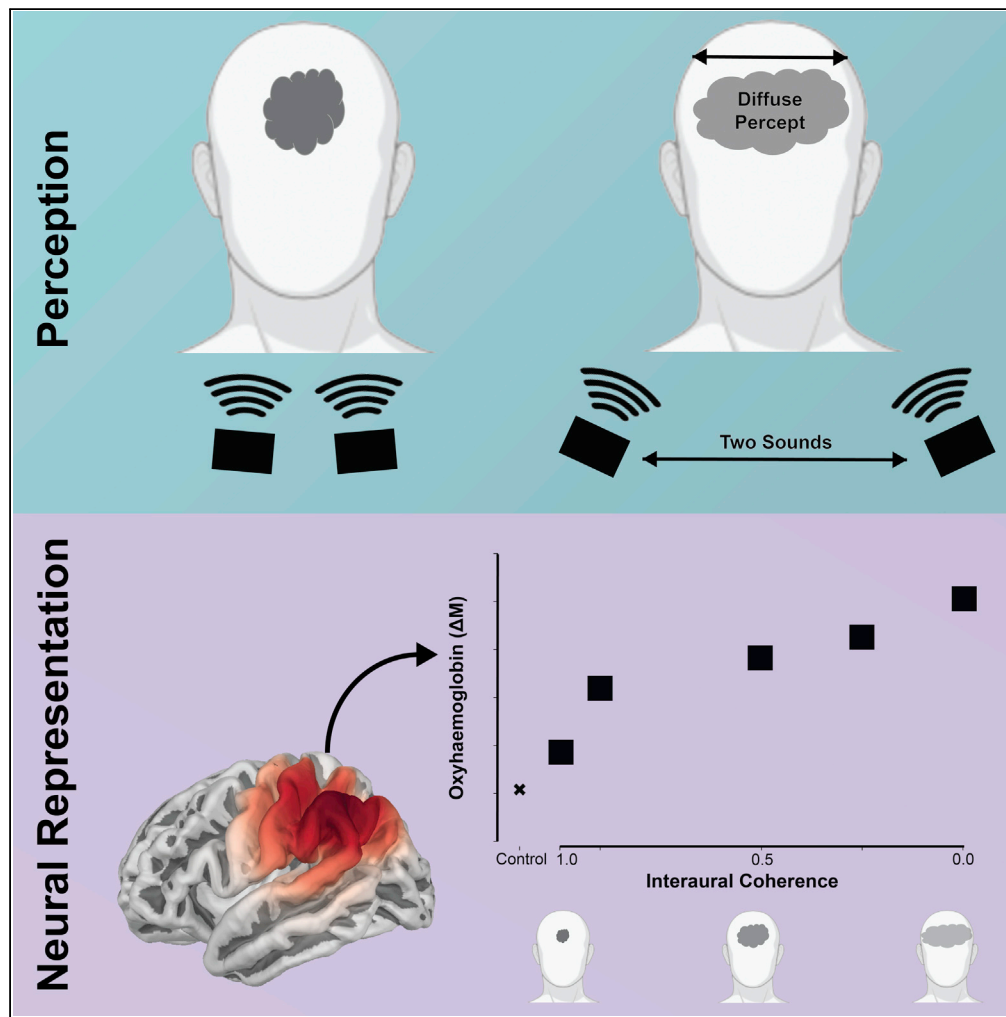**Article**

Human cortical processing of interaural coherence

Robert Luke,
Hamish Innes-
Brown, Jaime A.
Undurraga, David
McAlpine

robert.luke@mq.edu.au

Highlights

Listeners report
ambiguous spatial
percepts for sounds with
reduced interaural
coherence

Supra-ethological spatial
cues impose an additional
diffuse quality to sounds

A frequency-dependent
distribution of spatial
detectors explains
behavioral results

Sounds with low interaural
coherence impose
elevated cortical load

Article

Human cortical processing of interaural coherence

Robert Luke,[1,2,4,*] Hamish Innes-Brown,[3] Jaime A. Undurraga,[1] and David McAlpine[1]

SUMMARY

Sounds reach the ears as a mixture of energy generated by different sources. Listeners extract cues that distinguish different sources from one another, including how similar sounds arrive at the two ears, the interaural coherence (IAC). Here, we find listeners cannot reliably distinguish two completely interaurally coherent sounds from a single sound with reduced IAC. Pairs of sounds heard toward the front were readily confused with single sounds with high IAC, whereas those heard to the sides were confused with single sounds with low IAC. Sounds that hold supra-ethological spatial cues are perceived as more diffuse than can be accounted for by their IAC, and this is accounted for by a computational model comprising a restricted, and sound-frequency dependent, distribution of auditory-spatial detectors. We observed elevated cortical hemodynamic responses for sounds with low IAC, suggesting that the ambiguity elicited by sounds with low interaural similarity imposes elevated cortical load.

INTRODUCTION

Grouping sounds together and separating them from others underpins our ability to make sense of the world. Acoustic features such as harmonicity, common onset time, or co-modulation of amplitude or spectral cues all contribute to the ability to group sounds together, assign them to specific sources, and segregate them from others (Bregman, 1994). The similarity of sounds arriving at each ear—their interaural coherence—may also modulate the perception of auditory objects. Sounds with high interaural coherence (IAC) are generally heard as compact or focused whereas those with low IAC tend to be perceived as more diffuse. In quiet, non-reverberant rooms, most acoustic sources generate sounds with high IAC and are perceived as clearly defined auditory objects (Blauert and Lindemann, 1986; Griffiths and Warren, 2004). However, in noisy or highly reverberant rooms, reflections from hard surfaces and interference patterns from multiple sources reduce IAC and the perceived clarity of the sources generating those sounds. The degree to which sounds are interaurally coherent depends on the stability of binaural cues—differences in the timing and intensity of sounds at the two ears that underpin the ability to locate the sources of sounds along the horizontal plane (Grothe et al., 2010). Binaural cues that are stable over time are more likely to originate from distinct sources. Sounds with rapidly fluctuating binaural cues tend to arise from background features, including the physical properties of the listening space itself (Traer and McDermott, 2016). Listeners integrate these rapidly fluctuating cues—which can hold supra-ethological and incongruent values, i.e., values not possible for a direct sound path between source and listener and experienced for extended periods—to generate knowledge of the broader listening environment.

Commonly, sources exist in the physical environment and can generate sounds that propagate as acoustic waves. The physical properties of these sounds can be quantified using standardized metrics which, measured at one or both ears, are referred to as monaural and binaural cues, respectively. The transduction of the acoustic waveform at each ear elicits a neural representation from which the perception of an auditory image is generated, eliciting perceptual qualities that can be assessed behaviorally. The extent to which acoustic cues are congruent and sustained contributes to the perception of a stable auditory object, often externalized to beyond the head within the context of the physical environment (Best et al., 2020). To allow for precise control of sound properties and cues, however, sounds are often synthesized independent of the source, and presented directly to the ears via stereo headphones. Like real-world stimuli, these synthetic stimuli may elicit the percept of an acoustic image and can be systematically manipulated to modify their qualia. Here, we employ synthetically generated binaural stimuli to explore both the perception and neural representation of auditory spatial cues, including the interaural coherence (IAC)—the degree to which sounds are similar at the two ears. Interaural coherence can be quantified as the peak of the cross-correlation coefficient of the sounds at the two ears (Aaronson and Hartmann, 2010; Chait et al.,

[1]Macquarie University, Sydney, NSW, Australia

[2]The Bionics Institute, Melbourne, VIC, Australia

[3]Eriksholm Research Centre, Oticon A/S, Snekkersten, Denmark

[4]Lead contact

*Correspondence: robert.luke@mq.edu.au

https://doi.org/10.1016/j.isci.2022.104181

2005). In headphone listening, a fully coherent sound (IAC = 1.0) is perceived as a compact intracranial image located in toward the center of the head. The perceived width of the image broadens as IAC is reduced until, for highly incoherent sounds (IAC ≈ 0.0), it is reported as either spanning the head or as two distinct images, one toward either ear (Blauert and Lindemann, 1986; Hall et al., 2005). Sensitivity to changes in IAC is not uniform across the range, with human listeners an order of magnitude more sensitive to changes in IAC from fully correlated sounds than they are to changes from less correlated, or fully uncorrelated, sounds (Kaigham and Colburn, 1981; Pollack and Trittipoe, 1959). Listeners are also faster and more accurate at detecting rapid deviations from interaurally correlated than uncorrelated noise (Chait et al., 2005; McEvoy et al., 1991). This perceptual nonlinearity is consistent with the general framework of detecting changes from ordered to disordered states (Chait et al., 2005,2007), with the brain integrating information over time until a sufficient level of reliability is reached. Under this framework, common across sensory modalities (Julesz and Tyler, 1976), coherent sensory input reaches a certain level of reliability faster than incoherent input.

Changes in IAC elicit event-related potentials in auditory cortex that parallel behavioral detection and reaction times. Electrophysiological responses occur earlier for changes away from coherent compared to incoherent sounds, and larger changes in IAC evoke larger event-related potentials (Ando et al., 1987; Chait et al., 2005,2007; Jones et al., 1991; McEvoy et al., 1991; Soeta et al., 2004). Behavioral assays consider IAC in terms of a continuum, and electrophysiological and imaging studies assess it in terms of high and low/dis-ordered soundscapes. Nevertheless, listeners commonly report fully uncorrelated sounds as sounding like two intracranial images (Hall et al., 2005) if allowed to do so. The perception of uncorrelated sounds as comprising two images makes sense from the perspective of auditory spatial cues. In headphone listening, an IAC of zero is generated by presenting two completely independent sounds, one to each ear. In natural listening, two separate sources relatively proximate to the listener can generate interaural time and level differences (ITDs and ILDs; binaural cues for localizing sources on the horizontal plane) that, combined, would produce very low values of IAC. As such, a relationship exists between multiple acoustic objects and the perception of acoustic backgrounds.

