Title: Listening to the room: disrupting activity of dorsolateral prefrontal cortex impairs learning of room acoustics in human listeners

Heivet Hernández-Pérez * [1], Jessica J.M. Monaghan [2], Jason Mikiel-Hunter [1] James Traer [3], , Paul F Sowman [4], David McAlpine [1]

[1] Department of Linguistics and Macquarie University Hearing, The Australian Hearing Hub, Macquarie University, Australia, [2] National Acoustic Laboratories, Australia, [3] Psychological and Brain Sciences Department, The University of Iowa, US [4] Auckland University of Technology, Physiotherapy, Auckland, New Zealand.

* Corresponding author
Heivet Hernandez-Perez
Email: heivet.hernandez-perez@mq.edu.au

Abstract

Navigating complex sensory environments is critical to survival, and brain mechanisms have evolved to cope with the wide range of surroundings we encounter. To determine how listeners learn the statistical properties of acoustic spaces, we assessed their ability to perceive speech in a range of noisy and reverberant rooms. Listeners were also exposed to repetitive transcranial stimulation (rTMS) to disrupt the dorsolateral prefrontal cortex (dlPFC) activity, a region believed to play a role in statistical learning. Our data suggest listeners rapidly adapt to statistical characteristics of an environment to improve speech understanding. This ability is impaired when rTMS is applied bilaterally to the dlPFC. The data demonstrate that speech understanding in noise is best when exposed to a room with reverberant characteristics common to human-built environments, with performance declining for higher and lower reverberation times, including fully anechoic (non-reverberant) environments. Our findings provide evidence for a reverberation “sweet spot” and the presence of brain mechanisms that might have evolved to cope with the acoustic characteristics of listening environments encountered every day.

Introduction

Learning occurs over multiple time scales—evolutionary, developmental and moment-to-moment—to support a diverse range of abilities: navigating complex sensory environments (Bregman, 1994; Lewicki et al., 2014; Smith & Lewicki, 2006), expressing intricate behaviours in social settings (Gariépy et al., 2014; van den Bos et al., 2013), or acquiring communication skills such as bird song (Brainard & Doupe, 2002; Lauay et al., 2004; Marler, 1970) or spoken language (Aslin et al., 1998; Saffran et al., 1996; Saffran, 2003). Although some learning requires active or explicit involvement in a task (Huyck & Wright, 2011; Mathews et al., 1989; Rebuschat, 2015), perhaps with a system of rewards and punishments to make the learning 'stick' (Barberis, 2013; Schultz, 2002; Wächter et al., 2009), other forms of learning seem automatic or implicit (Reber, 1967), acquired with individuals seemingly unaware it is taking place. This type of learning, referred to as 'statistical learning' (Ambrus et al., 2020; Saffran et al., 1996), 'sequence learning' (Nissen & Bullemer, 1987; Vékony et al., 2022) or 'sequential learning' (Conway & Christiansen, 2001; Vékony et al., 2022), is thought to entail automatic and incidental extraction of regularities or patterns within external stimuli or in the environment (Conway, 2020; Takács et al., 2021). Evident across sensory modalities to support automatic learning of tonal (Saffran et al., 1999) or linguistic (Saffran et al., 1996) sequences, strings of letters (Reber, 1967), visual scenes and shapes (Fiser & Aslin, 2001), visual-motor patterns (Nissen & Bullemer, 1987), and even tactile input (Conway & Christiansen, 2005) statistical learning appears to be a unitary, domain-general phenomenon, potentially governed by a single mechanism or neurocognitive principle (Conway, 2020; Kirkham et al., 2002).

Increasing evidence suggests that statistical learning is a critical part of how listeners deal with complex and cluttered acoustic scenes. Potential background sounds such as rain or insects—referred to as sound textures (Hicks & McDermott, 2024; McDermott et al., 2013; McWalter & McDermott, 2018) as well as changes in the regularity of sound patterns (Barascud et al., 2016; Bianco et al., 2020) are processed—seemingly unconsciously—in terms of their summary statistics. This form of statistical learning likely contributes to our ability to follow conversations in background noise ('cocktail party listening'; Cherry, 1953) and to deal with reverberant spaces where multiple, delayed copies of the same sound reach a listener in the form

of reflections from walls and other acoustically opaque surfaces (Blesser & Salter, 2009; Sabine, 1953; Schroeder, 1962). Though listeners rely on early-arriving sound energy to determine source location—suppressing potentially conflicting localization cues in later-arriving, often more intense, sound energy (Blesser & Salter, 2009; Bradley et al., 1999; Culling et al., 2003; Houtgast & Steeneken, 1985; Nielsen & Dau, 2010)—the perception of reflected sound energy is informative of the listening environment more broadly (Bronkhorst & Houtgast, 1999; Shinn-Cunningham, 2000). This includes whether environments are indoors or outdoors, their identity and dimensions (Brumm & Naguib, 2009; Cabrera et al., 2005; Kolarik et al., 2021; Zahorik & Wightman, 2001), as well as the number of occupants or potential interfering sources (Bradley et al., 1999; Culling et al., 2003; Hawley et al., 2004; Houtgast & Steeneken, 1985; Nielsen & Dau, 2010; Peissig & Kollmeier, 1997).

Nevertheless, despite its potential utility for understanding background features of a sound environment, the accumulation over time of late-arriving, reverberant sound energy is thought to generate an additional burden on listening performance beyond that from sound energy direct from interfering sources (Houtgast & Steeneken, 1973; Knudsen, 1929; Lochner & Burger, 1961; Santon, 1976; Shinn-Cunningham & Kawakyu, 2003), smearing the acoustic waveform and occluding temporal gaps that might otherwise be helpful for ‘glimpsing’ speech in background noise (Cooke, 2006). If listeners are able to utilize late-arriving, reverberant energy of known acoustic environments to supress disruptive spatial cues and enhance speech understanding (Brandewie & Zahorik, 2010, 2013; Vlahou et al., 2019; Watkins, 2005a, 2005b), this suggests that the acoustic characteristics of an environment can be learned and stored for later use in complex listening tasks.

Here, using an ecologically relevant listening task—understanding speech in background noise—we assessed the ability of human listeners to learn the statistical structure of different sound environments defined by their $RT_{60,}$ the time it takes for reverberant energy to decay by 60 decibels (dB), and found that speech understanding in background noise improved over time. Specifically, when asked to report words from unfamiliar and semantically uninformative spoken sentences of varying duration in noise convolved with the reverberant qualities of different acoustic environments, listeners’ performance improved with increasing sentence duration, and with repeated

exposure to each environment (Brandewie & Zahorik, 2010, 2013). This capacity for learning the acoustic environment to aid speech understanding was diminished when repetitive transcranial magnetic stimulation (rTMS) was applied bilaterally to impair the function of dorsolateral prefrontal cortex (dlPFC), a cortical locus hypothesised to contribute to statistical learning (Ambrus et al., 2020; Vékony et al., 2022). Counter to expectations that reverberant energy can only harm speech understanding, listeners' abilities to leverage the knowledge of a sound environment (talker identity, speech corpus, noise characteristics, spatial configuration) to improve speech understanding was best when a 'typical' room with reverberant characteristics close to the average of a wide range of common (built) environments was included as one of the three environments presented in a single experimental run. Performance declined systematically when this room was switched for one with more, or less, reverberation, including a fully anechoic (i.e., non-reverberant or dry) environment in which listeners initially trained on the recall task. Importantly in the context of other potentially learnable acoustic features, talker identity (three male and three female) represented a random variable in our experimental design. Evidence for a reverberation 'sweet spot' suggests the existence of brain mechanisms adapted to the longer-term structure of common listening environments (Traer & McDermott, 2016).

Our data suggest listeners rapidly 'tune in' or adapt to the background (the statistical structure of the sound environment, including its reverberation profile) and use this knowledge to improve their performance in a listening task. This benefit of learned listening appears most evident when encountering levels of reverberation commonly experienced in everyday settings. The data also support the view that listeners retain information about the acoustic background long enough for it to influence listening performance at some later time, consistent with *in vivo* experimental evidence of increased adaptive capacity for neural learning of sound environments upon repeated exposure to those environments (Dean et al., 2005, 2008; Ivanov et al., 2022) and that cortical circuits modulate this capacity (Robinson et al., 2016). If long-term learning of reverberant environments relies on, or even establishes, a preferred range of RT_{60}s for speech understanding, it suggests modifications might be required to listening assessments and technologies—routinely developed under anechoic conditions—to generate optimal performance outcomes.

Materials and Methods

Participants

All 62 participants (53 females, ages between 19-26 years old (mean ± SD = 22 ± 2 years old)) included in the study were Australian native-English speakers, had normal pure tone thresholds (< 20 dB HL tested at octave intervals between 0.5-8 kHz (Hughson & Westlake, 1944); Interacoustics Hearing Aid Fitting Analyzer Affinity 2.0 Audiometry) and normal middle ear function (assessed using standard 226 Hz tympanometry; Titan, Interacoustics). To ensure normal outer hair cell function, all participants were screened for Distortion Product Otoacoustic Emissions (DPOAEs) between 0.5-10 kHz (stimulus parameters: f1/f2 =1.2, f1 =65 dB SPL; f2 =55 dB SPL; responses parameters: SNR > 6 dB, signal level > -10 dB SPL, and reliability > 98% (DPOAE440; Titan, Interacoustics). All participants had normal steady-state ipsilateral broadband noise middle ear muscle reflexes (MEMR) (Titan, Interacoustics).

Acoustic stimuli

To assess listeners' abilities to understand speech in background noise across different sound environments, they were asked to verbally report keywords spoken by a talker virtually located in front of the participant that were masked by white noise arriving from a virtual source 90° to the left (Figure 1A) (Brandewie & Zahorik, 2010, 2013). A binaural configuration was necessary as speech intelligibility improvements following exposure to room acoustics is significantly reduced under monaural listening conditions (Brandewie & Zahorik, 2010). The speech stimuli used in this study were from the Coordinate Response Measure (CRM) corpus (Bolia et al., 2000), and all combinations of, |Callsigns| ('Baron', 'Eagle', 'Charlie', 'Tiger', 'Arrow', 'Ringo', 'Laker', 'Hoper'), |Colors| (four monosyllabic choices, 'red', 'white', 'blue', or 'green'), and |Numbers| (the English digits between 'one' and 'eight') were used in this experiment (Figure 1B) (Brandewie & Zahorik, 2010, 2013). The masking white noise was randomly generated on a PC using MATLAB: *RRID:SCR_001622* (The MathWorks, Inc. of Natick, MA: https://au.mathworks.com/) and preceded the speech stimuli by 150 ms, during which its amplitude linearly increased from zero to full scale. The masker was present throughout the speech and ended (without ramping) with the speech.

Virtual Environments

Reverberant environments were reproduced by convolving the generated acoustic signals with Room Impulse Responses (RIRs) obtained using 62-channel microphone array recordings of real rooms that were subsequently decoded into 41 higher-order Ambisonic (HOA) channels (Badajoz-Davila et al., 2020; Weisser et al., 2019). These final decoded channels corresponded to the spherical array of 41 Tannoy V8 concentric loudspeakers (Tannoy) installed in the anechoic chamber at the Australian Hearing Hub where testing took place. To optimize the "directionality" of the acoustic stimuli, the speech and masking noise signals were separately convolved with RIRs that had enhanced direct sound components in loudspeakers associated with their respective virtual source locations (i.e., 0º azimuth loudspeaker for the target talker and 90º azimuth loudspeaker for the masking noise) (Badajoz-Davila et al., 2020; Weisser et al., 2019). In the anechoic condition, RIRs only incorporating these enhanced direct sound components were convolved with the speech and masking noise signals. The SNR for speech-in-noise during data collection was -15 dB to avoid ceiling performance and was manipulated after convolution with the RIR by adjusting the gain of the speech target relative to a fixed masker level of 70 dB SPL. All signal processing was performed on a PC using MATLAB (Mathworks) and the final spatialized signal were presented via the RME MADI sound card (RME Audio) to two RME 32-channel digital-to-analog converters (M-32, RME Audio). These, in turn, fed 11 Yamaha XM4180 power amplifiers (Yamaha) that drove the loudspeaker array.

To compare our recordings to RIR data recorded and analyzed by Traer and McDermott (2016), we also calculated RT_{60}s for each room by integrating across the 41 HOA channels of each RIR and then calculating the median RT_{60} value in 31 frequency sub-bands with centre frequencies ranging from 80 Hz to 10 kHz (Traer & McDermott, 2016). We refer to these measures of reverberation as $RT_{60traer}$. The range of sub-bands and their frequency selectivity was chosen to match that of the human ear (Glasberg & Moore, 1990; McDermott & Simoncelli, 2011) . See *SI Materials and Methods,* (Traer & McDermott, 2016), for further information concerning how individual $RT_{60traer}$s were extracted from each sub-band.

