNRs (discarding phase information) were then used to average across participants.

Participants

Three participants (one male) completed the psychophysics experiment, and twelve participants (3 male) completed the EEG experiment. All had self-reported normal hearing, and provided written informed consent. Experimental procedures were approved by the ethics committee of the Department of Psychology, University of York. Data are available online at: https://doi.org/10.6084/m9.figshare.5955904.

Results

Discrimination results are consistent with weak interaural suppression

The results of the amplitude modulation depth discrimination experiment are shown in Figure 3 for 3 participants (panels a-c). Although there were differences in absolute sensitivity between the participants (for example P2 has higher thresholds than P1 and P3), the overall pattern of thresholds was remarkably consistent across all three participants and is shown averaged in Figure 3d. A 4 (condition) x 8 (pedestal level) repeated measures ANOVA found significant main effects of condition ($F=10.31$, $p<0.01$, $\eta_p^2=0.84$) and pedestal level ($F=6.55$, $p<0.01$, $\eta_p^2=0.77$), and a significant interaction between the two factors ($F=2.46$, $p<0.01$, $\eta_p^2=0.55$).

When the results are plotted as thresholds on logarithmic axes, the results for binaurally presented modulations (red squares in Figure 3a-d) followed a ‘dipper’ shape [35], with thresholds decreasing from an average of around 4% at detection threshold to around 1% on a pedestal of 8% (a facilitation effect). At higher pedestal modulations, thresholds increased to around 16%, indicating a masking effect. Thresholds for the monaural modulation (blue circles in Figure 3a-d) followed a similar pattern, but were shifted vertically by an average factor of 1.77 across all pedestal levels. There was no evidence that monaural and binaural dipper handles converged at higher pedestal contrasts. At detection threshold (pedestal m=0), the average summation between binaural modulation and the three other conditions (which are identical in the absence of a pedestal) was a factor of 1.28 (2.14 dB). This level of summation is above that typically expected from probabilistic combination of independent inputs [36], and implies the presence of physiological summation between the ears.

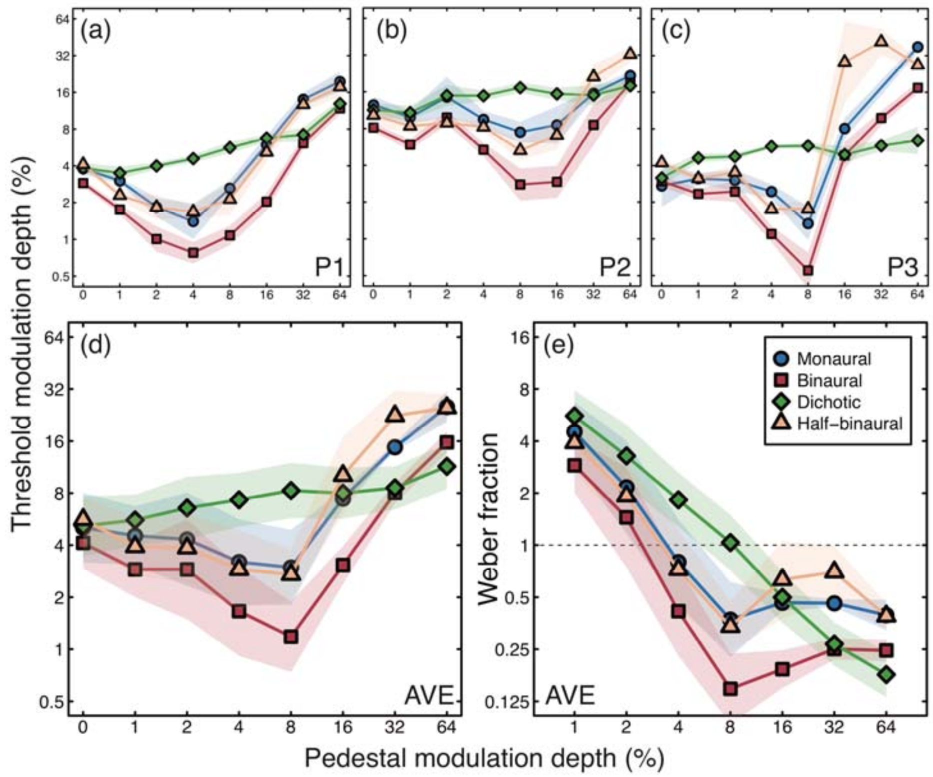

Figure 3: Results of the psychophysical AM depth discrimination experiment for three participants (panels a-c) and their average (panel d). Shaded regions give ±1 Standard Error (SE) of the Probit fit in panels a-c and ±1SE across participants in panel d. Arrangements of pedestal and target in different conditions were as illustrated in Figure 2a. It was not possible to measure a threshold for participant 3 in the monaural condition at 32% modulation depth, so this data point was omitted from panel c and when calculating the average in panel d. The data were replotted as Weber fractions in panel e by dividing each threshold by its accompanying pedestal modulation depth.