Here, we explored the ability of human listeners to discriminate foreground from background sounds, based on sensitivity to IAC. In a headphone-listening task, listeners had to identify which of two stimuli was more similar to a diffuse background. One stimulus comprised two independent sounds originating from different locations (each with IAC = 1), and the other contained a single frontal sound with reduced IAC. Stimuli composed of single sounds with IAC near 1.0 were easily distinguishable from the diffuse background, but this was not the case for stimuli composed of two sounds positioned to either side. Furthermore, stimuli composed of two sounds positioned closer to the front were more readily confused with stimuli composed of a single sound with moderate IAC. Stimuli composed of two sounds that were positioned further away from the midline maintained their relative similarity to fully incoherent reference sounds. The increase in perceptual ambiguity of sounds positioned further away from the auditory midline is consistent with the increase in minimum audible angle—the minimum distance a source must traverse before it is perceived to hold a different location (Mills, 1958)—for sources located more laterally.

Headphone listening enables access to ITDs beyond the range generated by the distance between the ears on an average human head and the speed of sound (equating to approx. ± 680 µs) (Feddersen et al., 1957; Kuhn, 1977). We observe that sounds with ITDs beyond the width of the head are perceived as more similar to diffuse background than sounds with ITDs confined to the width of the head. This increase in perceptual ambiguity for ITDs beyond the head width is explicable in terms of an additional reduction in IAC relative to the IAC of the sound signal at the two ears. We modeled the difference in perceptual ambiguity generated by sounds holding ethological and supra-ethological values of ITD in terms of the range of internal delays instantiated in the brain, and found that delay taps up to approximately 1/2 the period of the stimulus center frequency—the so-called π-limit (McAlpine et al., 2001; Thompson et al., 2006)—are required to account for the difference. Extending the range of taps beyond the π-limit provides no additional explanatory power. This π-limited representation of ITDs is consistent with the neural representation of ITDs reported from *in vivo* recordings in a wide range of smaller mammals (Brand et al., 2002; Harper et al., 2014; Joris et al., 2005; McAlpine et al., 2001).

Finally, using functional near-infrared spectroscopy (fNIRS) neuroimaging to explore the cortical representation of IAC, we find that sounds with low IAC—generating an ambiguous spatial percept—utilize more

resources than sounds with high IAC. Specifically, the oxyhemoglobin-related hemodynamic response in *planum temporale* increases with decreasing IAC, suggesting that perceptual ambiguity is associated with an increase in computational load. Our data are consistent with the view that spatial information is processed in the dorsal pathway (Bizley and Cohen, 2013), and that *planum temporale* acts as a 'computation hub' requiring greater neural resources to form a stable perceptual representation when less-reliable sensory input is available (Griffiths and Warren, 2002; Kumar et al., 2007).

RESULTS

Listeners exploit interaural coherence to judge the similarity of single and multiple intracranial images to diffuse noise

In headphone listening, sounds with low IAC are described as having a broad and diffuse intracranial image percept, or are perceived as two separate, lateralized intracranial images, one toward either ear (Blauert and Lindemann, 1986; Hall et al., 2005). We asked whether listeners could distinguish stimuli composed of two, fully coherent auditory sounds from stimuli composed of a single sound with reduced IAC. Seven listeners (4 female) performed a repeated two-interval, forced choice task in which each interval contained a pair of stimuli. The first half of each interval—the reference sound—was always a fully incoherent noise (IAC = 0.0). The second half of each interval was either a single sound with zero ITD and variable IAC (in the range 1.0 to 0.0), or it comprised two simultaneous sounds with ITDs of equal magnitude but opposite sign (van der Heijden and Trahiotis, 1999). Each sound consisted of band-limited noise of 400-Hz bandwidth centered at 500 Hz. Stimuli were presented in blocks, with all combinations of IAC presented in a block. An experimental session contained multiple blocks with a single value of unsigned ITD for the two sounds (Figure 1A). Presentation order—whether the interval in which the second sound contained two sounds or was a noise with varied IAC—was randomized, and listeners were asked to "select the interval with the two sounds that were most similar".

We assessed listeners' abilities to distinguish single-sound stimuli with reduced IAC from stimuli composed of two sounds with non-zero ITDs. We first explored ITDs ± 250 µs, ± 375 µs and ± 500 µs (McAlpine et al., 2001; Salminen et al., 2018; Thompson et al., 2006; von Kriegstein et al., 2008), all of which lie within the range generated by the size of the head (Figures 1B and 1D). When the IAC of the sound was 1.0, listeners consistently selected the interval which contained two lateralized sounds as being that in which the pair of sound images was most similar. That is, they indicated that the stimulus with two sounds was more similar to the reference sound with IAC = 0.0 (Figure 1D right). As the IAC of the single-sound was reduced, listeners increasingly selected the interval containing the single-sound with reduced IAC as being more similar to the reference sound. A value of 50% indicates that listeners selected the two-sound interval and the single-sound interval with reduced IAC with equal probability, suggesting that both stimuli were equally similar to the reference sound with IAC = 0. Values below 50% indicate that listeners perceived the one-sound interval with reduced coherence to be more similar to the reference sound with IAC = 0.0 than they did the two-sound interval. Below a particular value of IAC, considerably greater than 0.0, listeners no longer consistently selected intervals composed of two sounds as being more similar to the fully incoherent sound. Surprisingly, even when the single-sound and reference sound in one interval were identical, i.e., both sounds had IAC = 0.0, listeners selected the two-sound interval as most similar to the reference in 30% of trials for the 250 Hz condition and 50% of trials in the 500 Hz condition.

The highest value of single-sound IAC at which listeners reported as being equally similar to the reference sound as the two-sound interval varied systematically as a function of ITD; ~0.6 IAC for ITDs of ± 250 µs; ~0.4 for ITDs of ± 375 µs; and ~0.2 for ITDs of ± 500 µs (Figure 1D). Listeners' selection at IAC of zero also varied systematically as a function of ITD of the two sounds (Figure 1D left). Specifically, for two-sound interval with the most central pair of sources (ITD = ± 250 µs) listeners more readily judged single-sound interval with reduced IAC to be similar to the fully incoherent reference than they did the two sounds, although it was still the case that on nearly 30% of trials listeners could not distinguish completely incoherent sounds from two sounds with ITDs of ± 250 µs. This ambiguity increased for sources with ITDs of ± 375 µs, whereas for the most-lateralized sources with ITDs of ± 500 µs listeners judged two sounds as being equally similar to the reference as the fully uncorrelated sound.

Our data are consistent with sounds positioned close to the midline emerging more readily from a diffuse background, and with listeners showing greater acuity when localizing and tracking auditory objects toward the midline (Carlile and Leung, 2016; Strybel et al., 1992).
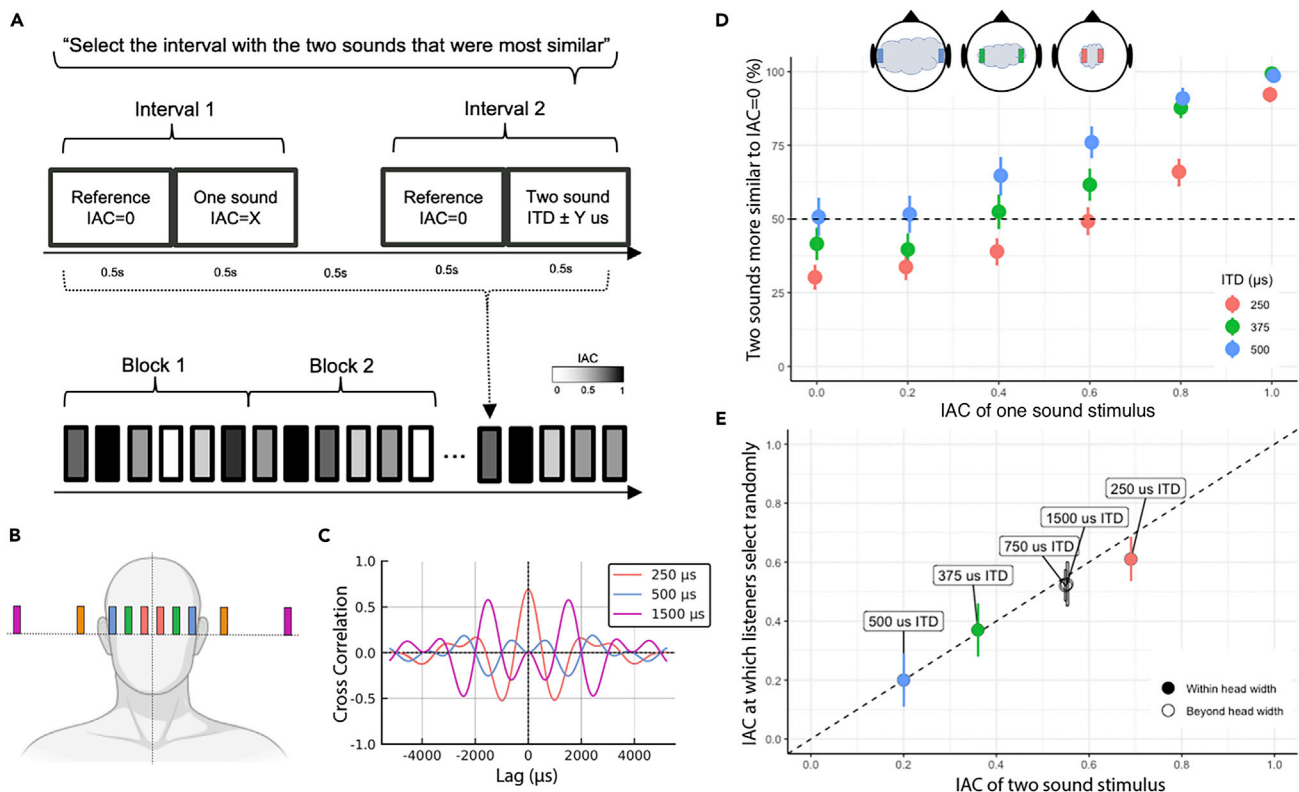
Figure 1. Interaural coherence describes the ability to distinguish diffuse noise from multiple sounds

(A). The two-interval, forced choice experimental paradigm. In this example the second stimulus in the second interval contained two sounds, each with an IAC = 1, but with equal and opposite ITD value.