Identity of the selected sound environments

Three rooms were selected, with similar Reverberation Times (RT_{60}s) to those employed by Brandewie and Zahorik (2013). As Bajadoz-Davila et al. (2020) and

Brandewie and Zahorik (2013) calculated these RT_{60}s according to ISO-3382 guidelines (ISO, 2009), they are referred to in the text as RT_{60iso}: Lecture room (RT_{60iso} = 0.46 s), Open-Plan Office (RT_{60iso} = 0.96 s) and Underground Car Park (RT_{60iso} = 2.42 s) (Table 1). Additionally, we assessed performance in different combinations of rooms to determine whether a specific combination was necessary to improve speech performance as 'carrier phrase length' was increased. For each combination, Lecture Room was swapped with one of three virtual environments: Anechoic (RT_{60iso} = n/a), Living Room (RT_{60iso} = 0.33 s) and Highly Reflective Room (RT_{60iso} of 1.55 s) (Table 1). Naïve participants were recruited for each room combination.

Continuous theta-burst stimulation

Applying repetitive Transcranial Magnetic Stimulation (rTMS)—specifically continuous theta burst stimulation (cTBS)—to the right and left dlPFC, elicits a period of depressed cortical excitability in the targeted area that lasts about an hour post-stimulation (Gamboa et al., 2010; Hoogendam et al., 2010; Huang et al., 2005), the time during which speech recall in reverberant rooms was assessed. Participants were screened for contra-indications to rTMS (see Supplemental Information. Transcranial Magnetic Stimulation screening form). Two cTBS conditions ('real' and 'sham' TMS) were counterbalanced across normal-hearing participants, and participants were blinded to the type of manipulation they might receive.

cTBS intensity was individually measured for every participant (i.e., both 'real' and 'sham' TMS conditions) as a function of their corticospinal excitability assessed with single-pulse TMS motor evoked potentials (MEPs). First, single-pulse TMS (Magstim Rapid2 system, Magstim) was delivered over the left and right primary motor cortex to determine the optimal site for MEP elicitation and to determine resting and active motor thresholds for each participant and each side of the brain. A 70 mm figure-of-eight coil was orientated at 45° to the scalp with current flowing posterior-anterior across the primary motor cortex. Coil position and angle was adjusted until the optimal site for consistent elicitation of MEPs was identified. Once the optimal site for stimulation was located, this was marked on the scalp. The resting motor threshold was then determined by the experimenter visually identifying the minimal single-pulse TMS intensity necessary to elicit a MEP from the right and left first dorsal interosseous

muscle, while the hand was at rest in 5 out of 10 consecutive stimulations (Rothwell et al., 1999; Sandrini et al., 2011).

Participants' resting left and right motor thresholds ranged from 46-57% of the maximum stimulator output (mean = 51 ± 5). Participants' individual motor threshold for left and right hemisphere were recorded to set the intensity for bilateral cTBS stimulation which was administered by placing the same figure-of-eight coil over the right and left dlFPC located over electrode positions F3 and F4 (10-20 system EEG) (Herwig et al., 2003; Jurcak et al., 2005). Bilateral TMS was performed serially i.e., first the right or the left dlFPC (order was alternated among subjects), was chosen to limit possible compensation effects of the non-stimulated hemisphere (Ambrus et al., 2020). Each cTBS burst consisted of three pulses at 50 Hz, with bursts repeated at a frequency of 5 Hz, applied continuously for 40s, and delivered at an intensity of 80% of the right or left resting motor threshold (Huang et al., 2005).

Procedure

All participants completed first a familiarization task that consisted of 10 trials (i.e., different 'carrier phrase lengths') presented in anechoic conditions at 0 dB SNR where listeners were asked to report |Color| and |Number|. Participants exposed to 'real' or 'sham' TMS completed the familiarization task after cTBS procedures. All participants performed the familiarization phase with 80-100% accuracy, confirming that all participants including exposed to cTBS understood the task and procedural learning was achieved and not affected by either 'sham' or 'real' TMS exposure. The CRM sentences were spoken by six talkers (3 males and 3 females) and were of varying duration where all, some, or none of the preceding sentence (the 'carrier phrase' (CP, Figure 1B)) before '|Color| |Number|' was included (Brandewie & Zahorik, 2010, 2013). Thus, for CP0, listeners heard '|Color| |Number| now', CP1—'go to |Color| |Number| now', CP2—'|Callsign| go to |Color| |Number| now', CP3—'Ready |Callsign| go to |Color| |Number| now' (Figure 1B and C). After each phrase was presented, participants reported the |Color| and |Number| they heard to the experimenter and performance was assessed based on keywords correctly identified. When collecting performance data in reverberant environments at -15 dB SNR (SNR that allows a level of performance similar to (Brandewie & Zahorik, 2010, 2011, 2013; Zahorik, 2009)), the room environment was selected randomly for each trial except that the same room

could not appear in two consecutive trials. The target |Color|, |Number| and talker were selected randomly for each trial. For each experimental session, three different listening environments were assessed, and listeners were presented 360 trials (30 repeats of each 'room' x 'carrier phrase length' condition) with a total test time of 45 min per session. Each listener participated in one session only. Throughout the session, listeners were required to verbally repeat the appropriate |Color| and |Number| combination they heard from corpus lists located to the sides of the 0º azimuth loudspeaker. The CRM is a close-set corpus, therefore participants were provided with |Color| and |Number| choices to select from (Brandewie & Zahorik, 2010, 2011, 2013; Pavel Zahorik, 2009). The experimenter continuously monitored participants' responses while scoring performance, but no feedback was provided.

Speech Performance Analysis and Time course-fittings of mean cumulative hit rates

d' was first calculated individually for both |Color| and |Number| in different room/carrier combinations using Equation 1, where *z(H)* and *z(F)* were the *z* transforms of the hit rate and false alarm, respectively. Hit rates consisted of the correct |Color| or |Number| being selected, whereas false alarms were considered when the same |Color| or |Number| was reported as an incorrect response to the presentation of any other |Color||Number| combination. To prevent infinite values of d', hit rates/false alarms of 1 were set to 0.99, and corresponding false alarms/hit rates of 0 were set to 0.01 leading to maximal/minimal d' values of ± 4.65. The total d' for a carrier phrase length in a particular room was calculated by averaging |Color| and |Number| d' values for that combination of room and 'carrier phrase length' across individuals.

Equation 1.

$$d' = z(H) + z(F)$$

Given how *d'* was calculated for |Color| and |Number|, it was not considered a useful metric to describe performance across time in different environments: the paucity of data for each |Color| and |Number| affecting the temporal resolution of any generated curve. Therefore, to understand how the average performance developed in each

listening environment as a function of time, we averaged cumulative hit rate across all subjects, after a 5-point moving average was implemented for each individual trace (equivalent to ~7 seconds) and plotted this as a function of the mean exposure time accumulated in each environment in question. Time courses of the developing performance were quantified by fitting a double exponential function to the curves, whose model is described as follows:

Equation 2.

$$f(t) = C + mag_{fast} \cdot \exp\left(-\frac{t}{\tau_{fast}}\right) + mag_{slow} \cdot \exp\left(-\frac{t}{\tau_{slow}}\right)$$

Where *f(t)* is the fit as a function of time; mag_{fast}/mag_{slow} are the magnitudes and τ_{fast}/τ_{slow} are the time constants for the two exponential fits. *C* is a constant term that describes their offset on the y-axis.

Constrained optimization of the fits was achieved using the fmincon function in Matlab to find local minima in mean-squared errors. Fits were reinitialized using a combination of 50 values linearly spaced between 50 s and 500 s for $\tau_{slow}$and between 0 s and 100 s for τ_{fast}; upper bounds of 2000 s and 200 s were set for τ_{slow} and τ_{fast} respectively. The optimal fit for a listening environment was selected from the resulting 2450 fits by identifying the fit with the smallest objective function value, fval. During optimization, fits were weighted by their variance; therefore, greater importance was attributed to performance at later exposure times. Adjusted R^2 values are quoted for each fit in Supplemental Figure 1, 4 and 5. For analysis of 'global learning', we calculated the absolute percentage deviation of cumulative hit rate curves relative to the final time point in each curve i.e., their Final Hit Rate (FHR), we then calculated back in time to determine where fitted curves to the data first deviated from FHR by a threshold of 10%.

Statistical analysis

Repeated measures ANOVAs (rANOVA) were performed to assess if speech understanding (d') was affected by the factors: rooms (levels: Lecture Room, Open-Plan Office and Carpark) and carrier phrase length (levels: CP0, CP1, CP2 and CP3),

in a within-subjects factor analysis. Univariate ANOVA was employed to assess whether speech understanding was affected by the factors: TMS exposure (levels Sham and real TMS), rooms (levels: Lecture Room, Open-Plan Office and Carpark) and carrier phrase length (levels: CP0, CP1, CP2 and CP3) in a between-subjects factor analysis.

Univariate ANOVA was also used to assess whether speech understanding was affected by room-context (i.e., the 3rd room environment in which Open-Plan Office and Underground Car Park are learnt). Only one factor was assessed in this between-subjects room context (levels: Anechoic room, Living room, Lecture room and Highly Reflectant room). Effect sizes were calculated for all statistical analysis (Partial Eta-squared (ηp2) or Cohen's d when appropriate). Further two-tailed t-tests (alpha=0.05, with Bonferroni corrections for multiple comparisons) were also performed. The statistical analysis was performed in SPSS: RRID:SCR_002865 (IBM Corp. Released 2023. IBM SPSS Statistics for Windows, Version 29.0.2.0 Armonk, NY).

Power Analysis

Initial sample size estimation (≥ 18) was computed using G*Power: *RRID:SCR_013726* (Faul et al., 2007) (Effect size f = 0.25; α err prob = 0.05; Power (1-β err prob) = 0.95). However, given the large effect size observed in previous studies (Brandewie & Zahorik, 2010, 2013; Srinivasan & Zahorik, 2012; Pavel Zahorik & Brandewie, 2016) using sample sizes = 9-16 listeners, we expected relatively large effect sizes with a sample ≥ 10. All variables variance were tested for normal distribution (Shapiro-Wilk test, (Shapiro & Wilk, 1965).

Data availability statement

All data and related metadata were deposited in Figshare, an appropriate public repository (https://doi.org/10.25949/24295342.v1)

Code Information

All custom code used for data collection and analysis will be available upon request.

Results

Human listeners can incorporate knowledge of a listening environment, specifically its reverberant characteristics, to improve speech understanding in noise (Brandewie & Zahorik, 2010, 2013; Srinivasan & Zahorik, 2012; Zahorik & Brandewie, 2016). To understand how this learning of sound environments is achieved, we recreated acoustic characteristics of real rooms using an array of loudspeakers located in an anechoic chamber (Figure 1A) and assessed listeners' performance in a speech-in-noise task using sentences from the Coordinate Response Measure (CRM) corpus—"Ready |Callsign| go to |Color| |Number| now" (Figure 1B). The CRM corpus is commonly used to quantify listening performance in noisy or cluttered environments (Bolia et al., 2000). Listeners reported the |Color| (one of four monosyllabic choices, 'red', 'white', 'blue', or 'green') and the |Number| (the English digits between 'one' and 'eight') they heard for sentences of varying duration where all, some, or none of the preceding sentence (the 'carrier phrase length') before '|Color| |Number|' was included (Figure 1B). Importantly, words in the CRM phrase preceding '|Color| |Number|' are uninformative as to the color and number (Brandewie & Zahorik, 2010, 2013). CRM sentences—complete or partial (see Methods and Figure 1B)—were presented as if originating from in front of a participant, whilst a randomly generated white noise was presented from a source 90° to their left. Listeners' abilities to recognize |Color| and |Number| from whole or partial CRM sentences were assessed. Although sentences varied in duration from trial to trial, they always contained the element '|Color| |Number| now' embedded in noise were convolved with the impulse response of the real rooms (Figure 1C). Sentence length varied by adjusting the duration of the preceding 'carrier phrase' (CP; Figure 1B-C), which was always constructed from the same CRM sentence embedded in noise and convolved with the same impulse response as the remainder of the phrase. Thus, for CP0, listeners heard only '|Color| |Number| now', CP1—'go to |Color| |Number| now', CP2—'|Callsign| go to |Color| |Number| now', CP3—'Ready |Callsign| go to |Color| |Number| now'. After each phrase was presented, participants verbally reported the |Color| and |Number| they heard to the experimenter and performance was assessed based on keywords correctly identified (Figure 1D).

We first confirmed that understanding speech in background noise depends on the reverberant characteristics of the listening spaces from which impulse responses were obtained. Overall, RT_{60iso}s [RT_{60}s according to ISO-3382 guidelines (ISO, 2009)] of

the (six) environments varied from fully anechoic to a highly reverberant underground car park with RT_{60iso} = 2.42 s. Our initial assessment (Figure 1D) examined listening performance in three environments: a Lecture Room (RT_{60iso} = 0.45 s), an Open-Plan Office (RT_{60iso} = 0.96 s), and an Underground Car Park (RT_{60iso} = 2.42 s). Listeners reported |Color| and |Number| from CRM phrases of varying length, spoken by one of 6 talkers (3 female, 3 male) in one of the three environments. 'Carrier phrase length' and talker were interleaved in a pseudorandom order to avoid the same environment being presented in consecutive trials (e.g., Figure 1C). Overall, listeners performed better (quantified in terms of d' values for reporting the correct |Color| and |Number|) in the Lecture Room—the least reverberant of the 3 environments (Figure 2A)—with a significant main effect of room [rANOVA: $F(2,42) = 76.75$, $p<0.001$, $\eta p2 = 0.79$]. Bonferroni-corrected *post-hoc* pairwise comparisons demonstrated that performance in the Lecture Room was significantly better than in the Open-Plan Office [mean difference = 0.41, $p < 0.001$] and Underground Car Park [mean difference = 1.05, $p < 0.001$]. Performance was also significantly better in the Open-Plan Office when compared to the Underground Car Park [mean difference = 0.65, $p < 0.001$].