Dichotic presentation (pedestal modulation in one ear and target modulation in the other) elevated thresholds by a factor of 2.21 at the highest pedestal modulation depths (green diamonds in Figure 3a-d), compared to baseline (0% pedestal modulation). This masking effect was substantially weaker than is typically observed for dichoptic pedestal masking in vision (see Figure 1a), which can elevate thresholds by around a factor of 30 [31]. The half-binaural condition (orange triangles in Figure 3a-d), where the pedestal was presented to both ears, but the target only to one ear, was not appreciably different from the monaural condition, with thresholds greater than in the binaural condition by a factor of 1.86 on average.

These results can be converted to Weber fractions by dividing the threshold increments by the pedestal modulation depths, for pedestals >0%. These values are shown for the average data in Figure 3e. At lower pedestal modulation depths (<8%), Weber fractions decreased as a function of pedestal level. At pedestal modulations above 8%, the binaural Weber fractions (red squares) plateaued at around 0.25, whereas the monaural and half-binaural Weber fractions (blue circles and orange triangles) plateaued around 0.5. The dichotic Weber fractions (green diamonds) continued to decrease throughout. Thus, the “near miss to Weber’s law” behaviour occurred over the lower range of pedestal modulations depths, but more traditional Weber-like behaviour was evident at higher pedestal levels. The exception is the dichotic condition, where the "near miss" behaviour was evident throughout.

Overall, this pattern of results is consistent with a weak level of interaural suppression between the left and right ears. This accounts for the lack of convergence of monaural and binaural dipper handles at high pedestal levels, and the relatively minimal threshold elevation in the dichotic masking condition. Our second experiment sought to measure modulation response functions directly using steady-state EEG to test whether this weak suppression is also evident in cortical responses.

Direct neural measures of binaural combination

Steady-state EEG signals were evident over central regions of the scalp, at both modulation frequencies tested, and for both monaural and binaural modulations (Figure 4). In particular, there was no evidence of laterality effects for monaural presentation to one or other ear. We therefore averaged steady-state SNRs across a region-of-interest (ROI) comprising nine fronto-central electrodes (Fz, F1, F2, FCz, FC1, FC2, Cz, C1, C2, highlighted white in Figure 4) to calculate modulation response functions.

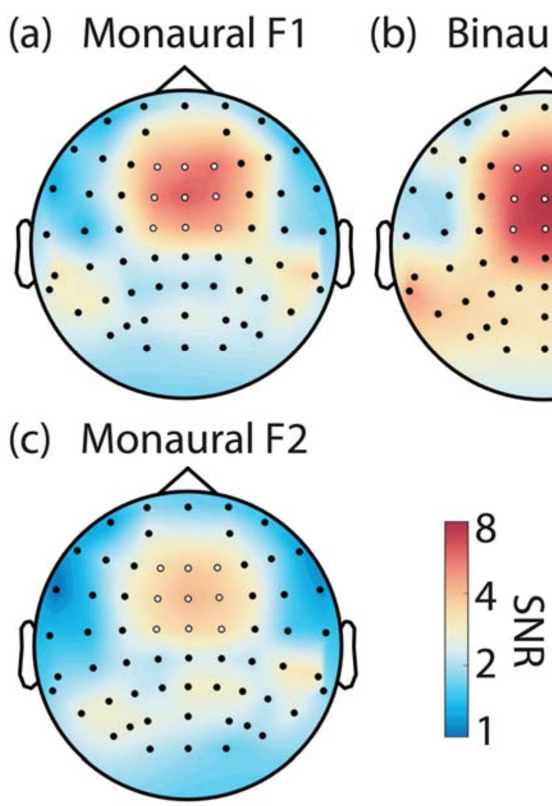

Figure 4: SNRs across the scalp at either 40 Hz (F1, panel a,b) or 35 Hz (F2, panel c). In panels a,c the signal modulation was presented to one ear (averaged across left and right), in panel b the modulation was presented to both ears. Note the log-scaling of the colour map. Dots indicate electrode locations, with those filled white showing the 9 electrodes that comprised the region-of-interest (ROI) used for subsequent analyses.