(B). Illustration of the ITDs used for the two sound stimuli.

(C). Example cross correlation functions for the two sound stimuli.

(D). The percentage of trials for which listeners reported stimuli composed of two sounds being more similar to a single, incoherent noise than a single sound with reduced IAC. Listeners consistently perceive the two-sound stimuli as being more similar to incoherent noise when compared to a single sound with IAC = 1 (right). Listeners provided a less consistent response when presented with an interval containing a pair of incoherent sounds (left). The dashed line indicates chance selection in the task. Error bars represent 95% confidence intervals. The inset provides an example of how the three single-sound IAC values may be perceived by listeners, with the single-sound IAC condition represented by a central image (gray) that broadens with reducing IAC, and the location of the two sounds indicated with bars.

(E). Relation between IAC of the two-sound interval and the IAC value in the experiment at which listeners selected randomly between the two-sound condition and the single sound with reduced coherence. Error bars represent 95% confidence intervals.

Sounds reach each eardrum as a unitary waveform, with no explicit, separable information about source identities or locations generating them. The extent to which sounds are perceived as originating from distinct sources, then, relies on stimulus features that contribute to perceptual grouping and segregation (Bregman, 1994). Although it seems intuitive that spatial properties such as source location would function as potent grouping cues, little evidence exists to support this contention, and neural mechanisms underpinning grouping and segregation remain poorly understood, including for well-established cues such as harmonicity. Because ITD information is derived several synaptic stages beyond the inner ear, the independently generated sources in our two-sound stimuli will first be combined as a single waveform at each eardrum before binaural processing. To this end, we assessed the extent to which the combined IAC of the sounds might explain listeners' preferences when judging the relative similarity of stimuli composed of two sounds to stimuli comprising a single, fully incoherent sound. To do so, we calculated the IAC—defined as the peak of the cross-correlation function of the signals at the two ears (Figure 1C)—of two-sound stimulus and partially coherent sounds judged to be perceptually similar. We found this to be highly predictable based on the IAC of the signals [Pearson correlation, $r(3) = 0.99$, $p < 0.001$. Figure 1E], indicating that listeners can exploit interaural differences, quantified in this study using the IAC metric, as a feature in their judgment to differentiate a single sound with reduced coherence from stimuli consisting of multiple sounds.
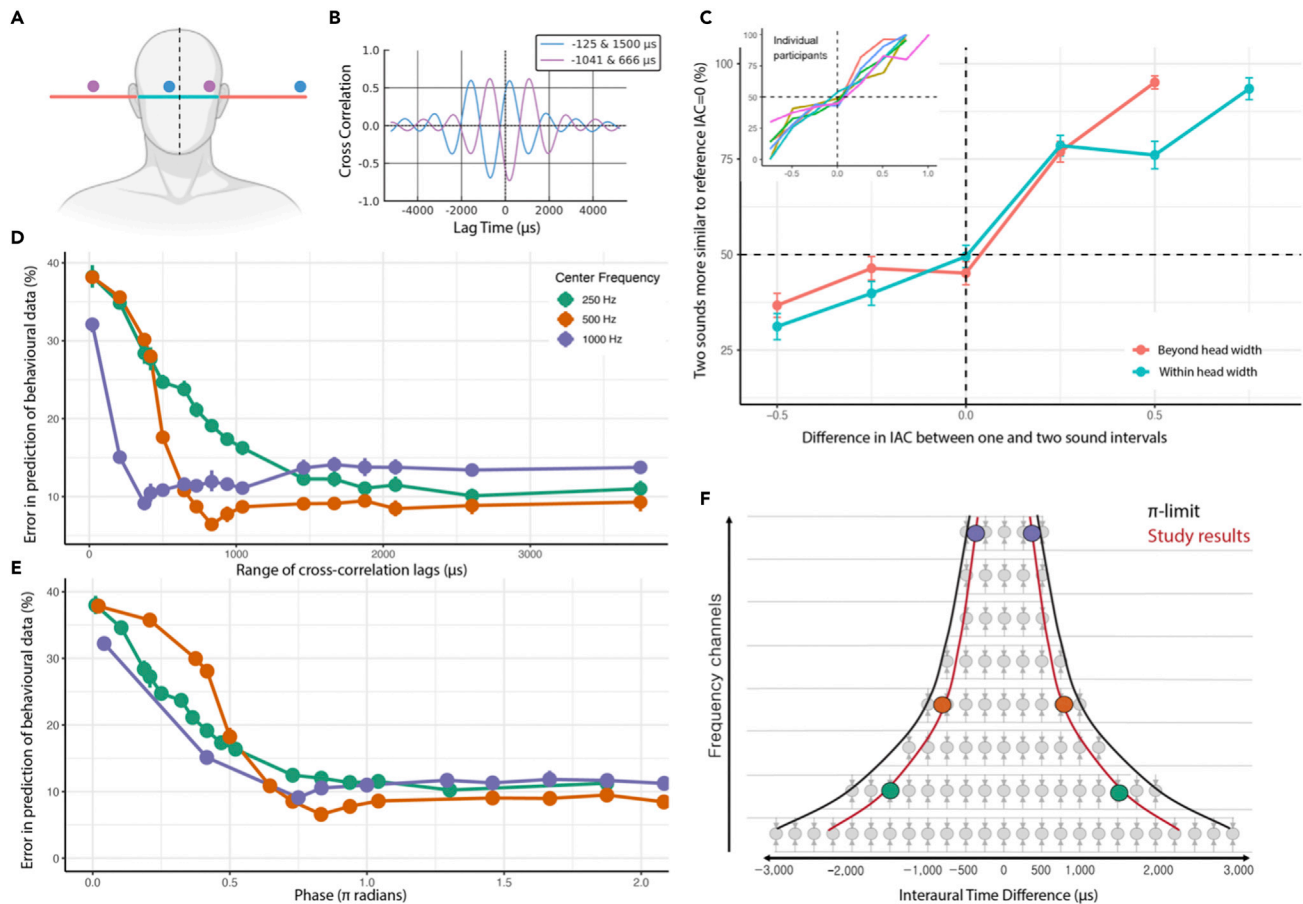
Figure 2. Perception and modeling of multiple sounds located within and beyond the human head width

(A). The horizontal bars illustrate the range of ITDs from which two values were randomly chosen for each trial, the color of the bar represents whether the ITD was considered within or beyond the head width. The dots represent the example conditions used in Figure 2B.

(B). Two examples of cross-correlation functions for two sounds that are not equally lateralized, note that the two examples have vastly different cross-correlation functions but similar maximum, IAC values.

(C). Behavioral responses from all conditions reported as a function of the difference between the IAC of the single-sound and the two-sound stimulus. A difference of zero indicates that the second sound in each interval had the same IAC. Note that the trials with ITDs beyond the head width (red) are reported as more like the reference than the within head width (blue) trials. The inset illustrates the results from each individual participant grouped across all trials (D). The difference between the beyond and within head width results when the cross-correlation computational model used a restricted range of lags to estimate the IAC.

(E). Same as in d, but reported as a function of stimulus phase relative to center frequency.

(F). Schematic of results. Arrangement of coincidence detector as described in von Kriegstein et al. (2008) (gray circles and arrows), overlaid with the lowest ITD in the computational model before prediction error increased. The black line represents the proposed upper limit—the π-limit—to the existence of coincidence detectors. The orange line represents the limit of ITDs predominantly utilized by participants in this experiment.

We next determined whether the IAC for which listeners judged stimuli with two sounds and sounds with reduced coherence to be of equal similarity to the reference signal arose for any combination of ITDs of the two-sound intervals. We assessed this for the same band of noise centered at 500-Hz center frequency and two additional noise bands centered at 250 Hz and 1000 Hz (see STAR Methods). Additional centre frequencies were included to determine if the effect of ITD on this judgment was frequency dependent. Neural recordings from a range of small mammals suggest a frequency-dependent representation of best ITD as observed in small mammals and supporting the existence of a π-limited range of ITD detectors (McAlpine et al., 2001; Thompson et al., 2006). ITDs of the left- and right-leading sounds were independently and randomly selected in the range 125 to 1,500 μs for the 500-Hz and 1000-Hz center frequencies and 125 to 3,000 μs for the 250-Hz center frequency (Figure 2A). Extending the range of ITDs beyond the head-width and center frequencies provides for a greater range of possible cross-correlation functions and finer

sampling of the IAC range (Figure 2B). It also allows us to determine whether the inability to distinguish two coherent sounds from an incoherent one depends on whether ITDs of the two sounds are symmetric around the midline. For trials where the IAC of the single and two-sound stimuli were identical (indicated as zero difference in Figure 2C), listeners selected randomly (binomial test, estimate = 0.474 p = 0.239, 95% CI = 0.433–0.516), confirming the result from the first experiment regarding the importance of IAC in making perceptual judgments about sound-image complexity. For trials in which the stimulus IAC of the single-sound was greater than the IAC of the combined two-sound stimulus (Figure 2C), listeners selected the interval containing two sounds as being more similar to the reference (binomial test, estimate = 0.822 p< 0.001, 95% CI = 0.795–0.847), confirming that listeners can exploit IAC as a cue to distinguish two coherent sounds from one sound with reduced coherence regardless of the symmetry of the two sounds about the midline.