Statistical learning of reverberant environments occurs over long and short-time courses

Statistical learning likely occurs over different time courses subject to a range of possible brain mechanisms controlling or modulating the different cadences over which learning emerges (Robinson et al., 2016; Simpson et al., 2014). We sought to distinguish short-term from longer-term learning by assessing benefit to word recall of prior exposure to the listening environment over multiple time courses. The design of our paradigm—with talker, length of carrier phrase, target words |Color| & |Number| and reverberation time (listening environment) constituting random variables—means that the initial phase of learning may also be highly variable; listeners, idiosyncratically, likely experience very different parameters over the first few trials, making it difficult to assess the rate at which learning accumulates over these early epochs, and we indeed found this to be the case (see Supplemental Figure 1). Given this, we assessed the rate at which listeners accumulate knowledge to improve listening performance across the task by assessing 'global likelihood learning'—defined here as the time at which participants achieved stable performance in each environment and quantified as the time point (backwards in time) at which performance reached ± 10% of the Final Hit

Rate (FHR)]. For the 22 listeners, this measure of global learning (Figure 2B) was similar across all rooms (a total of 180 trials per room and average trial length of 1.5s): Lecture Room: [42.95 ± 38.35 s]; Open-Plan Office: [54.50 ± 29.50 s] and Car Park: [56.95 ± 33.15 s], with a one-way ANOVA revealing no statistical differences in global learning for the three reverberant environments, suggesting listeners learned these environments at the same rate.

We next wondered whether the FHR achieved by a participant within an environment was related to the time point at which global learning had manifest. To this end, a Pearson's correlation analysis revealed FHR and the timepoint at which stable performance was achieved were negatively correlated for the Lecture Room: $r_{(22)} = -0.63$, $p = 0.002$, 95% CI [-0.84, -0.32] and Open Plan Office $r_{(22)} = -0.49$, $p = 0.02$, 95% CI [-0.74, -0.11], but not for the more-highly reverberant Car Park $r_{(22)} = 0.02$, $p = 0.94$, 95% CI [-0.29, 0.39], Figure 2C. This analysis suggests that the earlier in time a listener achieves asymptotic performance (i.e., the time point defined as global learning)—presumably by accumulating knowledge about these environments over time—the higher their overall performance (defined by FHR) in the task, at least in the two less-reverberant environments assessed here.

We next analysed whether speech understanding—quantified in terms of d'—improves as the length of the 'carrier phrase' was increased (Figure 2D). Consistent with Brandewie and Zahorik (2010; 2013), we found a significant main effect of length of 'carrier phrase': [rANOVA: $F(3,63) = 26.59$, $p<0.001$, $\eta p2 = 0.56$]. Bonferroni-corrected *post-hoc* pairwise comparisons revealed a significant effect for most carrier phrase lengths (Supplemental Table 1) other than between the two longest, CP2 *vs*. CP3. This plateau in performance between CP2 and CP3 suggests an upper limit in the ability to exploit/accumulate information in reverberant listening environments with increasing sound duration, consistent with Brandewie and Zahorik's (Brandewie & Zahorik, 2013) observation of a plateau (and a decline in some reverberant environments) in listening performance with increasing length of carrier phrase, and the existence of an upper limit of listeners to exploit prior exposure to sound environments to benefit listening (Hicks & McDermott, 2024; McDermott et al., 2013; McWalter & McDermott, 2018).

We hypothesized that if prior exposure to the statistics of reverberant rooms arises from an increase in the length of the carrier phrase, then performance to |Number| will always be better than to |Color| for the shortest carrier phrase (CP0: '|Color| |Number| now'; Figure 2E) as |Number| is subject to a longer preceding phrase than |Color|. Assessed in terms of reporting |Color| and |Number| alone, listeners performed significantly better (i.e., showed greater sensitivity) for |Number| compared to |Color|: main effect of target word: [rANOVA F (1,21) =48.79, p<0.001, ηp2 =0.70], [mean difference = 0.38, p < 0.001] despite there being twice as many (eight compared to four) possibilities. Moreover, a significant interaction 'carrier phrase length' x 'target word' was observed: [rANOVA F (1,21) =4.33, p=0.008, ηp2 =0.17]. *Post-hoc* pairwise comparisons (see Supplemental Table 2) indicate that for all carrier phrases except CP3, performance to |Number| was significantly higher than for |Color|. This result also suggests that a roll-over effect i.e., a lack or limit to the improvement in performance is evident for carrier phrases longer than CP2 across environments (Brandewie & Zahorik, 2013; McWalter & McDermott, 2018).

A potential explanation for poorer performance in accurately reporting |Color| relative to |Number| for the shortest carrier phrase (CP0) impact of the utterance of |Color|, namely its immediate appearance at the start of the phrase i.e., the impact of "order/certainty or predictability" in statistical learning (Conway et al., 2010; Daikoku & Yumoto, 2023). One way to determine whether |Color| presented in the context of the CP0 is at all subject to statistical learning (and is therefore not wholly reliant on the short-term accumulation of information within a CP0 trial; see Figure 3G) is to assess whether performance for |Color| for the shortest carrier phrase (CP0) improves from the longer-term accumulation of knowledge i.e., the process of meta-adaptation (Robinson et al., 2016) in which adaptation to short-term statistics improves with repeated exposure to those statistics. Specifically, we tested the hypothesis that, independent of the environment, performance for |Color| in later *steady* trials (here, trials 9-10) of the shortest carrier phrase, CP0, are better compared to *initial* (1-2) trials (Figure 2F). If this hypothesis is supported, it suggests performance for |Color| for CP0 benefits from meta-adaptive information conveyed in later trials as knowledge about the global structure of the environment accumulates over time. Consistent with this hypothesis, a Wilcoxon signed rank test revealed significant better performance (n=22, Z= -2.16, p=0.03) for |Color| for *steady* trials (9-10; assumed to reflect a meta-adaptive

state; Robinson et al., 2016) compared to *initial* trials (1-2; i.e., short-term adaptation). Interestingly, in anechoic conditions i.e., in the absence of reverberation, this meta-adaptive process—the expected improvement in performance between *initial* and *steady* trials—was not observed: Wilcoxon signed rank test [n=10, Z= -0.71, p=0.48]. This suggests that performance for |Color| conveyed in the shortest carrier phrase, CP0, improves over time even in the absence of immediate information in the form of any preceding carrier phrase, with knowledge of the statistical/acoustical properties of environments accumulating over the course of the experimental task.

Improvements in performance are explained by exposure to the environment not talker idiosyncrasies

Normal-hearing listeners are reported to understand speech slightly better when listening to male talkers in a mixture of male and female talkers (Larsby et al., 2015), and quickly adapt to talkers' idiosyncrasies such as non-native accented speech (Idemaru & Holt, 2011, 2014; Liu & Holt, 2015). Here, we wondered if exposure to different talkers (3 female and 3 male) could explain the improvement in speech as the length of the carrier phrase was increased (Supplemental Figure 1). However, this was not the case. Despite listening performance to some talkers appearing overall better than for others [main effect of talkers: rANOVA: $F(5,105) = 27.19$, $p<0.001$, $\eta p2 = 0.56$, i.e., significantly worse performance for Talker 1 (male) and Talker 6 (female), see Supplemental Figure 2A and Supplemental Table 3], overall differences in performance due to a specific talker did not explain the benefit to speech understanding as the length of the carrier phrase was increased in each environment (i.e., lack of significant interaction: or 'talker' x 'carrier' x 'room': $F(30,630) = 0.73$, $p=0.85$, $\eta p2 = 0.034$]).

To explore further the possibility of any talker-specific effects on performance, we assessed the potential benefit of experiencing the same talker on consecutive trials across an experiment, hypothesizing that, independent of room characteristics or length of carrier phrase, experiencing the same talker on consecutive trials would provide a benefit to listening performance. Across all our participants, a total of 1262, 217, and 40 trials were identified as having two, three, or four consecutive trials in which the same talker appeared (see Supplemental Figure 3). Despite the possibility that sustained experience of the same talker might lead to improved listening

performance, a Wilcoxon signed rank test revealed no significant differences in performance between trial 1 *vs*. 2 consecutive/same talker trials (n=1262, Z= -0.42, p=0.68), trial 1 *vs*. 3 consecutive/same talker trials (n=217, Z= -0.17, p=0.87) or trial 1 vs. 4 consecutive/same talker trials (n=40, Z= -0.159, p=0.11). Our data suggest that listeners do not use talker identity, at least in the task reported here, to benefit their speech-in-noise understanding.

We also explored the extent to which a talker and/or the task itself could be learned by analysing improvements in performance with increasing exposure to carrier phrases and talkers in the absence of reverberation i.e., under anechoic conditions. Although a significant rANOVA was observed: F (3,27) = 9.10, p=0.007, ηp2 =0.57; post-hoc pairwise comparisons, with Bonferroni corrections, revealed that only a significant improvement in performance was observed only between CP0 and CP2 [mean difference = 15.67, p = 0.02; see Supplemental Table 4], suggesting that in anechoic conditions very little improvement in performance is achieved by learning the talker/task as length of carrier phrase increases. These data support us employing a listening task with high ecological relevance—comprehending speech—to demonstrate the potential benefits of learning background acoustic features to leverage listening performance without requiring listeners to attend to or report features related to the statistics of presumed background sounds *per se* (e.g., Agus et al., 2014, McWalter & McDermott 2018, Bianco et al., 2020).

TMS disrupts long- but not short-term adaptation to an environment's reverberation profile

The capacity for learning the statistical structure of acoustic environments, the better to understand speech embedded in background noise, suggests a real-world benefit to this form of learned listening. To determine possible brain mechanisms underlying this ability, we reversibly impaired dorsolateral prefrontal cortex (dlPFC, Figure 3A) — a brain region implicated in listening performance in noise (Houtgast & Steeneken, 1973; Knudsen, 1929; Lochner & Burger, 1961)—using repetitive transcranial magnetic stimulation (rTMS) and then assessed the ability of listeners to recall |Color| and |Number| in our modified CRM sentences. Repetitive TMS is posited to elicit a period of depressed cortical excitability in the targeted area that persists for about an hour post-stimulation (Gamboa et al., 2010; Hoogendam et al., 2010; Huang et al.,

2005). Specifically, 10 naïve, normal-hearing listeners were subjected to ‘real’ TMS—continuous theta burst on dlPFC for 40s on each side—and then transferred to the anechoic chamber, where they were presented sequences of CRM phrases of varying phrase length in background noise convolved with one of the three listening environments (Lecture Room, Open-Plan Office, and Underground Car Park) as before. A second, control, group of 11 naïve participants underwent ‘sham’ TMS stimulation in which an otherwise-identical TMS procedure was performed with a ‘sham’ TMS coil. All participants were naïve to differences in the TMS procedure as well as to the listening task; indeed, as with the experimental group, participants in the ‘sham’ group experienced a short period of single-pulse TMS stimulation to obtain motor thresholds, followed by the ‘sham’ TMS with the stimulator positioned over dlPFC. This process familiarised all participants with the procedure and the influence of TMS in generating involuntary finger movements through direct stimulation of motor cortex. We assume that participants in the ‘sham’ group believed later assessment of speech recall occurred under the influence of ‘real’ TMS.

Given the reduction in sample size, as well as the potential placebo effects of a ‘sham’ TMS stimulation, we first confirmed that our sample of 11 ‘sham’ TMS participants exhibited similar behavioural performance to the larger sample of 22 participants who had not been exposed to any TMS manipulation. To avoid sample size imbalances in these comparisons, a random sample of 11 subjects was selected from the 22 participants. A Univariate analysis confirmed that overall performance in these two populations was comparable, with no significant main effect of conditions (‘sham’ *vs*. no exposure to TMS) observed: [$F(1,263) = 1.10$, $p = 0.29$, $\eta p2 = 0.005$]. As expected, significant main effect of rooms [$F(2,131) = 83.84$, $p < 0.001$, $\eta p2 = 0.41$] and carrier phrase length [$F(3, 43) = 20.52$, $p < 0.001$, $\eta p2 = 0.20$] were observed, confirming that participants experiencing ‘sham’ TMS did not perform significantly differently from the ‘no exposure to TMS’ population.