We conducted separate 6 (condition) x 5 (modulation depth) repeated measures ANOVAs at each modulation frequency using the SNRs averaged across the ROI. At 40Hz, we found significant main effects of condition (F=28.21, p<0.001, η_p^2=0.72, GG corrected) and modulation depth (F=29.04, p<0.001, η_p^2=0.73, GG corrected), and a significant interaction between the two variables (F=6.96, p<0.001, η_p^2=0.39). At 35Hz, we also found significant main effects of condition (F=14.87, p<0.01, η_p^2=0.58, GG corrected) and modulation depth (F=9.27, p<0.01, η_p^2=0.46, GG corrected), and a significant interaction between the two variables (F=5.97, p<0.001, η_p^2=0.35).

SNRs are plotted as a function of modulation depth in Figure 5. For a single modulation frequency (40 Hz), responses increased monotonically with increasing modulation depth, with SNRs >2 evident for modulation depths above 12.5%. Binaural presentation (red squares in Figure 5a) achieved SNRs of around 7 at the highest modulation depth, whereas monaural modulation produced weaker signals of SNR~5 (blue circles in Figure 5a). Assuming a nominal baseline of SNR=1 in the absence of any signal, this represents a binaural increase in response of a factor of 1.5. In the dichotic condition (green diamonds in Figure 5a), a masker with a fixed 50% modulation depth presented to one ear produced an SNR of 4 when the unmodulated carrier was presented to the other ear (see left-most point). As the dichotic signal modulation increased, responses increased to match the binaural condition at higher signal modulations (red squares and green diamonds in Figure 5a).

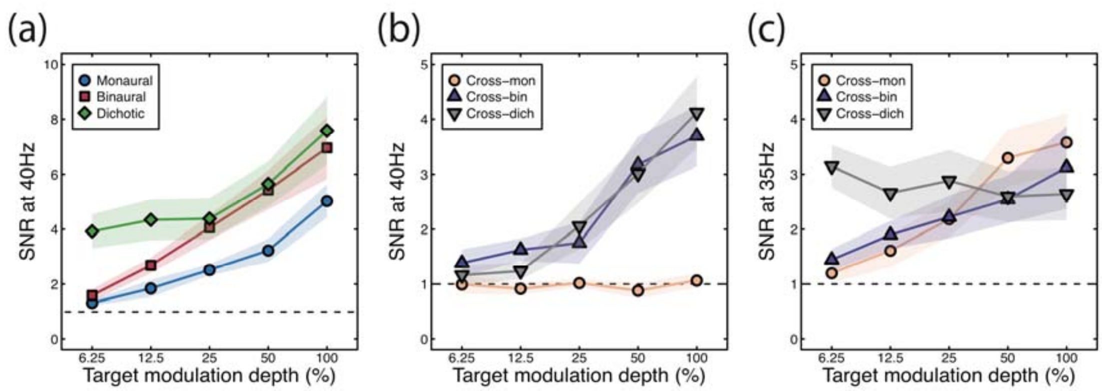

Figure 5: SNRs expressed as a function of signal modulation depth for six conditions at two frequencies. Shaded regions give ±1SE of the mean across participants (N=12). The dashed horizontal line in each plot indicates the nominal baseline of SNR=1.

When the carrier presented to one ear was modulated at a different frequency (35 Hz), several differences were apparent for the three conditions. Monaural modulation at 35 Hz (the cross-mon condition) evoked no measureable responses at 40Hz as expected (orange circles in Figure 5b). At the modulation frequency of 35Hz, this condition produced a monotonically increasing function peaking around SNR=4 (orange circles in Figure 5c). Binaural modulation with different modulation frequencies in each ear led to weaker responses (SNRs of 4 at 40 Hz and 3 at 35 Hz; purple triangles in Figure 5b,c) than for binaural modulation at the same frequency (SNR=8, red squares in Figure 5a). A 35-Hz AM masker with a fixed 50% modulation depth presented to one ear produced little change in the response to a signal in the other ear, which was amplitude-modulated with a modulation frequency of 40Hz (grey inverted triangles in Figure 5b), though increasing the signal modulation depth slightly reduced the neural response to the 35-Hz AM masker (grey inverted triangles in Figure 5c).

To summarise the EEG results, there is again evidence for weak interaural suppression owing to the non-overlapping monaural and binaural functions in Figure 5a, and the relatively modest masking effect in the cross-dichotic condition (grey inverted triangles in Figure 5b,c). We now consider model arrangements that are able to explain these results.