Sounds with ITDs beyond the head width are perceived as more diffuse than sounds within the head width

Most models of auditory-spatial processing include a network of internal delays, the function of which is to represent source location—either as a place or rate code—by offsetting externally applied ITDs. Whether implemented to account for ITDs generated by true sources in free field, or a broader range accessible through headphone listening, a key feature of these models is a reduced number of internal delay taps (or fidelity of the delay mechanism) with increasing magnitude of ITD. Without this reduction, internal delays could offset completely ITDs of infinite magnitude. *In vivo* recordings in a range of small mammals suggest that delay taps exist up to approximately 1/2 the period of the stimulus center frequency (Brand et al., 2002; Harper et al., 2014; Joris et al., 2005; McAlpine et al., 2001); however, neuroimaging studies have yet to specify the distribution in humans (Thompson et al., 2006; von Kriegstein et al., 2008). In our analysis, therefore, we distinguished between trials in which ITDs of both sounds in the two-sound stimuli were restricted to the range generated by the width of the human head (approx. ±680 μs) from those in which one or both ITDs lay beyond this range. In everyday listening, ITDs within the ethological range that persist for sustained periods are most often generated by a true source—localized in physical space—and are assumed to be fully represented by the binaural system. i.e., the IAC represented within the binaural system is the same as the IAC of the signal arriving at the ears. As such, in the context of our headphone-listening experimental data, these "within head-width" trials anchor a listener's response to an internally represented IAC generated by these sustained ITDs.

When the IAC of the two-sound stimulus in one interval was lower than the IAC of the single-sound stimulus in the other (Figure 2C right), two-sounds with long ITDs were perceived as more similar to a fully incoherent (IAC = 0.0) reference than were trials with ITDs confined to the ethological range. This indicates that listeners perceived the sound as if it had a lower IAC than the stimulus signal would suggest, i.e., the internal representation of the IAC for sounds with ITDs beyond the head width is reduced relative to sounds with ITDs confined to the head width. This interpretation is consistent with a range of psychoacoustic-based models suggesting a reduction in the fidelity of the internal representation of ITD detectors for ITDs of increasing magnitude, especially for ITDs beyond the range generated by the size of the head. Fewer or less-effective detectors representing longer ITDs reduce the IAC, leading to a less-accurate representation of the ITDs present in the binaural signal.

To understand if the apparent increase in similarity of sounds with ITDs beyond the head width to the fully incoherent reference sound could be explained by a restricted range of ITD detectors—as observed in other mammals (Brand et al., 2002; Harper et al., 2014; Joris et al., 2005; McAlpine et al., 2001)—the behavioral data were used as input to a computational model designed to determine the range of time delays that best accounted for the difference in perception of sounds with ITDs beyond the ethological range relative to within. For each of the trials containing ITDs beyond the head width, the model estimated an *internal IAC* by varying the range of lags used in the cross-correlation function in the range 0–4,000 μs. The average behavioral response was calculated for each *internal IAC*, and the difference in behavioral performance between the within- and beyond-head width conditions was computed for each model. A minimal difference in performance between the within- and beyond-head width conditions indicates that the model better estimates the cue that listeners exploited when responding to each trial. For the 500-Hz center frequency band of noise, the computational model predicted listeners' responses for trials with ITDs beyond the ethological range with 10% error rate, when the range of lags was greater than 3000 μs. Reducing the range of lags to 2000 μs did not change performance, consistent

with the cross-correlation functions having few maxima above 2000 μs (Figure 2B). However, reducing the range of lags to below 2000 μs improved the model prediction, with an optimal performance occurring for a range of lags up to 820 μs. This is despite the existence of maxima in the cross-correlation functions out to approximately 2000 μs. Reducing the range of lags to below 820 μs reduced performance once more, indicating that a minimum range of lags is required to extract relevant IAC cues. Similarly, at 1000 Hz, an extended range of lags great than 3000 μs provided no benefit beyond that provided by a range extending to 1500 μs, with optimal prediction performance obtained for a range of lags extending only to 400 μs. At 250 Hz, model performance reduced systematically—with no local minimum in error—as the range of lags was reduced. Together, these data indicate that the range of lags which the IAC metric is computed over the range that best accounts for the listener responses is dependent on stimulus centre-frequency, suggestive of a common, phasic representation of internal delays (Brand et al., 2002; Harper et al., 2014; McAlpine et al., 2001). Consequently, we analyzed the data as a function of phase, rather than time, delay (Figure 2E), which illustrated general agreement across the different centre-frequency conditions. The error in the model's prediction of the listener's behavioral responses followed a non-monotonic pattern, particularly for the 500 and 1000 Hz conditions, with a local minimum occurring below π radians, i.e., 1/2 the period of the stimulus centre-frequency. The model minimum error indicates the range of lags in the cross-correlation function that best allowed the model to describe the listener responses. The predictive performance of the model did not improve when the range of lags was increased above 1/2 the period of the stimulus centre-frequency, and the optimal range of lags that best accounted for data across all conditions occurred at 0.825 π. The frequency-dependent, π-limited, representation of ITD detectors required to account for human perception of IAC is consistent with the neural representation of ITDs recorded *in vivo* in a wide range of smaller mammals with various head sizes (Brand et al., 2002; Harper et al., 2014; McAlpine et al., 2001).

Incoherent sounds engage greater cortical resources than coherent sounds

Our behavioral data suggest that increasing the ITD of two sounds beyond the ethological range increases their perceptual ambiguity. ITDs beyond the head width engender additional perceptual ambiguity not accounted for by the IAC of the signal. Reasoning that increased perceptual ambiguity requires the use of more cortical resources, we employed functional near-infrared spectroscopy (fNIRS) to assess cortical neural activation when listening to auditory-spatial percepts with different degrees of interaural coherence. We focused our assessment on *planum temporale*, a cortical region suggested to reflect perception of auditory objects (Bizley and Cohen, 2013), one in which stimulus features are grouped in terms of abstract representations (Griffiths and Warren, 2002,2004). Activity to attended streams is enhanced compared to unattended streams in *planum temporale* (Mesgarani and Chang, 2012), and functional magnetic resonance imaging (fMRI) studies have demonstrated *planum temporale* to be sensitive to IAC (Barrett and Hall, 2006; Zimmer and Macaluso, 2005).

fNIRS measures hemodynamic responses generated by superficial cortical regions only, making it ideal for studying activity in *planum temporale* uncontaminated by signals generated by deeper cortical structures. We confirmed the suitability of fNIRS to investigate activity in *planum temporale* by calculating the specificity of our optode array using photon-migration simulations for different ROIs (Zimeo Morais et al., 2018). Our array was most sensitive to the inferior parietal lobe, encompassing *planum temporale*, with a specificity value of 55%. Specificity to Heschl's gyrus, implicated in the processing of more fundamental acoustic features, including ITD *per se* (Higgins et al., 2017; von Kriegstein et al., 2008), was just 1% (Figure 3A).

Twenty participants (11 female) listened passively to sounds with varying degrees of IAC. 400-Hz bandpass noises centered at 500 Hz were presented for 5 s in randomized order, with intervals of variable duration separating each presentation. Responses averaged across all values of IAC generated a canonical fNIRS hemodynamic response, with a significant increase in oxyhemoglobin concentration relative to the control (silent) condition between 2.2 and 7.8 s after stimulus onset (Figure 3B). Activation was observed over the entire superior temporal gyrus with an increase in activity toward the inferior parietal lobe (Figure 3D).

Our 20 listeners constituted two separate groups of 10. All 20 listeners were exposed to the conditions with perfectly correlated (IAC = 1) and uncorrelated (IAC = 0) sounds, and the control condition. For the combined group of twenty listeners the change in oxyhemoglobin concentration to uncorrelated sounds was 2.8 μM larger than to fully correlated ($z(40) = -3.073$, $p < 0.01$) (Figure 3C).
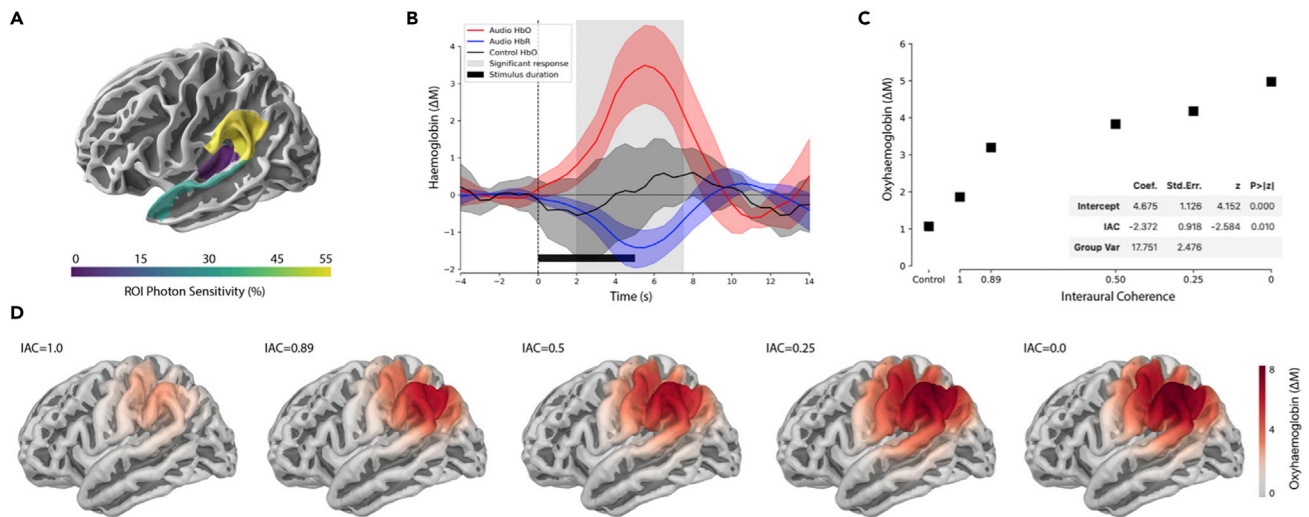
Figure 3. Sounds with reduced interaural coherence elicit larger cortical fNIRS responses

(A). The specificity of the fNIRS technique and optode array to each ROI as determined via photon migration simulations for planum temporale (yellow), superior temporal gyrus (green) and Heschl's gyrus (purple).