We first assessed learning over the time-course of the task itself—‘global learning'—as the mean cumulative hit rates for listeners reporting |Color| and |Number|, fitted for each TMS condition and environment with a two-phase association model (Figure 3B and Supplemental Figure 4 and 5). Our metric for ‘global learning’ was achieved for Lecture room at: ‘sham’ TMS [40.40 ± 32.57 s] and ‘real’ TMS [63.00 ± 20.82 s]; for

Open-Plan Office: ‘sham’ [53.70 ± 30.72 s] and ‘real’ TMS [36.40 ± 37.36 s]; and for Car Park: ‘sham’ [61.40 ± 41.42 s] and ‘real' TMS [69.80 ± 37.31 s]. A Univariate ANOVA between ‘sham' TMS and ‘real' TMS exposed participants revealed no statistical differences in global learning between conditions (mean global learning ‘sham' TMS= 47.93 ± 33.34s) and mean global learning ‘real' TMS= 56.40 ± 34.85s) suggesting that participants achieved a stable level of performance (± 10% of FHR) at a similar time point in the task under both 'real' and 'sham' TMS. However, Pearson’s correlation analysis (collapsed across all environments) revealed a negative correlation between FHR and ‘global learning’ for ‘sham' TMS listeners (see Figure 3C): [$r_{(11)}$ =-0.45, p = 0.01, 95% CI [-0.70, -0.17] but not for ‘real' TMS listeners: [$r_{(10)}$ = 0.08, p = 0.69, 95% CI [-0.22, 0.37]. Specifically, participants undergoing ‘sham’ TMS who achieved stable performance (±10% of the Final Hit Rate) relatively early in the task maintained this level of performance throughout. Relative delay in attaining stable performance was associated with a lower FHR, with the time at which early ‘global learning’ is achieved predicting overall task performance. This relationship was not evident in listeners subject to ‘real’ TMS, where the time to reach a stable performance was not predictive of the magnitude of final performance (FHR). Under TMS manipulation of dlPFC, ‘global learning’ had no predictive value, indicating that task performance was less stable and reliable for participants exposed to TMS.

We next employed a Univariate analysis to compare overall performance (d’)— including FHR and False Alarm Rate, (see Figure 3D)— of ‘sham’ participants with those who received ‘real’ TMS, and observed a main effect of TMS conditions [Univariate ANOVA [F (1,501) = 26.34, p<0.001, ŋp2 = 0.1], where ‘sham’ participants showed significantly better abilities to recall |Color| and |Number| compared to participants subjected to ‘real’ TMS [*post-hoc* pairwise comparisons with Bonferroni corrections: [mean difference = 0.32, p < 0.001]. In addition, a significant interaction ‘TMS condition’ x ‘carrier phrase length’ was observed, Figure 3E: [F (3,123) = 2.82, p=0.039, ŋp2 = 0.02]. A *post-hoc* pairwise comparison with Bonferroni corrections showed that performance in the CP0 condition was not statistically different between ‘sham’ and ‘real’ TMS conditions (mean difference = 0.02, p = 0.86), however performance in the recall task (see Figure 3E and Supplemental Table 5) was better found for ‘sham’ compared to ‘real’ TMS for CP1 (mean difference = 0.34, p = 0.006),

CP2 (mean difference = 0.41, $p = 0.001$), and CP3 (mean difference = 0.50, $p < 0.001$). Listeners exposed to 'sham' TMS retained the improvements of performance as the length of carrier phrase was increased whereas those listeners receiving 'real' TMS—in which activity in dlFPC is presumably impaired—appeared unable to accumulate over time knowledge of the sound environment.

The different time courses over which statistical learning might occur suggests the involvement of multiple neural mechanisms and brain circuits (Anderson & Malmierca, 2013; Antunes & Malmierca, 2011; Robinson et al., 2016), with different cadences of learning potentially controlled, or at least modulated, by different brain centres. The time-course of midbrain neurons to adapt to different sound environments, for example, is slowed, and the capacity for retaining a 'memory' of those sound environments disappears, when cortex is inactivated through cooling (Robinson et al., 2016). This suggests feed-forward and feed-back mechanisms, with their own time-courses or time-constants, contribute to overall performance, including the rate at which sound environments are learned. We specifically wondered if 'real' TMS had a detrimental effect on short-term adaptation i.e., the observed advantage in performance for |Number| compared to |Color|, for the shortest carrier phrase CP0 (Figure 2E and 3F). A Univariate analysis revealed a main effect of 'target word', where performance for |Number| was always significantly better than for |Color| for all the environments explored and across TMS conditions [$F(1,498) = 53.53$, $p<0.001$, $\eta p2 = 0.11$]. However, no interaction between 'TMS conditions' and 'target word' was observed, indicating that in both 'real' and 'sham' TMS conditions, the same trend of better performance to |Number| compared to |Color| was evident, Figure 3E: ['sham' TMS: mean difference = 0.81, $p < 0.001$ and 'real' TMS: mean difference = 0.66, $p < 0.001$]. However, overall d' has contributions of short-term and meta-adaptive (long-term accumulation of knowledge about the environments) across the task.

To dissociate specifically short-term adaptation from longer-term meta-adaptation, we compared performance (Hit Rate) to |Number| for *initial* trials (1-2), for CP0 only, as performance for this length of carrier phrase should be influenced only by short-term accumulation of information (i.e., within trial; see Figure 3G). A non-parametric Wilcoxon signed rank test revealed no significant differences in |Number| performance for *initial* trials between 'sham' and 'real' TMS exposed listeners ($n=10$, $Z= -0.28$,

p=0.78; Figure 3H), suggesting that rapid learning of the statistical structure of the reverberant environment—in the order of a few hundreds of milliseconds—is resistant to the effects of 'real' TMS applied to dlPFC. To determine how much meta-adaptative improvement in performance could be disrupted by applying 'real' TMS to dlPFC, we then compared *initial* trials (1-2) and *steady* trials (9-10) for CP0—our proxy for meta-adaptation (Figure 3G). We tested the hypothesis that disruption of dlPFC would affect late (meta-) but not early adaptation. To this end, we expected no differences between performance in *initial* trials between 'real' and 'sham' TMS (as *initial* trials are influenced only by short-term adaptation), and a relative lack of improvement in performance in later, *steady* trials (meta-adaptation influenced) in 'real' compared to 'sham' TMS. Consistent with our hypothesis, a Wilcoxon signed rank test revealed no significant differences in performance for *initial* trials between 'sham' and 'real' TMS exposed listeners (n=10, Z= -0.21, p=0.83), but improved performance for *steady* compared to *initial* trials (n=10, Z=-2.06, p=0.04) for 'sham' TMS only (Figure 3I). These data suggest that performance in |Color| for the shortest carrier phrase, CP0, was disrupted due to impaired accumulation of knowledge about the environments encountered. In turn, performance under 'TMS' in *steady* trials (presumed to be influenced by meta-adaptation) was reminiscent of performance observed in initial trials where only immediate knowledge could be used to improve performance. This is reminiscent of adaptation to sound environments reported *in vivo* (Dean et al., 2005, 2008; Robinson et al., 2016; Wen et al., 2009), where interrupting efferent feedback by cortical cooling impairs the capacity of midbrain neurons to learn the statistical structure of sound environments; though neurons adapt to the short-term statistical structure of a sound environment each time they are exposed to it, they fail to demonstrate the acceleration of adaptation and improvement in (neural) discrimination performance as the same environment is re-encountered, adapting to it only as if exposed to it the first time.

Statistical learning of room acoustics is tuned to commonly experienced reverberation times

Reverberation is a common feature of many acoustic environments—natural and built (Traer & McDermott, 2016). Interestingly, Zahorik and Brandewie (2016) reported that improvements in speech understanding when a 'simulated room' is repeatedly encountered were best when listeners experienced moderately reverberant

environments. This suggests that brain mechanisms responsible for listening in noise might be adapted to conditions of reverberation. Francl & McDermott (2022) described an auditory model able to reproduce several ‘human-like’ spatial hearing features when successfully trained under realistic listening conditions, such as in noise and reverberation. To this end, we wondered if the learning effects we observed similarly rely on the combination of real-world environments we employed, i.e., the RT_{60}s in which listening performance was assessed. Specifically, we hypothesized that the ability to understand speech in noise depends on exposure to specific values of RT_{60} across our experimental paradigm, including those in the range commonly experienced by human listeners in natural and built environments (Traer & McDermott, 2016). To test this hypothesis, we recruited a new population of naïve participants with no previous exposure to our experimental paradigm and assessed their ability to understand speech in noise using the same CRM corpus as before, but in different combinations of acoustic environments i.e., room contexts defined by their RT_{60}s.

We first tested the ability of 10 participants to recall |Color| and |Number| in Open-Plan Office and Underground Car Park as before, but with the Lecture Room (RT_{60iso} = 0.42 s) swapped out for a more highly reflective elevator lobby (, Figure 4) with an RT_{60iso} of 1.55 s—i.e., between that of the Open-Plan Office (0.96s) and the Underground Car Park (2.42s). We therefore changed the 'room-context' in which Open-Plan Office and Underground Car Park were learned, replacing the Lecture Room with a Highly Reflective Room. When contrasting (Univariate ANOVA) the performance (hit rate) of these 10 naïve participants *vs*. the 11 subjects randomly selected from the initial 22 participants sample exposed to the initial three rooms (i.e., Lecture Room, Open-Plan Office and Underground Car Park), overall performance was significantly better in the context of the Lecture Room compared to the Highly Reflective Room (Figure 4 A-B), with a main effect of room-context [$F(1,251) = 207.49$, $p<0.001$, $\eta p2 = 0.46$], mean difference = 20.09, $p < 0.001$] and a significant interaction of room-context x room [$F(2, 125) = 26.64$, $p<0.001$, $\eta p2 = 0.18$]. *Post hoc* comparisons focused only on the overlapping rooms i.e., Underground Carpark and Open-Plan Office as it was expected that longer reverberation times, such as those in the Highly Reflective Room, have detrimental effects in speech understanding compared to shorter reverberation times i.e., Lecture Room (Houtgast & Steeneken, 1973; Knudsen, 1929; Lochner & Burger, 1961) [mean difference = 34.48, $p < 0.001$]. Performance in the Underground

Car Park was significantly better when presented in the context of the Lecture Room than in the context of the Highly Reflective Room [mean difference = 13.30, $p < 0.001$]. A similar trend of improved performance in the Open-Plan Office was observed in the context of the Lecture Room when compared to the Highly Reflective Room context: [mean difference = 12.50, $p < 0.001$].

Wondering whether the 'room-context' effect was determined by task difficulty i.e., the more-highly reverberant lobby being intrinsically more difficult for listening than the Lecture Room, we recruited a further 11 naïve participants and performed the same task in which they were exposed to another combination of three rooms, here with the Lecture Room swapped for a Living Room with a shorter RT_{60iso} (0.33 s) i.e., an expected 'easier' room. Surprisingly, although performance was improved overall relative to that in the combination of the three highly reverberant rooms (Figure 4 A-B), it remained marginally poorer [mean difference = 4.04] than when the Lecture Room was included in combination with the two fixed rooms, i.e., the Open-Plan Office and Underground Car Park (main effect of room-context [$F(1,263) = 9.32$, $p=0.002$, $\eta p2 = 0.04$].

Although reverberation might be considered detrimental to speech-in-noise performance, short-term benefits of more, compared to less, room reverberation for speech understanding have, in fact, been reported, including for the task we report here (Brandewie & Zahorik, 2010, 2013). Anechoic spaces—rooms whose walls are treated to remove completely reflected sound energy—have been extensively reported to aid speech understanding, especially in background noise. Notwithstanding this possibility, however, anechoic listening environments are rare—whether natural or built—and it is possible that listening performance in noise is indeed better when listeners have access to more commonly experienced, or ethologically realistic, levels of reverberation, a possibility suggested by the small, but significant, reduction in performance in all three rooms when the living room replaced the Lecture Room. To test this hypothesis directly, we recruited a further 10 naïve participants and compared their speech-in-noise performance in a combination of an Anechoic Room and the original Open-Plan Office and Underground Car Park, with the performance of the 11 participants in the original three rooms. Although task difficulty might be expected to be at its lowest i.e., easiest, in the less-reverberant Anechoic Room, this

was not the case. Performance was poorer when the combination of three rooms contained an Anechoic Room and the other two, more highly reverberant rooms (Figure 4 A-B), with a main effect of 'room-context' [$F(1,251) = 34.42$, $p<0.001$, $\eta p2 = 0.12$], [mean difference = 9.19, $p < 0.001$].

Seeking to explain why some rooms might be better than others for the improvement of speech understanding over time, we recalculated reverberation times for our 5 reverberant rooms according to the method of Traer and McDermott (2016) where $RT_{60traer}$ equals the median RT_{60} measured in 31 sub-bands with centre frequencies between 80 Hz and 10 kHz (see Methods). We then compared these with the values collected for a wide range (n= 271) of built and natural environments by the same authors (Figure 4C-D. $RT_{60traer}$s for 5 rooms conditions and Traer's data, plotted as $\log_{10} RT_{60traer}$ for visualisation). In particular, the $RT_{60traer}$ of the Lecture Room, at 0.49 s, lies extremely close to the median and mean $RT_{60traer}$ values of the built environments [0.42 s and 0.50 s respectively generated by the skewed distribution (n= 199), Figure 4C]. This suggests that superior performance in understanding speech understanding in noise in the Lecture Room—and the positive impact on speech understanding when the Lecture Room is included in the three-room listening task—is related to its reverberant characteristics being commonly encountered in everyday listening situations. Notably, the mean and median $RT_{60traer}$s of the recorded natural environments were much lower (0.12 s and 0.16 s, respectively [n = 72], Figure 4C) than the $RT_{60traer}$s of the Lecture Room. This suggests that brain mechanisms contributing to effective speech understanding in reverberant environments might have adapted to the range of environments experienced over some longer time course than we assessed here.