A single model of signal combination predicts psychophysics and EEG results

To further understand our results, we fit the model described by equation 1 to both data sets. To fit the psychophysical data, we calculated the target modulation depth that was necessary to increase the model response by a fixed value, σ_{40}, which was a fifth free parameter in the model (the other four free parameters being p, q, Z and ω; note that all parameters were constrained to be positive, q was constrained to always be greater than 2 to ensure that the nonlinearity was strong enough to produce a dip, and we ensured that $p>q$). With five free parameters, the data were described extremely well (see Figure 6a), with a root mean square error (RMSE, calculated as the square root of the mean squared error between model and data points across all conditions displayed in a figure panel) of 1.2 dB, which compares favourably to equivalent model fits in vision experiments [31]. However, the value of the interaural suppression parameter was much less than 1 (ω=0.02, see Table 1). This weak interaural suppression changes the behaviour of the model shown in Figure 1c in two important ways, both of which are consistent with our empirical results. First, the degree of threshold elevation in the dichotic condition is much weaker, as is clear in the data (green diamonds in Figure 3, 6a). Second, the thresholds in the monaural condition are consistently higher than those in the binaural condition, even at high pedestal levels (compare blue circles and red squares in Figure 3, 6a).

To illustrate how the model behaves with stronger interaural suppression, we increased the weight to a value of ω=1, but left the other parameters fixed at the values from the previous fit. This manipulation (shown in Figure 6b) reversed the changes caused by the weaker suppression – masking became stronger in the dichotic condition, and the monaural and binaural dipper handles converged at the higher pedestal levels. These changes provided a poorer description of human discrimination performance, with the RMSE increasing from 1.2 dB to 5.5 dB. Finally, we held suppression constant (at ω=1), but permitted the other four parameters to vary in the fit. This somewhat improved the fit (see Figure 6c), but retained the qualitative shortcomings associated with strong interaural suppression, and only slightly improved the RMSE (from 5.5 dB to 4.6dB).

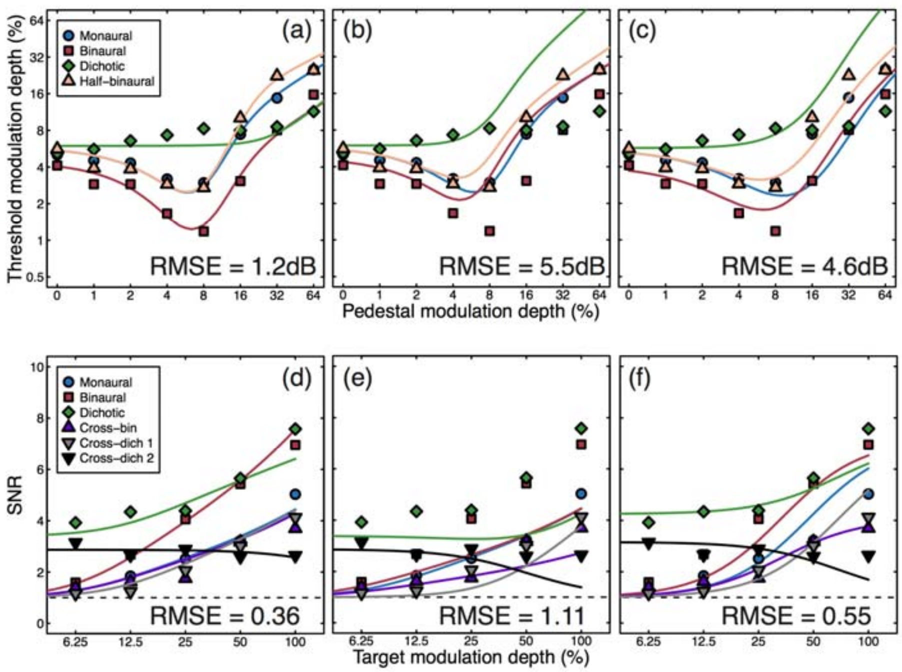

Figure 6: Fits of several variants of the signal combination model (curves) to empirical data (symbols), as described in the text. The data in the top panels are the averaged dipper functions duplicated from panel 3d, and those in the lower row are collapsed across the three panels of Figure 5 (omitting the mon-cross condition). Values in the lower right of each plot give the root mean square error (RMSE) of the fit in logarithmic (dB) units in the upper panels, and in units of SNR in the lower panels.

Table 1: Parameters for the model fits shown in Figure 6 with parameter constraints as described in the text.

Panel	p	q	Z	σ_{40}	σ_{35}	ω	RMSE
6a	3.28	2.90	7.66	0.63	-	0.02	1.2dB
6b	3.28	2.90	7.66	0.63	-	1	5.5dB
6c	2.23	2.00	14.30	0.20	-	1	4.6dB
6d	2.49	2.00	9.39	2.73	3.50	0.05	0.36
6e	2.49	2.00	9.39	2.73	3.50	1	1.11
6f	2.00	2.00	47.26	0.16	0.24	1	0.55