(B). Response morphology to all auditory stimuli and silent controls. Gray shading indicates the region where the response to auditory stimuli was significantly larger than the (silent) control condition.

(C). Oxyhemoglobin concentration as a function of IAC. The inset describes the results of the mixed effects model examining the effect of IAC on oxyhemoglobin concentration.

(D). Surface projection of GLM response estimates for all optodes, illustrating the response magnitude increases with decreasing IAC.

We also assessed the demand for cortical resources as a function of IAC by presenting listeners sounds containing a range of IAC values intermediate to 1.0 and zero. One group of 10 listeners was presented an additional intermediate IAC value of 0.89, the effective internal IAC suggested by computational models when a single ITD of 1500 μs is imposed on a 400-Hz band of noise centered at 500 Hz. von Kriegstein et al. (2008) reported ITDs of this magnitude to generate bilateral BOLD activation in primary auditory cortex, in contrast to ITDs of 500 μs—within the ethological range—which evoked contralaterally dominant (to the perceived location of the source) activation in Heschl's gyrus. The second group of 10 listeners was exposed to the additional intermediate IAC values of 0.5 and 0.25. This covers the range over which we observed sounds with variable IAC to be indistinguishable perceptually from fully coherent sounds containing two sounds with equal and opposite ITDs. Combined across listeners and IAC values, the change in oxyhemoglobin increased monotonically with decreasing interaural coherence (Figure 3C), with the factor IAC explaining a significant proportion of the variance in oxyhemoglobin concentration ($z(70) =$ 2.584, $p = 0.01$). This bilateral increase in hemodynamic responses with decreasing IAC is consistent with the elevated fMRI response in *planum temporale* reported for sounds with supra-ethological ITDs presumed to generate internally a similar reduction in IAC (von Kriegstein et al., 2008).

Our data, combined with the relative specificity of the optode array for *planum temporale*, suggest that greater neural resources are required to process less coherent compared to more-coherent sounds. Our data are also consistent with *planum temporale* operating as a computational hub for processing auditory objects (Griffiths and Warren, 2002), with greater neural resources required to form a stable perceptual representation in response to less reliable, potentially more-ambiguous, sensory input.

DISCUSSION

Separating foreground objects from background noise is crucial for understanding our acoustic environment. We sought to determine whether interaurally uncorrelated sounds are perceived as a diffuse background or as two sources, positioned toward either ear. We employed a behavioral task in which listeners were presented with pairs of stimuli containing a diffuse background and either multiple sounds or single sounds with reduced interaural coherence. We found that listeners did not reliably distinguish two sounds from single, interaurally incoherent sounds with the same long-term coherence. These data indicate that

the similarity of sounds at both ears—evaluated in this study using interaural coherence—is a readily available cue that listeners may exploit to distinguish between objects and diffuse noise.

We observed that sound images with relatively short ITDs, lateralized relatively closer to the midline, were easier to distinguish from a single sound with reduced IAC than were sound images with relatively longer ITDs. This suggests that variations in behavioral response with presumed location of sources along the azimuthal plane are well described by the interaural coherence properties of acoustic signals. Furthermore, we observed that ITDs beyond the human range impose an additional reduction in perceived coherence not present in the stimulus, and that this change in perceptual quality is well described by a computational model incorporating a restricted, and frequency-dependent, distribution of internal delay taps. This distribution is consistent with the reported distribution of internal delays in a wide range of smaller mammals, suggesting a conserved representation across species (Brand et al., 2002; Harper et al., 2014; Joris et al., 2005; McAlpine et al., 2001). Employing fNIRS to measure cortical hemodynamic responses, we observed increased hemodynamic responses for sounds with low interaural coherence in the *planum temporale*, a presumed cortical nexus in the parsing of auditory scenes. Together, our data illustrate that interaural coherence is a cue that listeners can readily employ to distinguish between multiple intracranial images and diffuse noise, and that sounds with low interaural coherence—which are perceived more ambiguously—require greater neural resources than unambiguously perceived intracranial images elicited by sounds with a high interaural coherence.

Interaural coherence and perceptual ambiguity

Our data are consistent with threshold measures demonstrating listeners have greater acuity when localizing and tracking auditory objects toward the anterior hemisphere (Carlile and Leung, 2016; Strybel et al., 1992). For the longest ITDs of ± 500 μs, listeners were unable to differentiate these from fully interaurally incoherent noise. This increase in perceptual ambiguity with increasing ITD is consistent with the increase in variability of ITD cues increasing with azimuth—sources located more frontally generate ITDs with less variability, and thus less ambiguity than sounds from the side. Pavão et al. (2020) suggest listeners factor the increased variance in ITD into their perceptual decision-making, such that when they hear sounds presented over headphones their behavioral ITD thresholds account for the natural statistics of ITDs not present in the stimulus itself. However, we were able to account for the within-head width behavioral results in this study, based purely on the interaural coherence properties of the stimulus, without any requirement for compensatory mechanisms that consider the (learned) statistical structure of binaural cues present in the environment

Perception of supra-ethological ITDs is consistent with a frequency-dependent distribution of auditory-spatial detectors

We account for the perceptual ambiguity of multiple sounds in terms of their combined IAC, particularly when the sounds are confined to the range generated by the size of the human head. Supra-ethological ITDs impose an additional apparent reduction in IAC not present in the stimuli. Given the constrained range of internal delays available to offset ITDs of larger magnitude in small mammals (Joris et al., 2005; McAlpine et al., 2001; Pecka et al., 2008) we hypothesized that the additional apparent reduction in IAC for longer ITDs was the result of reduced number or fidelity of internal delays available to offset ITDs of this magnitude. Modeling the range of internal delays required to account for the discrepancy between stimulus IAC and perceptual ambiguity, we found optimal performance arises when the range of internal delay taps is constrained to 1/2 the period of the stimulus centre-frequency, the so-called π-limit. This limit is consistent with the range of internal delays instantiated in the auditory brains of a wide range of smaller mammals including guinea pigs (McAlpine et al., 2001), cats (Hancock and Delgutte, 2004; Joris et al., 2005), and gerbils (Brand et al., 2002; Pecka et al., 2008). Our computational model tested a brick-wall distribution of internal delays and does not rule out a more complex distribution of detectors, for example with a long tail of detectors with lower density out to long ITDs. Nevertheless, consistent with smaller mammals, the majority of sensitivity is focused within the π-limit.

The existence of a π-limit to ITD processing is suggested to arise from the operation of biophysical mechanisms underpinning the sensitivity of neurons in the brainstem to microsecond time differences (Brand et al., 2002), mechanisms that seemingly obviate the requirement for dedicated internal delays that offset externally applied ITDs (McAlpine et al., 2001). The dominant models of human ITD-processing (Stern and Trahiotis, 1992; Stern et al., 1988), informed by headphone listening tasks, include internal delay taps that explicitly represent ITDs of many thousands of microseconds to account for performance in

headphone-listening tasks. Notably, these models specifically exclude sensitivity to IAC in their analysis. Our data support the view that neural mechanisms for spatial hearing reflect the ethological relevance of acoustic features, namely, sounds of sustained duration containing ITDs limited to the range generated by the size of the head, and transient, rapidly fluctuating ITDs, including those beyond the ethological range, that contribute to background features of the listening environment. The operation of a π-limit to ITD coding provides a succinct explanation for the spatial percept generated by two auditory objects with ethologically plausible spatial locations and the additional reduction in perceived IAC generated by supra-ethological ITDs in headphone-listening tasks.

Incoherent sounds require greater cortical neural resources than coherent sounds

Our fNIRS data indicating a largely monotonic increase in oxyhemoglobin utility in *planum temporale* as IAC was reduced suggests an increased neural demand for processing less-coherent sounds. *Planum temporale* is widely implicated in the formation of auditory objects (Bizley and Cohen, 2013; Griffiths and Warren, 2004), these objects are commonly considered to be processed within a parallel hierarchical-processing model in which space and motion—predominantly represented by binaural cues—are processed in the dorsal "where" pathway, and auditory objects in the ventral "what" pathway (Rauschecker and Tian, 2000). Our data are consistent with this model underestimating the mutual interaction between the two pathways and the extent to which spatial information alters object perception (Bizley and Cohen, 2013). In particular, we demonstrate that modifying binaural spatial cues directly alters intracranial image perception, consistent with studies illustrating spatial cues modulate segregation of auditory streams (Middlebrooks and Onsan, 2012), and neural activity in structures associated with object perception.