Discussion

We assessed the ability of listeners to understand speech embedded in background noise and convolved with the room impulse responses (RIRs) of different indoor environments defined by their reverberation decay time, or RT_{60}—the time taken for reverberant sound energy to decay 60 decibels (dB). We found speech understanding improved with repeated exposure to an environment, a form of learning that was impaired following continuous, bilateral theta-burst transcranial magnetic stimulation (TMS) of dorsolateral prefrontal cortex (dlPFC). Specifically, we observed rapid learning i.e., an improvement in speech understanding over the first few seconds of room exposure, likely due to listeners learning the acoustic reverberation—the only variable held constant across trials for a given environment. This learning, on the timescale of several seconds, was impaired by TMS, whilst learning at shorter or longer timescales was not, suggesting that TMS applied to dlPFC specifically disrupted learning of the reverberant characteristics of an environment. Listeners also showed better listening performance in moderately reverberant environments, and this performance was transferable between different acoustic environments. Specifically, the ability to correctly report keywords spoken in environments with the more extreme—lower or higher—RIRs was best when these environments were encountered in experimental blocks also containing the moderately reverberant Lecture Room or Living Room. This transference suggests an ethological tuning to more commonly encountered environments when learning acoustic backgrounds.

A role for dlPFC in statistical learning of room acoustics

Inactivation of dlPFC using theta-burst TMS reduced overall performance in our speech-in-noise task, assessed in terms of the hit rate for |Color| and |Number| in CRM phrases as a function of duration of exposure to a sound environment. Notably, however, TMS did not impair listening performance for the shortest-duration phrase, '|Color| |Number| now'. For this duration of carrier phrase, overall performance, and the tendency to report more accurately |Number| compared to |Color|—despite |Number| having twice as many options as |Color|—was unaffected by TMS stimulation of dlPFC. The robustness of listening performance to the shortest carrier phrase suggests that the ability to hear out speech in background noise and potentially to learn this over time in complex acoustic environments does not rely solely on cortical feedback from dlPFC. The ability to leverage performance on |Number| by exposure

to |Color| presented in background noise convolved with the room impulse response is evident even when the function of dlPFC is (presumably) impaired. It is the case, however, that the rate of learning—i.e., how rapidly performance improved with increasing time of exposure to an acoustic environment—was influenced when dlPFC was impaired. Under the influence of prior TMS stimulation, listeners were less able, and less reliable when learning the acoustic background features of an environment to enhance their speech understanding, particularly in the more challenging environments of the Underground Car Park and Open-Plan Office.

Dense connections between dlPFC and the sensory cortices provide the conduit for prediction-based, top-down regulation (Morrone, 2010). These model-based predictions rely on learned templates of sensory environments (Alexander & Brown, 2018), which are critical for making inferences under situations of sensory uncertainty such as those encountered in noisy and challenging listening environments (Bartolo & Averbeck, 2021). Dorsolateral prefrontal cortex can influence auditory processing by means of direct connections to auditory cortex (Hackett et al., 1999; Plakke & Romanski, 2014)—see Figure 3Ai—and to the auditory thalamic reticular nucleus, a key modulator of auditory thalamo-cortical loops (Zikopoulos & Barbas, 2006). In return, dlPFC receives projections from primary and secondary auditory cortex (Barbas & Pandya, 1987; Barbas & Pandya, 1991; Goldman-Rakic & Schwartz, 1982; Pandya & Barnes, 2019; Petrides & Pandya, 2002), indicating that a reciprocal auditory-to-prefrontal 'listening loop' exists to support hierarchical predictions and prediction-error feedback. These 'listening loops' could potentially extend to cortical efferent feedback (Blackwell et al., 2020; McAlpine & de Hoz, 2023) and modulate the function of the the inner-ear sensitivity in attended speech-in-noise tasks (de Boer & Thornton, 2008; Garinis et al., 2011; Giraud et al., 1997; Mishra & Lutman, 2014; Hernández-Pérez et al., 2021).

The specific mechanism by which TMS-induced modulation of dlPFC impairs listening performance remains unknown. Neurostimulation studies targeting dlPFC during learning paradigms have generated conflicting directions of effect i.e., stimulation of dlPFC is reported to impair implicit learning (Nydam et al., 2018; Pascual-Leone et al., 1996) or enhance it (Ambrus et al., 2020). Ambrus and colleagues argued that impairment of dlPFC during implicit/statistical learning allows for increased

engagement of a learning mechanism that is ‘model-free’ i.e., does not rely on inherent sensory representations and therefore takes longer to consolidate. Here, we speculate that the short periods over which learning is required in our task are subserved by a pre-frontally mediated, ‘model-based’ learning able to construct predictions ‘on the fly’ (Daw et al., 2005). We suggest, in our study, that disrupting a rapidly acting model-based form of learning supported by dlPFC, as part of the auditory-prefrontal ‘listening loop’, leads to impaired learning of reverberant environments.

Dorsolateral prefrontal cortex is also implicated in several, high-level cognitive functions including multi-sensory integration (Fuster et al., 2000), executive functions such as inhibition and working memory (Castro-Meneses et al., 2016; Coltheart et al., 2018; Wang et al., 2015) and listening in noisy environments (Du et al., 2016). The impact of TMS on the overall performance of our speech-in-noise task may be linked to the involvement of dlPFC in more generalized cognitive functions such as executive, memory, and attention processes. Dissociating the potential influence of TMS on such processes relative to a specific role in speech-in-noise understanding in noisy, reverberant environments would be an ideal next step. Specifically frontal regions have been linked to sensorimotor integration during speech-in-noise perception in young and older human participants (Du et al., 2016), suggesting that TMS might disrupt global functioning required for the analysis and understanding of speech. Whilst not entirely unrelated, all participants completed a familiarization task preceding the main experiment, which consisted of 10 trials presented in anechoic conditions at 0 dB Signal-to-Noise Ratio (SNR). Participants exposed to either ‘real’ or ‘sham’ TMS underwent this task following cTBS procedures. Notably, all participants, including those subjected to cTBS, achieved 75-100% accuracy during the familiarization phase, with no statistical differences between TMS groups [mean difference=1.25, $t_{(1, 9)}$=0.43, p=0.68, d=0.14], confirming their comprehension of the task and successful procedural learning (i.e. learning the task *per se*). In addition, speech understanding in the most challenging condition (CP0) was not disrupted following TMS manipulations, suggesting that the effect of TMS was specifically in terms of influencing the way these environments were learnt over time. Nevertheless, we acknowledge that the primary task may have imposed greater cognitive demands, potentially requiring a more significant involvement of the dlPFC to meet such task requirements.

What is being learned in statistical learning of acoustic features?

The need to communicate in reverberant environments is common to daily life. It is well known that long reverberation times have detrimental effects on speech quality and intelligibility (Brandewie & Zahorik, 2010, 2013; Srinivasan & Zahorik, 2012; Zahorik & Brandewie, 2016) and that, specifically, reverberation blurs phoneme boundaries, increasing the extent to which similar words might be confused, e.g., 'sir'/'stir' (Watkins, 2005b; Watkins & Makin, 2007). However, with sufficient exposure to longer reverberation times in the form of a carrier phrase, similar levels of word identification to those achieved in minimal reverberation are observed (Watkins, 2005b; Watkins & Makin, 2007). Based on the premise that providing sufficient contextual information enables the auditory system to adapt and compensate for the detrimental effects of reverberation, Zahorik and colleagues (Brandewie & Zahorik, 2010, 2013; Srinivasan & Zahorik, 2012; Zahorik & Brandewie, 2016) and, here, ourselves, demonstrate that prior exposure to the reverberant characteristics of a room can enhance sentences understanding when that environment is re-encountered. Moreover, Zahorik and colleagues, and our own data (see Figure 4A), demonstrate that improvements in speech understanding are always greater in reverberant compared to anechoic, environments (Brandewie & Zahorik, 2010, 2013; Srinivasan & Zahorik, 2012), i.e., improved performance arises only through exposure to a reverberant environment rather than to the speech material *per se*.

Adaptation to continuous noise is a well-known phenomenon in the auditory system (Costalupes et al., 1984; Gibson et al., 1985; Phillips, 1985; Rees & Palmer, 1988) that could potentially also explain the speech improvements observed among carrier lengths. This is particularly relevant in our experiments because the masking noise preceded the onset of the speech material, potentially providing a window for noise-adaptation to occur (Ainsworth & Meyer, 1994; Ben-David et al., 2016). However, Zahorik and Brandewie reported similar effect sizes in speech enhancements when 1s (Brandewie & Zahorik, 2010) or 150 ms (Brandewie & Zahorik, 2013) of white noise was presented prior to the carrier phrase (i.e., CP0—|Color| |Number|). From this the authors concluded that noise alone did not convey sufficient information about the acoustic environment to elicit improvements in speech understanding.

Adult listeners can also acclimatise to speech acoustics that deviate from the norm or from long-term language regularities such as dialects and foreign accents, with sufficient exposure (Idemaru & Holt, 2011, 2014; Liu & Holt, 2015). Even if our participants were all Australian-English native speakers exposed to an American-English speech corpus—the CRM—some perceptual learning of individual talker's idiosyncratic speech patterns might arise over the course of the task (Choi & Perrachione, 2019; Liu & Holt, 2015; Stilp, 2020). In addition, it has been reported that 'target voice continuity' i.e., the same talker within trials—similar to our design—can enhance the build-up of selective attention and improve listeners abilities to report correctly digits sequences in the presence of other talkers (Best et al., 2008). However, when the room build-up effect—the effect assessed here—is explored in an isolated anechoic space i.e., not intermingled with room acoustic of other reverberant spaces, there is little or no improvement in speech understanding with increasing exposure to the talker *per se*; i.e., in the form of an increasing length of carrier phrase (Brandewie & Zahorik, 2010, 2013). Interrupting the continuity of the room—it's reverberation profile—rather than the continuity of the target talker, is the factor reported to disrupt significantly the improvement in speech understanding with increasing exposure (Brandewie & Zahorik, 2018). If listeners adapted to a talker in the course of our experiment, this occurred rapidly, within a few speech tokens (Cooke et al., 2022; Kakehi, 1992; Kato & Kakehi, 1988) and contributed little to the enhancement of speech understanding we observed in reverberant rooms.

A key feature of our study, one that distinguishes it from previous assessments of statistical learning of acoustic features in human listeners, is our use of an ethologically valid listening task—understanding speech in background noise. Given the perceptual biases inherent in sensory processing, it is notable that investigators have previously reported different noise tokens to be differently learnable (Agus et al., 2014; Daikhin et al., 2017), for example, and that textures must be carefully controlled to ensure listeners are not exploiting subtle spectro-temporal features in their judgments of statistical similarity (McDermott et al., 2013). The potential for listeners to hear out specific—potentially unique to them—spectro-temporal features of otherwise statistically identical sound tokens, or to make judgments on the regularity or similarity of tone sequences based on unique listening experiences (Barascud et al., 2016; Bianco et al., 2020), presents a potential confound for these studies. Our study

countermands this problem by actively exploiting the propensity for human listeners to ascribe meaning to spectro-temporal fluctuations (here, speech) in acoustic waveforms (Brandewie & Zahorik, 2010, 2013). Though still targeting the learning of background acoustic features, it engages listeners in a highly relevant listening task, one for which humans have rapidly evolved—i.e., understanding speech in background noise. Listeners' attention, therefore, was called away from the background acoustic features to the ethologically relevant foreground task of attending to human speech, void of semantics; additionally, length of phrase, talker, and |Callsign| were uninformative to |Color| and |Number| in our utterances, as were |Color| and |Number| to each other. This misdirection allows us to exploit speech as a 'reporter' or biomarker for statistical learning of background, more-abstract, acoustic features, and without the potential confounding factor of (individualised) perceptual bias. Further, we have demonstrated that prior exposure to ethologically relevant environments matters when learning less-commonly encountered acoustic scenes.

The impact on speech understanding of less-ecologically realistic reverberant environments is reminiscent of the detrimental effect of TMS, with lower overall listening performance. This suggests that learning the statistical structure of sound environments, the better to understand speech in reverberant background noise, requires some longer form of memory (experience, developmental, or evolutionary). Brain circuits, therefore, might act to improve listening performance, operating best when at least some of the exposure to listening environments includes plausible, and commonly experienced, reverberation times (Traer & McDermott, 2016). When the acoustics of commonly encountered rooms were replaced with those of other environments having higher or lower amounts of reverberation, performance declined over the course of an experimental session. Intriguingly, the most effective reverberation time we employed, at least in terms of its capacity to be learned and potentially from which learning could be transferred—the $RT_{60traer}$ of 0.49 s of the small Lecture Room is very close to the median and mean $RT_{60traer}$ of 0.42 and 0.50 s previously recorded for a wide range of built environments (indoors and outdoors). This suggests that statistical learning of acoustic background features might be tuned to ethologically relevant environments. Alternatively, built environments might be constructed to generate reverberation subjectively most suited to maximising speech-in-noise understanding, though we are unaware as to whether such a constraint is

consciously applied in the design of listening spaces given contingencies such as the range and relative consistency and constancy of potential contents of any given space that influence its reverberant characteristics. Compared to built environments, however, the $RT_{60traer}$s of natural environments were much lower than all the rooms we assessed, save for that of the anechoic room, which was suboptimal in terms of listening performance.