To fit the EEG data, we converted the model response to an SNR by adding and dividing by the noise parameter (σ). Because maximum SNRs varied slightly across the two modulation frequencies (40 and 35Hz, see Figure 5), we permitted this noise parameter to take a different value at each frequency (σ_{40} and σ_{35}). Model predictions for the conditions described in Figure 2c are shown in Figure 6d for a version of the model with six free parameters. This produced an excellent fit [comparable to those for visual signals, see 32], which included the main qualitative features of the empirical amplitude response functions, with an RMSE of 0.36. The model captures the increased response to binaural modulations compared with monaural modulations (blue circles vs red squares in Figure 6d), the relatively modest suppression in the cross-bin (purple triangles) and cross-dichotic (grey triangles) conditions at 40 Hz relative to the monaural condition, and the gentle decline in SNR in the cross dichotic condition at the masker frequency (black triangles in Figure 6d). Most parameters took on comparable values to those for the dipper function fits described above. Of particular note, the weight of interaural suppression remained weak (ω=0.05).

We again explored the effect of increasing the weight of suppression (to ω=1) whilst keeping other parameters unchanged. This resulted in a reduction of amplitudes in the binaural and cross-binaural conditions, which worsened the fit (to an RMSE of 1.11). Permitting all other parameters to vary freely improved the fit (to RMSE=0.55), but there were still numerous shortcomings. In particular the monaural and binaural response functions were more similar than in the data, and the reduction in SNR in the cross-binaural and cross-dichoptic conditions was more extensive than found empirically.

Our modelling of the data from two experimental paradigms therefore support the empirical finding that interaural suppression is relatively weak (by more than an order of magnitude) compared with similar phenomena in vision (interocular suppression).

Discussion

We have presented converging evidence from two experimental paradigms (psychophysics and steady-state EEG) concerning the architecture of the human binaural auditory system. A single computational model, in which signals from the two ears inhibit each other weakly before being combined, provided the best description of data sets from both experiments. This model architecture originates from work on binocular vision, showing a commonality between these two sensory systems. We now discuss these results in the context of related empirical results, previous binaural models, and ecological constraints that differentially affect vision and hearing.

A unified framework for understanding binaural processing

Our psychophysical experiment replicates the classical finding [1,6–9] of slightly less than 3 dB of summation at detection threshold (here 2.14 dB) using amplitude-modulated stimuli. This is somewhat weaker than is typically reported for binocular summation of contrast in vision, where summation ratios of 3-6 dB are typical. One explanation for this is a stronger early nonlinearity before signal combination, because the amount of summation is given by $2^{1/m}$, where m is the sum of all exponents before signal combination [37]. Indeed, the best fitting model numerator exponent for the psychophyscial data was 3.28 (see Table 1), much higher than the value of around 1.3 often used in vision [31], but consistent with previous hearing studies [7].

Above threshold, this difference persisted, with monaural stimulation producing higher discrimination thresholds (Figure 3) and weaker EEG responses (Figures 4, 5) than binaural stimulation. This is consistent with previous EEG work [17,21], and also the finding that perceived loudness and modulation depth are higher for binaural than monaural presentation [11–14,26]. However, these auditory effects are dramatically different from the visual domain, where both discrimination performance and perceived contrast are invariant of the number of eyes stimulated [30]. We discuss possible reasons for this modality difference below.

Suppression between the ears has been measured previously with steady-state magnetoencephalography (MEG) using amplitude-modulated stimuli of different frequencies in the left and right ears [22,23]. This is equivalent to the cross-binaural and cross-dichotic conditions from the EEG experiment reported here, though we observed somewhat less suppression than in the MEG studies. This difference could be due to the hemispheric differences reported in both MEG studies, the source localisation technique they used, or differences in the modulation frequencies across studies.

Another widely-studied phenomenon that might involve suppression between the ears is the binaural masking level difference [BMLD; 38]. In this paradigm, a signal embedded in noise is detected more easily when either the signal or the noise in one ear is inverted in phase [39]. Contemporary explanations of this effect [40,41] invoke cross-correlation of binaural signals, but lack explicit inhibition between masker and test signals. However, more elaborate versions of the model described here include mechanisms tuned to opposite spatial phases of sine-wave grating stimuli [42], and a similar approach in the temporal domain might be capable of predicting BMLD effects. Alternatively, since the BMLD phenomenon involves segmentation of target and masker, it might be more akin to 'unmasking' effects that occur in vision when stimuli are presented in different depth planes [43,44]. Modelling such effects would likely require additional mechanisms representing different spatial locations, far beyond the scope of the architecture proposed here.

The model shares features with previous binaural models

An influential model of binaural loudness perception [45,46] has some architectural similarities to the model shown in Figure 1a. For example, recent implementations incorporate binaural inhibition across multiple timescales [45]. However this model was designed with a focus on explaining perception of loudness across a range of frequencies (and for inputs of arbitrary frequency content), rather than attempting to understand performance on tasks (i.e. increment detection thresholds) or the precise mapping between stimulus and cortical response (i.e. the amplitude response functions measured using steady-state EEG). At threshold it predicts minimal levels of binaural summation (~1dB) in line with probabilistic combination of inputs [46], but below that found experimentally. The model would therefore likely require modification (i.e. the inclusion of physiological summation and early nonlinearities) to explain the data here, though it is possible that such modifications could be successful, given the other similarities between the models.