Our finding of increased neural demand for less-coherent sounds is consistent with electrophysiological findings in humans (Ando et al., 1987; Chait et al., 2005; Soeta et al., 2004) that report larger onset responses for uncorrelated noise compared to correlated noise, and delayed responses for changes from incoherent sounds compared to coherent sounds. However, findings from fMRI studies are less consistent, with some studies observing increased responses for sounds with greater IAC (Budd et al., 2003; Zimmer and Macaluso, 2005,2009), and others the reverse (Hall et al., 2005). Previous discrepancies between electrophysiology and hemodynamic results have been attributed to "a lack of understanding of how hemodynamic BOLD responses are related to the electrical physiological brain responses measured by MEG." (Chait et al., 2005). Our fNIRS data provide an alternative measure of the hemodynamic response to fMRI and indicates that discrepancies may instead be due to methodological differences. The sounds presented in this study had spectral content between 300 and 700 Hz. The frequency range, centered at 500 Hz, is the most advantageous for spatial listening in natural environments (Hartmann et al., 2005), above this range sensitivity to IAC is considerably reduced (Culling et al., 2001; Walther and Faller, 2013). Previous BOLD studies employed stimuli with frequency content up to 1.5 kHz (Budd et al., 2003; Hall et al., 2005) and even 22 kHz (Zimmer and Macaluso, 2005,2009). The inclusion of higher-frequency spectral components changes the perception of sounds, with sound images generated reported as being more perceptually compact (Blauert and Lindemann, 1986).

Interaural coherence is an ideal tool with which to assess the representation of auditory images independent of changes in the power or spectral content of sounds presented to the ears. Simply by reducing the moment-by-moment similarity of the signal at each ear the perception can be modified between a punctate unambiguous image and a diffuse percept. Within the predictive coding framework (Friston, 2010), modulating the interaural coherence varies the reliability the ITD and ILD cues, such that the features contained in signals with high IAC are more predictable than signals with low IAC. This framework would suggest that when listening to sounds with a low IAC the system model is unable to predict upcoming input, and thus requires constant updating and error processing which, consistent with our observations, would requires greater cortical resources than predictable sounds with high IAC.

Limitations of the study

Although attention is not specifically required for the perception of auditory objects (Alain, 2007; Alain et al., 1994; Bizley and Cohen, 2013; Micheyl et al., 2003), attention modulates listeners' awareness of them (Alain and Arnott, 2000; Shinn-Cunningham, 2008) and of their neural representation (Mesgarani and Chang, 2012). As such, differences in attention may account for in the reported discrepancies across studies, and may limit direct comparison of the behavioral and neuroimaging results in our own study, which constituted active and passive tasks, respectively.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

AUTHOR CONTRIBUTIONS

Robert Luke: Conceptualization, Data Collection, Methodology, Formal Analysis, Writing. Hamish Innes-Brown: Conceptualization, Data Collection, Methodology, Formal Analysis, Writing, Jaime Undurraga: Conceptualization, Data Collection, Methodology, Formal Analysis, Writing, David McAlpine: Conceptualization, Resources, Methodology, Formal Analysis, Writing

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Aaronson, N.L., and Hartmann, W.M. (2010). Interaural coherence for noise bands: waveforms and envelopes. J. Acoust. Soc. Am. *127*, 1367–1372.

Alain, C. (2007). Breaking the wave: effects of attention and learning on concurrent sound perception. Hear. Res. *229*, 225–236.

Alain, C., and Arnott, S.R. (2000). Selectively attending to auditory objects. Front. Biosci. *5*, D202–D212.

Alain, C., Woods, D.L., and Ogawa, K.H. (1994). Brain indices of automatic pattern processing.

Neuroreport: Int. J. Rapid Commun. Res. Neurosci. *6*, 140–144.

Ando, Y., Kang, S.-H., and Nagamatsu, H. (1987). On the auditory-evoked potential in relation to the IACC of sound field. J. Acoust. Soc. Jpn. *8*, 183–190.

Barrett, D.J., and Hall, D.A. (2006). Response preferences for "what" and "where" in human non-primary auditory cortex. Neuroimage *32*, 968–977.

Best, V., Baumgartner, R., Lavandier, M., Majdak, P., and Kopčo, N. (2020). Sound externalization: a

review of recent research. Trends Hear. *24*, 2331216520948390.

Bizley, J.K., and Cohen, Y.E. (2013). The what, where and how of auditory-object perception. Nat. Rev. Neurosci. *14*, 693–707.

Blauert, J., and Lindemann, W. (1986). Spatial mapping of intracranial auditory events for various degrees of interaural coherence. J. Acoust. Soc. Am. *79*, 806–813. https://doi.org/10.1121/1.393471.

Brand, A., Behrend, O., Marquardt, T., McAlpine, D., and Grothe, B. (2002). Precise inhibition is

essential for microsecond interaural time difference coding. Nature 417, 543–547.

Bregman, A.S. (1994). Auditory Scene Analysis: The Perceptual Organization of Sound (MIT press).

Budd, T.W., Hall, D.A., Goncalves, M.S., Akeroyd, M.A., Foster, J.R., Palmer, A.R., Head, K., and Summerfield, A.Q. (2003). Binaural specialisation in human auditory cortex: an fMRI investigation of interaural correlation sensitivity. Neuroimage 20, 1783–1794. https://doi.org/10.1016/j.neuroimage.2003.07.026.

Carlile, S., and Leung, J. (2016). The perception of auditory motion. Trends Hear. 20. https://doi.org/10.1177/2331216516644254.

Chait, M., Poeppel, D., de Cheveigne, A., and Simon, J.Z. (2005). Human auditory cortical processing of changes in interaural correlation. J. Neurosci. 25, 8518–8527. https://doi.org/10.1523/JNEUROSCI.1266-05.2005.

Chait, M., Poeppel, D., de Cheveigne, A., and Simon, J.Z. (2007). Processing asymmetry of transitions between order and disorder in human auditory cortex. J. Neurosci. 27, 5207–5214. https://doi.org/10.1523/JNEUROSCI.0318-07.2007.

Culling, J.F., Colburn, H.S., and Spurchise, M. (2001). Interaural correlation sensitivity. J. Acoust. Soc. Am. 110, 1020–1029. https://doi.org/10.1121/1.1383296.

Feddersen, W., Sandel, T.T., Teas, D., and Jeffress, L. (1957). Localization of high-frequency tones. J. Acoust. Soc. Am. 29, 988–991.

Friston, K. (2010). The free-energy principle: a unified brain theory? Nat. Rev. Neurosci. 11, 127–138.

Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., and Hamalainen, M. (2013). MEG and EEG data analysis with MNE-Python. Front. Neurosci. 7, 267. https://doi.org/10.3389/fnins.2013.00267.

Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Parkkonen, L., and Hamalainen, M.S. (2014). MNE software for processing MEG and EEG data. Neuroimage 86, 446–460. https://doi.org/10.1016/j.neuroimage.2013.10.027.

Griffiths, T.D., and Warren, J.D. (2002). The planum temporale as a computational hub. Trends Neurosci. 25, 348–353.

Griffiths, T.D., and Warren, J.D. (2004). What is an auditory object? Nat. Rev. Neurosci. 5, 887–892.

Grothe, B., Pecka, M., and McAlpine, D. (2010). Mechanisms of sound localization in mammals. Physiol. Rev. 90, 983–1012.

Hall, D.A., Barrett, D.J., Akeroyd, M.A., and Summerfield, A.Q. (2005). Cortical representations of temporal structure in sound. J. Neurophysiol. 94, 3181–3191. https://doi.org/10.1152/jn.00271.2005.

Hancock, K.E., and Delgutte, B. (2004). A physiologically based model of interaural time

difference discrimination. J. Neurosci. 24, 7110–7117.

Harper, N.S., Scott, B.H., Semple, M.N., and McAlpine, D. (2014). The neural code for auditory space depends on sound frequency and head size in an optimal manner. PLoS One. 9, e108154.

Hartmann, W.M., Rakerd, B., and Koller, A. (2005). Binaural coherence in rooms. Acta Acustica. 91, 451–462.

Higgins, N.C., McLaughlin, S.A., Rinne, T., and Stecker, G.C. (2017). Evidence for cue-independent spatial representation in the human auditory cortex during active listening. Proc. Natl. Acad. Sci. 114, E7602–E7611.

Huppert, T.J. (2016). Commentary on the statistical properties of noise and its implication on general linear models in functional near-infrared spectroscopy. Neurophotonics 3, 010401. https://doi.org/10.1117/1.NPh.3.1.010401.

Jones, S.J., Pitman, J.R., and Halliday, A.M. (1991). Scalp potentials following sudden coherence and discoherence of binaural noise and change in the inter-aural time difference: a specific binaural evoked potential or a"mismatch" response? Electroencephalogr. Clin.Neurophysiol. 80, 146–154.

Joris, P.X., van der Heijden, M., Louage, D.H., Van de Sande, B., and Van Kerckhoven, C. (2005). Dependence of binaural and cochlear "best delays" on characteristic frequency. In Auditory Signal Processing, D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet, eds. (Springer), pp. 477–483. https://link.springer.com/chapter/10.1007/0-387-27045-0_58.

Julesz, B., and Tyler, C.W. (1976). Neurontropy, an entropy-like measure of neural correlation, in binocular fusion and rivalry. Biol.Cybernetics. 23, 25–32.

Kaigham, G.J., and Colburn, H.S. (1981). Interaural correlation discrimination: i. bandwidth and level dependence. J. Acoust. Soc. Am. 69, 1394–1401.

Kuhn, G.F. (1977). Model for the interaural time differences in the azimuthal plane. J. Acoust. Soc. Am. 62, 157–167.

Kumar, S., Stephan, K.E., Warren, J.D., Friston, K.J., and Griffiths, T.D. (2007). Hierarchical processing of auditory objects in humans. PLoS Comput. Biol. 3, e100.

Luke, R. (2021). AuditoryStimuli.jl: A julia package for generating real-time auditory stimuli. J.Open Source Softw. 6, 3613. https://doi.org/10.21105/joss.03613.

Luke, R., Larson, E., Shader, M.J., Innes-Brown, H., Van Yper, L., Lee, A.K., Sowman, P.F., and McAlpine, D. (2021a). Analysis methods for measuring passive auditory fNIRS responses generated by a block-design paradigm. Neurophotonics 8, 025008.