Neural adaptation as a contributing mechanism to the learning of room acoustics

Our data are consistent with improved performance in a relevant 'foreground' task emerging as the brain adapts to the statistics of background features of the listening environment. Adaptation to stimulus statistics has been proposed as a neurophysiological mechanism underlying selective suppression of background noise in auditory cortex (Fuglsang et al., 2017; Kell & McDermott, 2019; Khalighinejad et al., 2019; Mesgarani et al., 2014). More recently, it has been shown that spectro-temporal receptive fields of auditory-cortical neurons are sensitive to the RT_{60}s of reverberant noise i.e., a form of adaptation specific to reverberation and consistent with a de-reverberation of the environment encountered (Ivanov et al., 2022). Cortical adaptation to reverberation has been also reported in awake listeners (Fuglsang et al., 2017; Mesgarani et al., 2014) and to this end, our data suggest a role for top-down mechanisms in improving over time an attended task—recalling spoken words—in reverberant environments. It is therefore possible that goal-directed behaviours and the feedback the auditory cortex receives form areas such as dlPFC may contribute to fine-tuning the adaptation to reverberant environments.

Our data are also consistent with reports of midbrain auditory neurons recorded *in vivo* adapting to the statistical structure of evolving sound environments over the course of several hundred milliseconds (Dean et al., 2005, 2008) and this capacity for adaptation increases with repeated exposures to the same statistically structured environment. Importantly, whilst rapid adaptation is retained when descending cortical influences are disrupted (through cortical cooling), the longer-term learning effect—a speeding up of adaptation referred to as meta-adaptation—is abolished (Robinson et al., 2016). As with auditory neurons recorded *in vivo* (Bakay et al., 2018; Dean et al., 2005, 2008; Ivanov et al., 2022), it seems listeners might adapt to the statistical structure of a sound

environment within a few hundred milliseconds of exposure to enhance performance in a listening task, an initial, rapid, learning phase impervious to inactivation of the dorsolateral prefrontal cortex by theta-burst TMS. The nested set of temporal sensitivities to room acoustics we observe is consistent with features of statistical learning that emerges from the level of the auditory nerve to primary cortex *in vivo* (Dean et al., 2005, 2008; Watkins & Barbour, 2008; Wen et al., 2009), and likely modulated by efferent influences that span the entire auditory pathway all the way to the sensory receptors of the inner ear (Hernández-Pérez et al., 2021; Perrot et al., 2006; Terreros & Delano, 2015) and even the mechanical sensitivity to sound of the eardrum and ossicles (middle-ear bones) (Gruters et al., 2018).

Dissociating auditory cortical feedback circuits to the midbrain impairs the capacity of midbrain neurons to 'recall' previous experience of that environment (Bajo et al., 2019; Robinson et al., 2016). Like the performance in our listening tasks, these neurons were still able to adapt rapidly (within hundreds of milliseconds) each time an environment was encountered as if for the first time, but showed no capacity to exploit a memory of that environment to improve neural coding (i.e., meta-adaptation) (Robinson et al., 2016). Meta-adaptation was consistent with neurons in lower brain centres adapting to the current sound environment, and those in higher brain centres learning the longer-term statistical structure of changing environments. Once higher brain centres learnt the experienced environment, this information, conveyed to early brain centres, ensures they were 'primed' for coding an environment when it is re-encountered.

Inclusion and Ethics statement

This study was approved by the Human Research Ethics Committee of Macquarie University (ref: 5201833344874). Each participant signed a written informed consent form and was given a small financial remuneration for their time.

Acknowledgements

The study was supported by the Australian Research Council (DP180102524 and FL 160100108 awarded to D.M.). The authors would like to thank Jörg Bulcholz and Javier Badajoz-Davila for their assistance with the spatialization and sound field simulation of the acoustic stimuli. They would also like to thank Kurt Shulver for his assistance during rTMS manipulations. We sincerely thank Yuranny Cabral-Calderin for her valuable insights and constructive feedback on improving this manuscript.

Authors' contributions

Conceptualization, H.H.P., D.M., J.J.M.M. and P.F.S. Methodology, H.H.P., D.M., J.J.M.M. and P.F.S. Investigation, H.H.P. Software, J.J.M.M., J.M-H. and J.T. Formal Analysis, H.H.P., J.M-H. and J.T. Visualization, H.H.P., J.M-H. and J.T. Writing – Original Draft, H.H.P. and D.M. Writing – Review & Editing, H.H.P., D.M., J.J.M.M., P.F.S., J.M-H. and J.T.

Declaration of interests

The authors declare no competing interests

References

Agus, T. R., Carrión-Castillo, A., Pressnitzer, D., & Ramus, F. (2014). Perceptual Learning of Acoustic Noise by Individuals With Dyslexia. *Journal of Speech, Language, and Hearing Research: JSLHR*, *57*(3), 1069–1077. https://doi.org/10.1044/1092-4388(2013/13-0020)

Ainsworth, W. A., & Meyer, G. F. (1994). Recognition of plosive syllables in noise: comparison of an auditory model with human performance. *The Journal of the Acoustical Society of America*, *96*(2 Pt 1), 687–694. https://doi.org/10.1121/1.410306

Alexander, W. H., & Brown, J. W. (2018). Frontal cortex function as derived from hierarchical predictive coding. *Scientific Reports*, *8*(1), 3843. https://doi.org/10.1038/s41598-018-21407-9

Ambrus, G. G., Vékony, T., Janacsek, K., Trimborn, A. B. C., Kovács, G., & Nemeth, D. (2020). When less is more: Enhanced statistical learning of non-adjacent dependencies after disruption of bilateral DLPFC. *Journal of Memory and Language*, *114*, 104144.

Anderson, L. A., & Malmierca, M. S. (2013). The effect of auditory cortex deactivation on stimulus-specific adaptation in the inferior colliculus of the rat. *The European Journal of Neuroscience*, *37*(1), 52–62. https://doi.org/10.1111/ejn.12018

Antunes, F. M., & Malmierca, M. S. (2011). Effect of auditory cortex deactivation on stimulus-specific adaptation in the medial geniculate body. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(47), 17306–17316. https://doi.org/10.1523/JNEUROSCI.1915-11.2011

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of Conditional Probability Statistics by 8-Month-Old Infants. *Psychological Science*, *9*(4), 321–324. https://doi.org/10.1111/1467-9280.00063

Badajoz-Davila, J., Buchholz, J. M., & Van-Hoesel, R. (2020). Effect of noise and reverberation on speech intelligibility for cochlear implant recipients in realistic sound environments. *The Journal of the Acoustical Society of America*, *147*(5), 3538. https://doi.org/10.1121/10.0001259

Bajo, V. M., Nodal, F. R., Korn, C., Constantinescu, A. O., Mann, E. O., Boyden, E. S., 3rd, & King, A. J. (2019). Silencing cortical activity during sound-localization training impairs auditory perceptual learning. *Nature Communications*, *10*(1), 3075. https://doi.org/10.1038/s41467-019-10770-4

Bakay, W. M. H., Anderson, L. A., Garcia-Lazaro, J. A., McAlpine, D., & Schaette, R. (2018). Hidden hearing loss selectively impairs neural adaptation to loud sound environments. *Nature Communications*, *9*(1), 4298. https://doi.org/10.1038/s41467-018-06777-y

Barascud, N., Pearce, M. T., Griffiths, T. D., Friston, K. J., & Chait, M. (2016). Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. *Proceedings of the National Academy of Sciences*, *113*(5), E616–E625. https://doi.org/10.1073/pnas.1508523113

Barbas, H., & Pandya, D. N. (1987). Architecture and frontal cortical connections of the premotor cortex (area 6) in the rhesus monkey. *The Journal of Comparative Neurology*, *256*(2), 211–228. https://doi.org/10.1002/cne.902560203

Barbas, Helen, & Pandya, D. N. (1991). Patterns of connections of the prefrontal cortex in the rhesus monkey associated with cortical architecture. In H. S. Levin (Ed.), *Frontal lobe function and dysfunction , (pp* (Vol. 427, pp. 35–58). Oxford University Press, xv. https://psycnet.apa.org/fulltext/1992-97203-002.pdf

Barberis, N. C. (2013). Thirty Years of Prospect Theory in Economics: A Review and Assessment. *The Journal of Economic Perspectives: A Journal of the American Economic Association*, *27*(1), 173–196. https://doi.org/10.1257/jep.27.1.173

Bartolo, R., & Averbeck, B. B. (2021). Inference as a fundamental process in behavior. *Current Opinion in Behavioral Sciences*, *38*, 8–13. https://doi.org/10.1016/j.cobeha.2020.06.005

Ben-David, B. M., Avivi-Reich, M., & Schneider, B. A. (2016). Does the degree of linguistic experience (native versus nonnative) modulate the degree to which listeners can benefit from a delay between the onset of the maskers and the onset of the target speech? *Hearing Research*, *341*, 9–18. https://doi.org/10.1016/j.heares.2016.07.016

Best, V., Ozmeral, E. J., Kopco, N., & Shinn-Cunningham, B. G. (2008). Object continuity enhances selective auditory attention. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(35), 13174–13178. https://doi.org/10.1073/pnas.0803718105

Bianco, R., Harrison, P. M. C., Hu, M., Bolger, C., Picken, S., Pearce, M. T., & Chait, M. (2020). Long-term implicit memory for sequential auditory patterns in humans. *ELife*, *9*, e56073. https://doi.org/10.7554/eLife.56073

Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., & Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience*, *7*(3), 295–301. https://doi.org/10.1038/nn1198

Blackwell, J. M., Lesicko, A. M., Rao, W., De Biasi, M., & Geffen, M. N. (2020). Auditory cortex shapes sound responses in the inferior colliculus. *ELife*, *9*. https://doi.org/10.7554/eLife.51890

Blesser, B., & Salter, L.-R. (2009). *Spaces Speak, Are You Listening?: Experiencing Aural Architecture*. MIT Press. https://play.google.com/store/books/details?id=5aY1nrVTAZIC

Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D. (2000). A speech corpus for multitalker communications research. *The Journal of the Acoustical Society of America*, *107*(2), 1065–1066. https://doi.org/10.1121/1.428288

Bradley, J. S., Reich, R. D., & Norcross, S. G. (1999). On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility. *The Journal of the Acoustical Society of America*, *106*(4 Pt 1), 1820–1828. https://doi.org/10.1121/1.427932

Brainard, M. S., & Doupe, A. J. (2002). What songbirds teach us about learning. *Nature*, *417*(6886), 351–358. https://doi.org/10.1038/417351a

Brandewie, E. J., & Zahorik, P. (2018). Speech intelligibility in rooms: Disrupting the effect of prior listening exposure. *The Journal of the Acoustical Society of America*, *143*(5), 3068. https://doi.org/10.1121/1.5038278

Brandewie, E., & Zahorik, P. (2010). Prior listening in rooms improves speech intelligibility. *The Journal of the Acoustical Society of America*, *128*(1), 291–299. https://doi.org/10.1121/1.3436565

Brandewie, E., & Zahorik, P. (2011). Adaptation to Room Acoustics Using the Modified Rhyme Test. *Proceedings of Meetings on Acoustics Acoustical Society of America*, *129*(4), 2487. https://doi.org/10.1121/1.3588198

Brandewie, E., & Zahorik, P. (2013). Time course of a perceptual enhancement effect for noise-masked speech in reverberant environments. *The Journal of the Acoustical Society of America*, *134*(2), EL265-70. https://doi.org/10.1121/1.4816263

Bregman, A. S. (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press. https://play.google.com/store/books/details?id=jI8muSpAC5AC

Bronkhorst, A. W., & Houtgast, T. (1999). Auditory distance perception in rooms. *Nature*, *397*(6719), 517–520. https://doi.org/10.1038/17374

Brumm, H., & Naguib, M. (2009). Chapter 1 Environmental Acoustics and the Evolution of Bird Song. In *Advances in the Study of Behavior* (Vol. 40, pp. 1–33). Academic Press. https://doi.org/10.1016/S0065-3454(09)40001-9

Cabrera, D., Jeong, C., Kwak, H. J., & Kim, J.-Y. (2005). Auditory room size perception for modeled and measured rooms. *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, *2005*, 2995–3004. https://www.academia.edu/download/41945460/Auditory_room_size_perception_for_modele20160203-29142-1gyb08z.pdf

Castro-Meneses, L. J., Johnson, B. W., & Sowman, P. F. (2016). Vocal response inhibition is enhanced by anodal tDCS over the right prefrontal cortex. *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale*, *234*(1), 185–195. https://doi.org/10.1007/s00221-015-4452-0

Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *The Journal of the Acoustical Society of America*, *25*(5), 975–979. https://doi.org/10.1121/1.1907229

Choi, J. Y., & Perrachione, T. K. (2019). Time and information in perceptual adaptation to speech. *Cognition*, *192*, 103982. https://doi.org/10.1016/j.cognition.2019.05.019

Coltheart, M., Cox, R., Sowman, P., Morgan, H., Barnier, A., Langdon, R., Connaughton, E., Teichmann, L., Williams, N., & Polito, V. (2018). Belief, delusion, hypnosis, and the right dorsolateral prefrontal cortex: A transcranial magnetic stimulation study. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *101*, 234–248. https://doi.org/10.1016/j.cortex.2018.01.001