Several previous neural models of binaural processing have focussed on excitatory and inhibitory processes of neurons in subcortical auditory structures such as the lateral superior olive. These models (reviewed in Colburn, 1996) are concerned with lateralised processing, in which interaural interactions are purely inhibitory, and so do not typically feature excitatory summation. However, models of inferior colliculus neurons do typically involve binaural summation, and have the same basic structure as the architecture shown in Figure 1a. In general these models are designed to explain responses to diotically asynchronous stimuli, and so typically feature asymmetric delays across the excitatory and inhibitory inputs from the two ears [e.g. 47]. Since a time delay is not a critical component of the divisive suppression on the denominator of equation 1, and because a mechanism with broad temporal tuning is equivalent to the envelope of many mechanisms with different delays, the architecture proposed here can be considered a generalised case of such models.

Ecological constraints on vision and hearing

This study reveals an important but subtle difference between hearing and vision – suppression between the ears is far weaker than suppression between the eyes. Why should this be so? In the visual domain, the brain attempts to construct a unitary percept of the visual environment from two overlapping inputs, termed binocular single vision. For weak signals (at detection threshold) it is beneficial to sum the two inputs to improve the signal-to-noise ratio. But above threshold, there is no advantage for a visual object to appear more intense when viewed with two eyes compared with one. The strong interocular suppression prevents this from occurring by normalizing the signals from the left and right eyes to achieve 'ocularity invariance' – the constancy of perception through one or both eyes [30].

In the human auditory system the ears are placed laterally, maximising the disparity between the signals received (and minimising overlap). This incurs benefits when determining the location of lateralised sound sources, though reporting the location of pure tone sources at the midline (i.e. directly in front or behind) is very poor [2]. Hearing a sound through both ears at once therefore does not necessarily provide information that it comes from a single object, and so the principle of invariance should not be applied (and interaural suppression should be weak). However other cues that signals come from a single auditory object (for example interaural time and level differences consistent with a common location) should result in strong suppression, and cues that signals come from multiple auditory objects should release that suppression. This is the essence of the BMLD effects discussed above – suppression is strongest when target and masker have the same phase offsets (consistent with a common source), and weakest when their phase offsets are different. The distinct constraints placed on the visual and auditory systems therefore result in different requirements, which are implemented in a common architecture by changing the weight of suppression between channels.

Conclusions

A combination of psychophysical and electrophysiological experiments, and computational modelling have converged on an architecture for the binaural fusion of amplitude-modulated tones. This architecture is identical to the way that visual signals are combined across the eyes, with the exception that the weight of suppression between the ears is weaker than that between the eyes. This is likely because the ecological constraints governing suppression of multiple sources aim to avoid signals from a common source being over-represented. Such a high level of consistency across sensory modalities is unusual, and illustrates how the brain can adapt generic neural circuits to meet the demands of a specific situation.

References

1. Shaw WA, Newman EB, Hirsh IJ. 1947 The difference between monaural and binaural thresholds. *J. Exp. Psychol.* **37**, 229–242.

2. Rayleigh, Lord. 1907 XII. On our perception of sound direction. *Philos. Mag. Ser. 6* **13**, 214–232. (doi:10.1080/14786440709463595)

3. Kolarik AJ, Cirstea S, Pardhan S, Moore BCJ. 2014 A summary of research investigating echolocation abilities of blind and sighted humans. *Hear. Res.* **310**, 60–68. (doi:10.1016/j.heares.2014.01.010)

4. Thaler L, Arnott SR, Goodale MA. 2011 Neural Correlates of Natural Human Echolocation in Early and Late Blind Echolocation Experts. *PLoS ONE* **6**, e20162. (doi:10.1371/journal.pone.0020162)

5. Rowan D, Papadopoulos T, Edwards D, Allen R. 2015 Use of binaural and monaural cues to identify the lateral position of a virtual object using echoes. *Hear. Res.* **323**, 32–39. (doi:10.1016/j.heares.2015.01.012)

6. Babkoff H, Gombosh D. 1976 Monaural and binaural temporal integration of noise bursts. *Psychol. Res.* **39**, 137–145.