Luke, R., Shader, M.J., and McAlpine, D. (2021b). Characterization of Mayer-wave oscillations in functional near-infrared spectroscopy using a physiologically informed model of the neural power spectra. Neurophotonics 8, 041001. https://doi.org/10.1117/1.NPh.8.4.041001.

Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods. 164, 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024.

McAlpine, D., Jiang, D., and Palmer, A.R. (2001). A neural code for low-frequency sound localization in mammals. Nat. Neurosci. 4, 396–401. https://doi.org/10.1038/86049.

McEvoy, L.K., Picton, T.W., and Champagne, S.C. (1991). The timing of the processes underlying lateralization: psychophysical and evoked potential measures. Ear Hearing. 12, 389–398.

Mesgarani, N., and Chang, E.F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. Nature 485, 233–236.

Micheyl, C., Carlyon, R.P., Shtyrov, Y., Hauk, O., Dodson, T., and Pullvermüller, F. (2003). The neurophysiological basis of the auditory continuity illusion: a mismatch negativity study. J. Cogn. Neurosci. 15, 747–758.

Middlebrooks, J.C., and Onsan, Z.A. (2012). Stream segregation with high spatial acuity. J. Acoust. Soc. Am. 132, 3896–3911.

Mills, A.W. (1958). On the minimum audible angle. J. Acoust. Soc. Am. 30, 237–246.

Neurobehavioral Systems Presentation®Software. https://www.neurobs.com.

Pavão, R., Sussman, E.S., Fischer, B.J., and Peña, J.L. (2020). Natural ITD statistics predict human auditory spatial perception. Elife 9, e51927.

Pecka, M., Brand, A., Behrend, O., and Grothe, B. (2008). Interaural time difference processing in the mammalian medial superior olive: the role of glycinergic inhibition. J. Neurosci. 28, 6914–6925.

Pollack, I., and Trittipoe, W.J. (1959). Binaural listening and interaural noise cross correlation. J. Acoust. Soc. Am. 31, 1250–1252.

Rauschecker, J.P., and Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc. Natl. Acad. Sci. 97, 11800–11806.

Salminen, N.H., Jones, S.J., Christianson, G.B., Marquardt, T., and McAlpine, D. (2018). A common periodic representation of interaural time differences in mammalian cortex. Neuroimage 167, 95–103. https://doi.org/10.1016/j.neuroimage.2017.11.012.

Santosa, H., Zhai, X., Fishburn, F., and Huppert, T. (2018). The NIRS brain analyzIR toolbox. Algorithms 11. https://doi.org/10.3390/a11050073.

Shader, M.J., Luke, R., Gouailhardou, N., and McKay, C.M. (2021). The use of broad vs restricted regions of interest in functional near-infrared spectroscopy for measuring cortical activation to auditory-only and visual-only speech. Hear Res. 406, 108256. https://doi.org/10.1016/j.heares.2021.108256.

Shinn-Cunningham, B.G. (2008). Object-based auditory and visual attention. Trends Cogn. Sci. 12, 182–186.

Soeta, Y., Hotehama, T., Nakagawa, S., Tonoike, M., and Ando, Y. (2004). Auditory evoked magnetic fields in relation to interaural cross-correlation of band-pass noise. Hear. Res. *196*, 109–114. https://doi.org/10.1016/j.heares.2004.07.002.

Stern, R.M., and Trahiotis, C. (1992). The role of consistency of interaural timing over frequency in binaural lateralization. In Auditory Physiology and Perception, Y. Cazals, K. Horner, and L. Demany, eds. (Elsevier), pp. 547–554. https://doi.org/10.1016/B978-0-08-041847-6.50067-8.

Stern, R.M., Zeiberg, A.S., and Trahiotis, C. (1988). Lateralization of complex binaural stimuli: a weighted-image model. J. Acoust. Soc. Am. *84*, 156–165.

Strangman, G., Franceschini, M.A., and Boas, D.A. (2003). Factors affecting the accuracy of near-infrared spectroscopy concentration calculations for focal changes in oxygenation parameters. Neuroimage *18*, 865–879.

Strybel, T.Z., Manllgas, C.L., and Perrott, D.R. (1992). Minimum audible movement angle as a function of the azimuth and elevation of the source. Hum. Factors. *34*, 267–275.

Thompson, S.K., Von Kriegstein, K., Deane-Pratt, A., Marquardt, T., Deichmann, R., Griffiths, T.D., and McAlpine, D. (2006). Representation of interaural time delay in the human auditory midbrain. Nat. Neurosci. *9*, 1096–1098.

Traer, J., and McDermott, J.H. (2016). Statistics of natural reverberation enable perceptual separation of sound and space. Proc. Natl. Acad. Sci. *113*, E7856–E7865.

van der Heijden, M., and Trahiotis, C. (1999). Masking with interaurally delayed stimuli: the use of "internal" delays in binaural detection. J. Acoust. Soc. Am. *105*, 388–399.

von Kriegstein, K., Griffiths, T.D., Thompson, S.K., and McAlpine, D. (2008). Responses to interaural time delay in human cortex. J. Neurophysiol. *100*, 2712–2718. https://doi.org/10.1152/jn.90210.2008.

Walther, A., and Faller, C. (2013). Interaural correlation discrimination from diffuse field reference correlations. J. Acoust. Soc. Am. *133*, 1496–1502.

Wiggins, I.M., Anderson, C.A., Kitterick, P.T., and Hartley, D.E. (2016). Speech-evoked activation in adult temporal cortex measured using functional near-infrared spectroscopy (fNIRS): are the measurements reliable? Hear. Res. *339*, 142–154. https://doi.org/10.1016/j.heares.2016.07.007.

Zimeo Morais, G.A., Balardin, J.B., and Sato, J.R. (2018). fNIRS Optodes' Location Decider (fOLD): a toolbox for probe arrangement guided by brain regions-of-interest. Sci. Rep. *8*, 3341. https://doi.org/10.1038/s41598-018-21716-z.

Zimmer, U., and Macaluso, E. (2005). High binaural coherence determines successful sound localization and increased activity in posterior auditory areas. Neuron *47*, 893–905. https://doi.org/10.1016/j.neuron.2005.07.019.

Zimmer, U., and Macaluso, E. (2009). Interaural temporal and coherence cues jointly contribute to successful sound movement perception and activation of parietal cortex. Neuroimage *46*, 1200–1208. https://doi.org/10.1016/j.neuroimage.2009.03.022.

STAR★METHODS

KEY RESOURCES TABLE

RESOURCE	SOURCE	IDENTIFIER
Software		
MNE-Python v0.24.0	MNE Developers	https://mne.tools
MNE-NIRS v0.0.6	MNE Developers	https://mne.tools/mne-nirs
AuditoryStimuli.jl v0.0.12	Robert Luke	https://rob-luke.github.io/AuditoryStimuli.jl/
Deposited data		
Human data	This paper	https://doi.org/10.17632/gky2sj9y6x.1

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Robert Luke (robert.luke@mq.edu.au).

Materials availability

This study did not generate new unique reagents

Data and code availability

● Raw data were deposited on Mendeley at: Luke, Robert (2021), "Separating Sources from Background", Mendeley Data, V1, https://doi.org/10.17632/gky2sj9y6x.1.

● Code is included in the *code* directory alongside the data deposited in Mendeley Data.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All data were collected under the Macquarie University Ethics Application Reference 52020640814625 and all participants provided informed consent to participate in the experiments.

METHOD DETAILS

Participants listened to auditory stimuli presented binaurally via Etymotic Research ER-2 insert-phones connected to an RME Fireface UCX sound card (16 bits, 44.1 kHz sampling rate). Stimuli were presented at 75 dBA as measured with a Norsonic sound level meter (Norsonic SA, Norway) connected to an ear simulator (RA0045 G.R.A.S., Denmark). The sound level meter was calibrated to an Casella Cel-110/2 sound source.

Behavioral experiments

Two behavioural experiments were performed that employed the same overall task. The task consisted of a two alternative forced choice procedure. Each of the two intervals contained two consecutive stimuli, each stimulus consisted of 400 Hz bandwidth noise with a centre frequency of 500 Hz, and had a duration of 0.5 seconds with an onset and offset ramp of 0.01 seconds. In total each interval had a duration of 1 s and intervals were separated by 0.5 s of silence. The first stimulus in each interval was incoherent noise (IAC = 0). The second stimulus in each interval was either a single-sound with variable IAC selected in the range of 0 to 1 in steps of 0.2, or a two-sound stimulus, which was generated by applying various ITDs (as described below in section Behavioural experiment 1 and Behavioural experiment 2 respectively). The order of the intervals was randomised, and the listeners were asked to "*select the interval with the two sounds that were most similar.*" The responses were converted to scores indicating the percentage of trials in which participants rated the interval containing the 'two-sound stimulus as the one which contained the two most similar sounds. If (for a given combination of ITD for the two-sound stimulus and IAC for the single-sound stimulus) participants had a high score, this would indicate that they consistently rated a

two-sound stimulus with a given ITD as more similar to the reference sound (a diffuse sound with IAC = 0) than a stimulus containing a single sound with the modified IAC. Participants were able to take breaks whenever they wished, but the software encouraged a break at three equally spaced moments in each session. Each session lasted approximately 30 minutes. All stimuli were generated and presented using the Julia programming language and the package *AuditoryStimuli.jl* (Luke, 2021).