Conway, C. M. (2020). How does the brain learn environmental structure? Ten core principles for understanding the neurocognitive mechanisms of statistical learning. *Neuroscience and Biobehavioral Reviews*, *112*, 279–299. https://doi.org/10.1016/j.neubiorev.2020.01.032

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: word predictability is the key. *Cognition*, *114*(3), 356–371. https://doi.org/10.1016/j.cognition.2009.10.009

Conway, C. M., & Christiansen, M. H. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences*, *5*(12), 539–546. https://doi.org/10.1016/s1364-6613(00)01800-3

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(1), 24–39. https://doi.org/10.1037/0278-7393.31.1.24

Cooke, M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America*, *119*(3), 1562–1573. https://doi.org/10.1121/1.2166600

Cooke, M., Scharenborg, O., & Meyer, B. T. (2022). The time course of adaptation to distorted speech. *The Journal of the Acoustical Society of America*, *151*(4), 2636. https://doi.org/10.1121/10.0010235

Costalupes, J. A., Young, E. D., & Gibson, D. J. (1984). Effects of continuous noise backgrounds on rate response of auditory nerve fibers in cat. *Journal of Neurophysiology*, *51*(6), 1326–1344. https://doi.org/10.1152/jn.1984.51.6.1326

Culling, J. F., Hodder, K. I., & Toh, C. Y. (2003). Effects of reverberation on perceptual segregation of competing voices. *The Journal of the Acoustical Society of America*, *114*(5), 2871–2876. https://doi.org/10.1121/1.1616922

Daikhin, L., Raviv, O., & Ahissar, M. (2017). Auditory Stimulus Processing and Task Learning Are Adequate in Dyslexia, but Benefits From Regularities Are Reduced. *Journal of Speech, Language, and Hearing Research: JSLHR*, *60*(2), 471–479. https://doi.org/10.1044/2016_JSLHR-H-16-0114

Daikoku, T., & Yumoto, M. (2023). Order of statistical learning depends on perceptive uncertainty. *Current Research in Neurobiology*, *4*(100080), 100080. https://doi.org/10.1016/j.crneur.2023.100080

Davis, M. H., Ford, M. A., Kherif, F., & Johnsrude, I. S. (2011). Does semantic context benefit speech understanding through “top--down” processes? Evidence from time-resolved sparse fMRI. *Journal of Cognitive Neuroscience*, *23*(12), 3914–3932. https://direct.mit.edu/jocn/article-abstract/23/12/3914/5277

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*(12), 1704–1711. https://doi.org/10.1038/nn1560

de Boer, J., & Thornton, A. R. D. (2008). Neural correlates of perceptual learning in the auditory brainstem: efferent activity predicts and reflects improvement at a speech-in-noise discrimination task. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *28*(19), 4929–4937.

Dean, I., Harper, N. S., & McAlpine, D. (2005). Neural population coding of sound level adapts to stimulus statistics. *Nature Neuroscience*, *8*, 1684. https://doi.org/10.1038/nn1541

Dean, I., Robinson, B. L., Harper, N. S., & McAlpine, D. (2008). Rapid Neural Adaptation to Sound Level Statistics. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *28*(25), 6430–6438. https://doi.org/10.1523/jneurosci.0470-08.2008

Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2016). Increased activity in frontal motor cortex compensates impaired speech perception in older adults. *Nature Communications*, *7*, 12241. https://doi.org/10.1038/ncomms12241

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/bf03193146

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499–504. https://doi.org/10.1111/1467-9280.00392

Fuglsang, S. A., Dau, T., & Hjortkjær, J. (2017). Noise-robust cortical tracking of attended speech in real-world acoustic scenes. *NeuroImage*, *156*, 435–444. https://doi.org/10.1016/j.neuroimage.2017.04.026

Fuster, J. M., Bodner, M., & Kroger, J. K. (2000). Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature*, *405*(6784), 347–351. https://doi.org/10.1038/35012613

Gamboa, O. L., Antal, A., Moliadze, V., & Paulus, W. (2010). Simply longer is not better: reversal of theta burst after-effect with prolonged stimulation. *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale*, *204*(2), 181–187. https://doi.org/10.1007/s00221-010-2293-4

Gariépy, J.-F., Watson, K. K., Du, E., Xie, D. L., Erb, J., Amasino, D., & Platt, M. L. (2014). Social learning in humans and other animals. *Frontiers in Neuroscience*, *8*, 58. https://doi.org/10.3389/fnins.2014.00058

Garinis, A. C., Glattke, T., & Cone, B. K. (2011). The MOC reflex during active listening to speech. *Journal of Speech, Language, and Hearing Research: JSLHR*, *54*(5), 1464–1476.

Gibson, D. J., Young, E. D., & Costalupes, J. A. (1985). Similarity of dynamic range adjustment in auditory nerve and cochlear nuclei. *Journal of Neurophysiology*, *53*(4), 940–958. https://doi.org/10.1152/jn.1985.53.4.940

Giraud, A. L., Garnier, S., Micheyl, C., Lina, G., Chays, A., & Chéry-Croze, S. (1997). Auditory efferents involved in speech-in-noise intelligibility. *Neuroreport*, *8*(7), 1779–1783.

Glasberg, B. R., & Moore, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, *47*(1–2), 103–138. https://doi.org/10.1016/0378-5955(90)90170-t

Goldman-Rakic, P. S., & Schwartz, M. L. (1982). Interdigitation of contralateral and ipsilateral columnar projections to frontal association cortex in primates. *Science*, *216*(4547), 755–757. https://doi.org/10.1126/science.6177037

Gruters, K. G., Murphy, D. L. K., Jenson, C. D., Smith, D. W., Shera, C. A., & Groh, J. M. (2018). The eardrums move when the eyes move: A multisensory effect on the mechanics of hearing. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(6), E1309–E1318. https://doi.org/10.1073/pnas.1717948115

Hackett, T. A., Stepniewska, I., & Kaas, J. H. (1999). Prefrontal connections of the parabelt auditory cortex in macaque monkeys. *Brain Research*, *817*(1–2), 45–58. https://doi.org/10.1016/s0006-8993(98)01182-2

Hawley, M. L., Litovsky, R. Y., & Culling, J. F. (2004). The benefit of binaural hearing in a cocktail party: effect of location and type of interferer. *The Journal of the Acoustical Society of America*, *115*(2), 833–843. https://doi.org/10.1121/1.1639908

Hernández-Pérez, H., Mikiel-Hunter, J., McAlpine, D., Dhar, S., Boothalingam, S., Monaghan, J. J. M., & McMahon, C. M. (2021). Understanding degraded speech leads to perceptual gating of a brainstem reflex in human listeners. *PLoS Biology*, *19*(10), e3001439. https://doi.org/10.1371/journal.pbio.3001439

Hertrich, I., Dietrich, S., Blum, C., & Ackermann, H. (2021). The role of the dorsolateral prefrontal cortex for speech and language processing. *Frontiers in Human Neuroscience*, *15*, 645209. https://doi.org/10.3389/fnhum.2021.645209

Herwig, U., Satrapi, P., & Schönfeldt-Lecuona, C. (2003). Using the international 10-20 EEG system for positioning of transcranial magnetic stimulation. *Brain Topography*, *16*(2), 95–99. https://link.springer.com/content/pdf/10.1023%2FB%3ABRAT.0000006333.93597.9d.pdf

Hicks, J. M., & McDermott, J. H. (2024). Noise schemas aid hearing in noise. *Proceedings of the National Academy of Sciences of the United States of America*, *121*(47). https://doi.org/10.1073/pnas.2408995121

Hoogendam, J. M., Ramakers, G. M. J., & Di Lazzaro, V. (2010). Physiology of repetitive transcranial magnetic stimulation of the human brain. *Brain Stimulation*, *3*(2), 95–118. https://doi.org/10.1016/j.brs.2009.10.005

Houtgast, T., & Steeneken, H. J. M. (1985). A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *The Journal of the Acoustical Society of America*, *77*(3), 1069–1077. https://doi.org/10.1121/1.392224

Houtgast, To, & Steeneken, H. J. (1973). The modulation transfer function in room acoustics as a predictor of speech intelligibility. *Acta Acustica United with Acustica*, *28*(1), 66–73. https://www.ingentaconnect.com/content/dav/aaua/1973/00000028/00000001/art00012

Huang, Y.-Z., Edwards, M. J., Rounis, E., Bhatia, K. P., & Rothwell, J. C. (2005). Theta Burst Stimulation of the Human Motor Cortex. *Neuron*, *45*(2), 201–206. https://doi.org/10.1016/j.neuron.2004.12.033

Hughson, W., & Westlake, H. (1944). Manual for program outline for rehabilitation of aural casualties both military and civilian. *Transactions - American Academy of Ophthalmology and Otolaryngology. American Academy of Ophthalmology and Otolaryngology*.

Huyck, J. J., & Wright, B. A. (2011). Late maturation of auditory perceptual learning. *Developmental Science*, *14*(3), 614–621. https://doi.org/10.1111/j.1467-7687.2010.01009.x

Idemaru, K., & Holt, L. L. (2011). Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology. Human Perception and Performance*, *37*(6), 1939–1956. https://doi.org/10.1037/a0025641

Idemaru, K., & Holt, L. L. (2014). Specificity of dimension-based statistical learning in word recognition. *Journal of Experimental Psychology. Human Perception and Performance*, *40*(3), 1009–1021. https://doi.org/10.1037/a0035269

Ivanov, A. Z., King, A. J., Willmore, B. D. B., Walker, K. M. M., & Harper, N. S. (2022). Cortical adaptation to sound reverberation. *ELife*, *11*. https://doi.org/10.7554/eLife.75090

Jurcak, V., Okamoto, M., Singh, A., & Dan, I. (2005). Virtual 10–20 measurement on MR images for inter-modal linking of transcranial and tomographic neuroimaging methods. *NeuroImage*, *26*(4), 1184–1192. https://ac.els-cdn.com/S1053811905001862/1-s2.0-S1053811905001862-main.pdf?_tid=87d3462c-b7fa-4780-b7b0-289d930895bd&acdnat=1544065940_494439b3ca37d9077258347769daf7e0

Kakehi, K. (1992). Adaptability to differences between talkers in Japanese monosyllabic perception. *Speech Perception, Speech Production, and Linguistic Structure*, 135–142. https://books.google.com.au/books?hl=en&lr=&id=eYrOR7VHnnQC&oi=fnd&pg=PA135&dq=Kakehi,+K.+(1992).+%E2%80%9CAdaptability+to+differences+between+talkers+in+Japanese+monosyllabic+perception,%E2%80%9D+in+Speech+Perception,+Speech+Production+and+Linguistic+Structure,+edited+by+Y.+Tohkura,+E.+Vatikiotis-Bateson,+and+Y.+Sagisaka+(OHM,+Tokyo),+pp.+135&ots=p9p7gXYBWu&sig=GPeZemflBK64RT6sNEwoGteaYXw

Kato, K., & Kakehi, K. (1988). Listener adaptability to individual speaker differences in monosyllabic speech perception. *J. Acoust. Soc. Jpn*, *44*(3), 180–186.

Kell, A. J. E., & McDermott, J. H. (2019). Invariance to background noise as a signature of non-primary auditory cortex. *Nature Communications*, *10*(1), 3958. https://doi.org/10.1038/s41467-019-11710-y

Khalighinejad, B., Herrero, J. L., Mehta, A. D., & Mesgarani, N. (2019). Adaptation of the human auditory cortex to changing background noise. *Nature Communications*, *10*(1), 2509. https://doi.org/10.1038/s41467-019-10611-4

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35-42. https://doi.org/10.1016/s0010-0277(02)00004-5

Knudsen, V. O. (1929). THE HEARING OF SPEECH IN AUDITORIUMS. *The Journal of the Acoustical Society of America*, *1*(1), 56–82. https://doi.org/10.1121/1.1901470

Kolarik, A. J., Moore, B. C. J., Cirstea, S., Aggius-Vella, E., Gori, M., Campus, C., & Pardhan, S. (2021). Factors Affecting Auditory Estimates of Virtual Room Size: Effects of Stimulus, Level, and Reverberation. *Perception*, *50*(7), 646–663. https://doi.org/10.1177/03010066211020598

Larsby, B., Hällgren, M., Nilsson, L., & McAllister, A. (2015). The influence of female versus male speakers' voice on speech recognition thresholds in noise: Effects of low- and high-frequency hearing impairment. *Speech, Language and Hearing*, *18*, 83–90. https://doi.org/10.1179/2050572814Y.0000000053

Lauay, C., Gerlach, N. M., Adkins-Regan, E., & DeVoogd, T. J. (2004). Female zebra finches require early song exposure to prefer high-quality song as adults. *Animal Behaviour*, *68*(6), 1249–1255. https://doi.org/10.1016/j.anbehav.2003.12.025

Lewicki, M. S., Olshausen, B. A., Surlykke, A., & Moss, C. F. (2014). Scene analysis in the natural environment. *Frontiers in Psychology*, *5*, 199. https://doi.org/10.3389/fpsyg.2014.00199

Liu, R., & Holt, L. L. (2015). Dimension-based statistical learning of vowels. *Journal of Experimental Psychology. Human Perception and Performance*, *41*(6), 1783–1798. https://doi.org/10.1037/xhp0000092

Lochner, J. P. A., & Burger, J. F. (1961). The intelligibility of speech under reverberant conditions. *Acta Acustica United with Acustica*, *11*(4), 195–200. https://www.ingentaconnect.com/content/dav/aaua/1961/00000011/00000004/art00004

Marler, P. (1970). A comparative approach to vocal learning: Song development in white-crowned sparrows. *Journal of Comparative and Physiological Psychology*, *71*(2p2), 1–25. https://doi.org/10.1037/h0029144

Mathews, R. C., Buss, R. R., Stanley, W. B., Blanchard-Fields, F., Cho, J. R., & Druhan, B. (1989). Role of implicit and explicit processes in learning from examples: A synergistic effect. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *15*(6), 1083–1100. https://doi.org/10.1037/0278-7393.15.6.1083

McAlpine, D., & de Hoz, L. (2023). Listening loops and the adapting auditory brain. *Frontiers in Neuroscience*, *17*, 1081295. https://doi.org/10.3389/fnins.2023.1081295

McDermott, J. H., Schemitsch, M., & Simoncelli, E. P. (2013). Summary statistics in auditory perception. *Nature Neuroscience*, *16*(4), 493–498.