7. Heil P. 2014 Towards a Unifying Basis of Auditory Thresholds: Binaural Summation. *J. Assoc. Res. Otolaryngol.* **15**, 219–234. (doi:10.1007/s10162-013-0432-x)

8. Pollack I. 1948 Monaural and Binaural Threshold Sensitivity for Tones and for White Noise. *J. Acoust. Soc. Am.* **20**, 52–57. (doi:10.1121/1.1906347)

9. Hirsh IJ. 1948 Binaural summation; a century of investigation. *Psychol. Bull.* **45**, 193–206.

10. Jesteadt W, Wier CC. 1977 Comparison of monaural and binaural discrimination of intensity and frequency. *J. Acoust. Soc. Am.* **61**, 1599–1603. (doi:10.1121/1.381446)

11. Fletcher H, Munson WA. 1933 Loudness, Its Definition, Measurement and Calculation. *J. Acoust. Soc. Am.* **5**, 82–108. (doi:10.1121/1.1915637)

12. Hellman RP, Zwislocki J. 1963 Monaural Loudness Function at 1000 cps and Interaural Summation. *J. Acoust. Soc. Am.* **35**, 856–865. (doi:10.1121/1.1918619)

13. Reynolds GS, Stevens SS. 1960 Binaural Summation of Loudness. *J. Acoust. Soc. Am.* **32**, 1337–1344. (doi:10.1121/1.1907903)

14. Treisman M, Irwin RJ. 1967 Auditory Intensity Discriminal Scale I. Evidence Derived from Binaural Intensity Summation. *J. Acoust. Soc. Am.* **42**, 586–592. (doi:10.1121/1.1910626)

15. Mäkelä J., Hari R. 1987 Evidence for cortical origin of the 40 Hz auditory evoked response in man. *Electroencephalogr. Clin. Neurophysiol.* **66**, 539–546. (doi:10.1016/0013-4694(87)90101-5)

16. Picton TW, Vajsar J, Rodriguez R, Campbell KB. 1987 Reliability estimates for steady-state evoked potentials. *Electroencephalogr. Clin. Neurophysiol. Potentials Sect.* **68**, 119–131. (doi:10.1016/0168-5597(87)90039-6)

17. Rees A, Green GG, Kay RH. 1986 Steady-state evoked responses to sinusoidally amplitude-modulated sounds recorded in man. *Hear. Res.* **23**, 123–133.

18. Ross B, Borgmann C, Draganova R, Roberts LE, Pantev C. 2000 A high-precision magnetoencephalographic study of human auditory steady-state responses to amplitude-modulated tones. *J. Acoust. Soc. Am.* **108**, 679–691. (doi:10.1121/1.429600)

19. Kuwada S, Batra R, Maher VL. 1986 Scalp potentials of normal and hearing-impaired subjects in response to sinusoidally amplitude-modulated tones. *Hear. Res.* **21**, 179–192.

20. Farahani ED, Goossens T, Wouters J, van Wieringen A. 2017 Spatiotemporal reconstruction of auditory steady-state responses to acoustic amplitude modulations: Potential sources beyond the auditory pathway. *NeuroImage* **148**, 240–253. (doi:10.1016/j.neuroimage.2017.01.032)

21. Poelmans H, Luts H, Vandermosten M, Ghesquière P, Wouters J. 2012 Hemispheric Asymmetry of Auditory Steady-State Responses to Monaural and Diotic Stimulation. *J. Assoc. Res. Otolaryngol.* **13**, 867–876. (doi:10.1007/s10162-012-0348-x)

22. Fujiki N, Jousmaki V, Hari R. 2002 Neuromagnetic responses to frequency-tagged sounds: a new method to follow inputs from each ear to the human auditory cortex during binaural hearing. *J. Neurosci. Off. J. Soc. Neurosci.* **22**, RC205.

23. Kaneko K, Fujiki N, Hari R. 2003 Binaural interaction in the human auditory cortex revealed by neuromagnetic frequency tagging: no effect of stimulus intensity. *Hear. Res.* **183**, 1–6. (doi:10.1016/S0378-5955(03)00186-2)

24. Danilenko L. 1969 Binaural hearing in non-stationary diffuse sound field (in German). *Kybernetik* **6**, 50–57.

25. Zahorik P, Kim DO, Kuwada S, Anderson PW, Brandewie E, Collecchia R, Srinivasan N. 2013 Amplitude modulation detection by human listeners in reverberant sound fields: Carrier bandwidth effects and binaural versus monaural comparison. pp. 050002–050002. (doi:10.1121/1.4733848)

26. Ozimek E, Konieczny J, Sone T. 2008 Binaural perception of the modulation depth of AM signals. *Hear. Res.* **235**, 125–133. (doi:10.1016/j.heares.2007.10.008)

27. Rutschmann J, Rubinstein L. 1965 Binaural Beats and Binaural Amplitude-Modulated Tones: Successive Comparison of Loudness Fluctuations. *J. Acoust. Soc. Am.* **38**, 759–768. (doi:10.1121/1.1909802)