Behavioural experiment 1: equally lateralised sources

In Experiment 1, the two sound stimuli consisted of two independent noise sources that had equal and opposite ITDs applied to them – often described as double delayed noise (van der Heijden and Trahiotis, 1999). This experiment consisted of 5 conditions; each condition was conducted in a single session. In each session a single ITD was applied in the two sound stimuli. The ITDs tested were ±250 µs, ±375 µs, ±500 µs, ±750 µs, or $\pm1,500$ µs. Additionally, a control condition was included where the one interval contained two sounds with IAC = 0, and the other interval contained a single sound with a negative ITD of -680 µs. Seven adult participants (4 female, 3 male) volunteered for this study, and each condition was completed by a total of 5 participants. The order in which each participant conducted the conditions was randomised. Each session was conducted in blocks, with each block containing all seven conditions (6 IAC values and 1 control), blocks were repeated until at least 210 trials were completed. All listeners achieved over 97% accuracy in the control condition.

Behavioural experiment 2: not equally lateralised sources

In Experiment 2, the two sound stimuli consisted of two independent noise sources which had independently selected ITDs with opposite signs applied for the left and right sounds. This is in contrast to Experiment 1, where the ITD for the left and right source were symmetrical about the midline. The experiment consisted of three sessions in which the centre frequency and bandwidth of the stimulus was varied. The centre frequencies were 250 Hz (200 Hz bandwidth), 500 Hz (400 Hz bandwidth), and 1,000 Hz (400 Hz bandwidth). For each trial the ITDs were selected randomly from a uniform distribution between 125 µs and 1,500 µs for the 500 and 1000 Hz centre frequencies, and 125 µs and 3000 µs for the 250 Hz centre frequency. All trials were presented in a randomised sequence, participants completed 150 trials per session. For each centre frequency, 5 adult participants (3 female, 2 male) who did not participate in experiment 1 volunteered for this experiment.

Neuroimaging protocol

Participants were seated in a sound treated booth in a comfortable chair for the duration of the experiment. Participants watched a silent subtitled film of their choice and were instructed to watch the film and not pay attention to the sounds. A NIRx NIRScout device was used to collect data continuously throughout the experiment. Sixteen sources and sixteen detectors were positioned to cover the superior temporal gyrus and inferior parietal lobe (Shader et al., 2021). Light sources were placed at positions PP05h, P5, TPP7h, CPP3h, CP3, CCP5h, C5, FTT7h, PPO6h, P6, TPP8h, CPP4h, CP4, CCP6h, C6, FTT8h. Detectors were placed at positions P3, CPP5h, CP5, TTP7h, T7, CCP3h, C3, FCC5h, P4, CPP6h, CP6, TTP8h, T8, CCP4h, C4, FCC6h.

All stimuli were 5 s in duration and consisted of bandpass noise with a centre frequency of 500 Hz and 400 Hz bandwidth. A 0.1 s onset and offset ramp applied to each stimulus. The stimuli varied in the value of the IAC, and there was a control condition that consisted of 5 s of silence. Twenty five trials of each stimulus (including the silent control) were presented to each participant in a randomised order. The delivery of stimuli was controlled by Presentation (Neurobehavioral Systems). The maximum duration of the neuroimaging experiment was 70 min.

This study consists of two datasets, each containing 10 adult participants (11 female, 9 male). Both datasets contained the silent control condition, and interaural coherence conditions of IAC of 0 and 1. The first set additionally contained intermediate IAC conditions of 0.5 and 0.25. And the second set contained the intermediate IAC condition of 0.89.

QUANTIFICATION AND STATISTICAL ANALYSIS

Behavioural experiment 1: equally lateralised sources

Data was analysed using the R (version 4.0.2) software with the tidyverse packages (version 1.3.0). Participant responses were converted to a percentage score per participant, that indicated the percentage of

trials in which participants rated the interval containing the 'two-sound' stimulus as the one which contained the two most similar sounds. The mean response value and 95% confidence interval across participants is illustrated in Figure 1D per IAC and centre frequency. To determine the IAC value at which listeners selected randomly, a linear line was fitted between the mean percentage scores and the point at which the fitted line crossed the 50% value was extracted.

Behavioural experiment 2: not equally lateralised sources

For each trial, the IAC of the two-sound interval was determined from the signal. Then, for each trial the difference between the IAC of the two-sound interval and the fixed IAC sound was computed and stored as $IAC_{difference}$. As above, participant responses were converted to a percentage score, that indicated the percentage of trials in which participants rated the interval containing the 'two-sound' stimulus as the one which contained the two most similar sounds per value of $IAC_{difference}$. The trial data was split into two groups, the first "within head width" group contained trials where the absolute value of both ITDs as less 680 μs, the "beyond head width" group contained trials where the absolute value of both ITDs was more than 680 μs. The mean participant response per $IAC_{difference}$ is displayed with 95% confidence intervals in Figure 2C per group of trials. To determine if listeners selected randomly when the IAC difference was zero, a binomial test was used to compare the trials with zero IAC difference to 50%.

Computational modelling

The aim of the modelling was to determine if the increase in similarity of sounds with ITDs beyond the head width to the fully incoherent refence sound could be explained by a restricted range of ITD detectors. As such, the within head width trials were treated as reference performance, as the IAC of these signals are assumed to be fully represented by the binaural system. Then, for the beyond head width trials, the cross-correlation function was computed from the signal over a restricted range of lags to simulate a restricted range of ITD detectors (as predicted by the π-limit model (McAlpine et al., 2001; Thompson et al., 2006)). For each range of lags an *internal* IAC metric was then extracted for the beyond head width trials. After which, the average behavioural response was calculated for each *internal IAC* for each range of lags, and the difference in behavioural performance between the within- and beyond-head width conditions was computed. A minimal difference in behavioural responses between the within- and beyond-head width conditions, per IAC and *internal IAC*, indicates that the model better estimates the effective IAC that users exploited to perform the behavioural task, and indicates the likely range of lags over which the listeners extracted information to perform the task. Specifically, for each of the two sound stimulus waveforms, the cross-correlation function was calculated with lags up to 4000 μs. The maximum value of the cross-correlation function was then extracted from within a restricted lag range and termed $IAC_{restricted(lag)}$. The range of restricted lags was systematically evaluated between 21 and 4000 μs. As above, the trials were split into two groups depending on if both ITDs were less than or greater to 680 μs. For the "within head width" trials, corresponding to ITDs less than 680 μs, the difference between the IAC of the two-sound interval and the fixed IAC sound was computed as described above and stored as $IAC_{difference}$. For the "beyond head width" trials, the difference between the $IAC_{restricted(lag)}$ of the two-sound interval and the fixed IAC sound was computed and stored as $IAC_{difference\_restricted(lag)}$. Participant responses were converted to a percentage score, that indicated the percentage of trials in which participants rated the interval containing the 'two-sound' stimulus as the one which contained the two most similar sounds per $IAC_{difference}$ and $IAC_{difference\_restricted(lag)}$ for the within- and beyond head width trials respectively. The percentage scores of the "within head width" trials per $IAC_{difference}$ were used as a reference with which to compare the "beyond head width" responses. For each range of restricted lags, the difference in percentage scores per $IAC_{difference\_restricted(lag)}$ was compared to the "within head width" percentage score. Then the mean absolute difference between within- and beyond-head width was computed over the range of differences and is reported per range of lags in Figures 2D and 2E. Computational modelling was performed using the Julia language (version 1.3.1).

Neuroimaging analysis

We present both a qualitative and quantitative analysis of the data. The analysis procedure is based on Luke et al. (2021a), and representative analysis code is provided at the first authors website (https://mne.tools/mne-nirs). The qualitative analysis visualises the grand average waveform. The quantitative statistical analysis is based on the approach of Huppert (2016)which accounts for the unique properties of the fNIRS signal (Luke et al., 2021b). We present group level fNIRS analysis as this has been shown to be reliable for auditory evoked responses (Wiggins et al., 2016). All analysis was performed using MNE (version 0.24)

(Gramfort et al., 2013, 2014) and MNE-NIRS (version 0.0.6) (Luke et al., 2021a), all functions used in this analysis are publicly available. The FOLD toolbox (Zimeo Morais et al., 2018) was used to determine the channels most sensitive to the planum temporale, this bilateral region of interest contained the optodes CPP3h, CP3, CCP5h, CPP4h, CP4, CCP6h and P3, CPP5h, CP5, CCP3h, C3, P4, CPP6h, CP6, CCP4h, C4.

Qualitative analysis

Data was resampled to 4 Hz. Source detector pairs with distances between 2 and 4 cm were retained, all others were discarded. The signal was then converted to optical density and channels with a scalp coupling index of less than 0.9 were discarded. The signal was then converted to haemoglobin concentration using the modified Beer Lambert law and filtered between 0.05 and 0.45 Hz. Epochs were extracted from 3 seconds before to 11 seconds after stimulus onset, with a linear detrend applied to each epoch. Any epoch with a peak to peak value greater than 60 μM was discarded.

To confirm responses were being measured to auditory, but not control conditions, all the auditory conditions were collapsed together. A cluster-level statistical permutation test was used to determine the time window over which the NIRS signal was significantly different to zero (Maris and Oostenveld, 2007).

Quantitative analysis

The data was first resampled to 0.6 Hz (Luke et al., 2021a), followed by conversion to optical density, then conversion to haemoglobin using the modified Beer-Lambert law with the toolbox default partial path length factor of 0.1 (Santosa et al., 2018; Strangman et al., 2003). GLM analysis was then performed utilising a first order autoregressive model (Huppert, 2016), the glover canonical haemodynamic response model, and including a cosine drift model. The channel level results were combined in to a region of interest (ROI) by taking a weighted average of the beta estimates, with the weighting equal to the inverse of the standard error. First a mixed effects model was run on the conditions recorded in all 20 participants (control, IAC=0, IAC=1). The mixed effects model examined the effect of condition with participant as a random factor. Second, a mixed effects model was run on all IAC conditions (silent control excluded). This model examined the effect of IAC with participant as a random factor.