28. Wojtczak M, Viemeister NF. 1999 Intensity discrimination and detection of amplitude modulation. *J. Acoust. Soc. Am.* **106**, 1917–1924. (doi:10.1121/1.427940)

29. McGill WJ, Goldberg JP. 1968 A study of the near-miss involving Weber's law and pure-tone intensity discrimination. *Percept. Psychophys.* **4**, 105–109. (doi:10.3758/BF03209518)

30. Baker DH, Meese TS, Georgeson MA. 2007 Binocular interaction: contrast matching and contrast discrimination are predicted by the same model. *Spat. Vis.* **20**, 397–413. (doi:10.1163/156856807781503622)

31. Meese TS, Georgeson MA, Baker DH. 2006 Binocular contrast vision at and above threshold. *J. Vis.* **6**, 7–7. (doi:10.1167/6.11.7)

32. Baker DH, Wade AR. 2017 Evidence for an Optimal Algorithm Underlying Signal Combination in Human Visual Cortex. *Cereb. Cortex* **27**, 254–264. (doi:10.1093/cercor/bhw395)

33. Ewert SD, Dau T. 2004 External and internal limitations in amplitude-modulation processing. *J. Acoust. Soc. Am.* **116**, 478–490. (doi:10.1121/1.1737399)

34. Viemeister NF. 1979 Temporal modulation transfer functions based upon modulation thresholds. *J. Acoust. Soc. Am.* **66**, 1364–1380. (doi:10.1121/1.383531)

35. Legge GE, Foley JM. 1980 Contrast masking in human vision. *J. Opt. Soc. Am.* **70**, 1458–1471.

36. Tyler CW, Chen CC. 2000 Signal detection theory in the 2AFC paradigm: attention, channel uncertainty and probability summation. *Vision Res.* **40**, 3121–3144.

37. Baker DH, Wallis SA, Georgeson MA, Meese TS. 2012 Nonlinearities in the binocular combination of luminance and contrast. *Vision Res.* **56**, 1–9. (doi:10.1016/j.visres.2012.01.008)

38. Webster FA. 1951 The Influence of Interaural Phase on Masked Thresholds I. The Role of Interaural Time-Deviation. *J. Acoust. Soc. Am.* **23**, 452–462. (doi:10.1121/1.1906787)

39. Hirsh IJ. 1948 The Influence of Interaural Phase on Interaural Summation and Inhibition. *J. Acoust. Soc. Am.* **20**, 536–544. (doi:10.1121/1.1906407)

40. Colburn HS. 1996 Computational Models of Binaural Processing. In *Auditory Computation* (eds HL Hawkins, TA McMullen, AN Popper, RR Fay), pp. 332–400. New York, NY: Springer New York. (doi:10.1007/978-1-4612-4070-9_8)

41. Gilbert HJ, Shackleton TM, Krumbholz K, Palmer AR. 2015 The Neural Substrate for Binaural Masking Level Differences in the Auditory Cortex. *J. Neurosci.* **35**, 209–220. (doi:10.1523/JNEUROSCI.1131-14.2015)

42. Georgeson MA, Wallis SA, Meese TS, Baker DH. 2016 Contrast and lustre: A model that accounts for eleven different forms of contrast discrimination in binocular vision. *Vision Res.* **129**, 98–118. (doi:10.1016/j.visres.2016.08.001)

43. Moraglia G, Schneider B. 1990 Effects of direction and magnitude of horizontal disparities on binocular unmasking. *Perception* **19**, 581–593. (doi:10.1068/p190581)

44. Wardle SG, Cass J, Brooks KR, Alais D. 2010 Breaking camouflage: Binocular disparity reduces contrast masking in natural images. *J. Vis.* **10**, 38–38. (doi:10.1167/10.14.38)

45. Moore BCJ, Glasberg BR, Varathanathan A, Schlittenlacher J. 2016 A Loudness Model for Time-Varying Sounds Incorporating Binaural Inhibition. *Trends Hear.* **20**, 233121651668269. (doi:10.1177/2331216516682698)

46. Moore BCJ, Glasberg BR. 2007 Modeling binaural loudness. *J. Acoust. Soc. Am.* **121**, 1604–1612. (doi:10.1121/1.2431331)

47. Sujaku Y, Kuwada S, Yin TCT. 1981 Binaural Interaction in the Cat Inferior Colliculus: Comparison of the Physiological Data with a Computer Simulated Model. In *Neuronal Mechanisms of Hearing* (eds J Syka, L Aitkin), pp. 233–238. Boston, MA: Springer US. (doi:10.1007/978-1-4684-3908-3_24)