McDermott, J. H., & Simoncelli, E. P. (2011). Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. *Neuron*, *71*(5), 926–940.

McWalter, R., & McDermott, J. H. (2018). Adaptive and Selective Time Averaging of Auditory Scenes. *Current Biology: CB*, *28*(9), 1405-1418.e10. https://doi.org/10.1016/j.cub.2018.03.049

Mesgarani, N., David, S. V., Fritz, J. B., & Shamma, S. A. (2014). Mechanisms of noise robust representation of speech in primary auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(18), 6792–6797. https://doi.org/10.1073/pnas.1318017111

Mishra, S. K., & Lutman, M. E. (2014). Top-down influences of the medial olivocochlear efferent system in speech perception in noise. *PloS One*, *9*(1), e85756.

Morrone, M. C. (2010). Brain development: critical periods for cross-sensory plasticity [Review of *Brain development: critical periods for cross-sensory plasticity*]. *Current Biology: CB*, *20*(21), R934-6. Elsevier. https://doi.org/10.1016/j.cub.2010.09.052

Nielsen, J. B., & Dau, T. (2010). Revisiting perceptual compensation for effects of reverberation in speech identification. *The Journal of the Acoustical Society of America*, *128*(5), 3088–3094. https://doi.org/10.1121/1.3494508

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1–32. https://doi.org/10.1016/0010-0285(87)90002-8

Nydam, A. S., Sewell, D. K., & Dux, P. E. (2018). Cathodal electrical stimulation of frontoparietal cortex disrupts statistical learning of visual configural information. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *99*, 187–199. https://doi.org/10.1016/j.cortex.2017.11.008

Pandya, D. N., & Barnes, C. L. (2019). Architecture and connections of the frontal lobe. *The Frontal Lobes Revisited*. https://doi.org/10.4324/9781315788975-3/architecture-connections-frontal-lobe-deepak-pandya-clifford-barnes

Pascual-Leone, A., Wassermann, E. M., Grafman, J., & Hallett, M. (1996). The role of the dorsolateral prefrontal cortex in implicit procedural learning. *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale*, *107*(3), 479–485. https://doi.org/10.1007/BF00230427

Peissig, J., & Kollmeier, B. (1997). Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners. *The Journal of the Acoustical Society of America*, *101*(3), 1660–1670. https://doi.org/10.1121/1.418150

Perrot, X., Ryvlin, P., Isnard, J., Guénot, M., Catenoix, H., Fischer, C., Mauguière, F., & Collet, L. (2006). Evidence for corticofugal modulation of peripheral auditory activity in humans. *Cerebral Cortex* , *16*(7), 941–948.

Petrides, M., & Pandya, D. N. (2002). Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. *The European Journal of Neuroscience*, *16*(2), 291–310. https://doi.org/10.1046/j.1460-9568.2001.02090.x

Phillips, D. P. (1985). Temporal response features of cat auditory cortex neurons contributing to sensitivity to tones delivered in the presence of continuous noise. *Hearing Research*, *19*(3), 253–268. https://doi.org/10.1016/0378-5955(85)90145-5

Plakke, B., & Romanski, L. M. (2014). Auditory connections and functions of prefrontal cortex. *Frontiers in Neuroscience*, *8*, 199. https://doi.org/10.3389/fnins.2014.00199

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*(6), 855–863. https://doi.org/10.1016/S0022-5371(67)80149-X

Rebuschat, P. (2015). *Implicit and Explicit Learning of Languages*. John Benjamins Publishing Company. https://play.google.com/store/books/details?id=jI-ACgAAQBAJ

Rees, A., & Palmer, A. R. (1988). Rate-intensity functions and their modification by broadband noise for neurons in the guinea pig inferior colliculus. *The Journal*

of the Acoustical Society of America. https://pubs.aip.org/asa/jasa/article-abstract/83/4/1488/826235

Robinson, B. L., Harper, N. S., & McAlpine, D. (2016). Meta-adaptation in the auditory midbrain under cortical influence. *Nature Communications*, *7*, 13442. https://doi.org/10.1038/ncomms13442

Rothwell, J. C., Hallett, M., Berardelli, A., Eisen, A., Rossini, P., & Paulus, W. (1999). Magnetic stimulation: motor evoked potentials. The International Federation of Clinical Neurophysiology. *Electroencephalography and Clinical Neurophysiology. Supplement*, *52*, 97.

Sabine, H. (1953). Room acoustics. *Trans IRE*, *1*, 4–12. https://www.theatrecrafts.com/archive/cue/cue_18_5.pdf

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. https://doi.org/10.1016/s0010-0277(98)00075-4

Saffran, Jenny R. (2003). Statistical Language Learning: Mechanisms and Constraints. *Current Directions in Psychological Science*, *12*(4), 110–114. https://doi.org/10.1111/1467-8721.01243

Salvi, R. J., Lockwood, A. H., Frisina, R. D., Coad, M. L., Wack, D. S., & Frisina, D. R. (2002). PET imaging of the normal human auditory system: responses to speech in quiet and in background noise. *Hearing Research*, *170*(1–2), 96–106. https://doi.org/10.1016/s0378-5955(02)00386-6

Sandrini, M., Umiltà, C., & Rusconi, E. (2011). The use of transcranial magnetic stimulation in cognitive neuroscience: a new synthesis of methodological issues. *Neuroscience and Biobehavioral Reviews*, *35*(3), 516–536.

Santon, F. (1976). Numerical prediction of echograms and of the intelligibility of speech in rooms. *The Journal of the Acoustical Society of America*, *59*(6), 1399–1405. https://doi.org/10.1121/1.381027

Schroeder, M. R. (1962). Frequency-Correlation Functions of Frequency Responses in Rooms. *The Journal of the Acoustical Society of America*, *34*(12), 1819–1823. https://doi.org/10.1121/1.1909136

Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron*, *36*(2), 241–263. https://doi.org/10.1016/s0896-6273(02)00967-4

Shapiro, S., & Wilk, M. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3–4), 591–611. https://doi.org/10.1093/BIOMET/52.3-4.591

Shinn-Cunningham, B. (2000, July). *Learning Reverberation: Considerations for Spatial Auditory Displays*. https://www.researchgate.net/profile/Barbara-Shinn-Cunningham/publication/2414695_Learning_Reverberation_Considerations_for_Spatial_Auditory_Displays/links/0f31752f91fb83f924000000/Learning-Reverberation-Considerations-for-Spatial-Auditory-Displays.pdf

Shinn-Cunningham, B., & Kawakyu, K. (2003). Neural representation of source direction in reverberant space. *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE Cat. No.03TH8684)*, 79–82. https://doi.org/10.1109/ASPAA.2003.1285824

Simpson, A. J. R., Harper, N. S., Reiss, J. D., & McAlpine, D. (2014). Selective adaptation to “oddball” sounds by the human auditory system. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(5), 1963–1969. https://doi.org/10.1523/JNEUROSCI.4274-13.2013

Smith, E. C., & Lewicki, M. S. (2006). Efficient auditory coding. *Nature*, *439*(7079), 978–982. https://doi.org/10.1038/nature04485

Srinivasan, N. K., & Zahorik, P. (2012). Prior listening exposure to a reverberant room improves open-set intelligibility of high-variability sentences. *The Journal of the Acoustical Society of America*, *133*(1), EL33–EL39. https://doi.org/10.1121/1.4771978

Stilp, C. (2020). Acoustic context effects in speech perception. *Wiley Interdisciplinary Reviews. Cognitive Science*, *11*(1), e1517. https://doi.org/10.1002/wcs.1517

Takács, Á., Kóbor, A., Kardos, Z., Janacsek, K., Horváth, K., Beste, C., & Nemeth, D. (2021). Neurophysiological and functional neuroanatomical coding of statistical and deterministic rule information during sequence learning. *Human Brain Mapping*, *42*(10), 3182–3201. https://doi.org/10.1002/hbm.25427

Terreros, G., & Delano, P. H. (2015). Corticofugal modulation of peripheral auditory responses. *Frontiers in Systems Neuroscience*, *9*. https://doi.org/10.3389/fnsys.2015.00134

Traer, J., & McDermott, J. H. (2016). Statistics of natural reverberation enable perceptual separation of sound and space. *Proceedings of the National Academy of Sciences*, *113*(48), E7856–E7865. https://doi.org/10.1073/pnas.1612524113

van den Bos, R., Jolles, J. W., & Homberg, J. R. (2013). Social modulation of decision-making: a cross-species review. *Frontiers in Human Neuroscience*, *7*, 301. https://doi.org/10.3389/fnhum.2013.00301

Vékony, T., Ambrus, G. G., Janacsek, K., & Nemeth, D. (2022). Cautious or causal? Key implicit sequence learning paradigms should not be overlooked when assessing the role of DLPFC (Commentary on Prutean et al.). *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *148*, 222–226. https://doi.org/10.1016/j.cortex.2021.10.001

Vlahou, E., Seitz, A. R., & Kopčo, N. (2019). Nonnative implicit phonetic training in multiple reverberant environments. *Attention, Perception & Psychophysics*, *81*(4), 935–947. https://doi.org/10.3758/s13414-019-01680-0

Wächter, T., Lungu, O. V., Liu, T., Willingham, D. T., & Ashe, J. (2009). Differential effect of reward and punishment on procedural learning. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *29*(2), 436–443. https://doi.org/10.1523/JNEUROSCI.4132-08.2009

Wang, L., Li, X., Hsiao, S. S., Lenz, F. A., Bodner, M., Zhou, Y.-D., & Fuster, J. M. (2015). Differential roles of delay-period neural activity in the monkey dorsolateral prefrontal cortex in visual–haptic crossmodal working memory. *Proceedings of the National Academy of Sciences*, *112*(2), E214–E219. https://doi.org/10.1073/pnas.1410130112

Watkins, A. J. (2005a). Perceptual compensation for effects of echo and of reverberation on speech identification. *Acta Acustica United with Acustica*, *91*(5), 892–901. https://www.ingentaconnect.com/content/dav/aaua/2005/00000091/00000005/art00010

Watkins, A. J. (2005b). Perceptual compensation for effects of reverberation in speech identification. *The Journal of the Acoustical Society of America*, *118*(1), 249–262. https://doi.org/10.1121/1.1923369

Watkins, A. J., & Makin, S. J. (2007). Steady-spectrum contexts and perceptual compensation for reverberation in speech identification. *The Journal of the*

Acoustical Society of America, *121*(1), 257–266. https://doi.org/10.1121/1.2387134

Watkins, P. V., & Barbour, D. L. (2008). Specialized neuronal adaptation for preserving input sensitivity. *Nature Neuroscience*, *11*(11), 1259–1261. https://doi.org/10.1038/nn.2201

Weisser, A., Buchholz, J. M., Oreinos, C., Badajoz-Davila, J., Galloway, J., Beechey, T., & Keidser, G. (2019). The Ambisonic Recordings of Typical Environments (ARTE) database. *Acta Acustica United with Acustica: The Journal of the European Acoustics Association (EEIG)*, *105*(4), 695–713. https://doi.org/10.3813/aaa.919349

Wen, B., Wang, G. I., Dean, I., & Delgutte, B. (2009). Dynamic range adaptation to sound level statistics in the auditory nerve. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *29*(44), 13797–13808. https://doi.org/10.1523/JNEUROSCI.5610-08.2009

Wong, P. C. M., Ettlinger, M., Sheppard, J. P., Gunasekera, G. M., & Dhar, S. (2010). Neuroanatomical Characteristics and Speech Perception in Noise in Older Adults. *Ear & Hearing*, *31*(4), 471–479. https://doi.org/10.1097/aud.0b013e3181d709c2

Zahorik, P., & Wightman, F. L. (2001). Loudness constancy with varying sound source distance. *Nature Neuroscience*, *4*(1), 78–83. https://doi.org/10.1038/82931

Zahorik, Pavel. (2009). Perceptually relevant parameters for virtual listening simulation of small room acoustics. *The Journal of the Acoustical Society of America*, *126*(2), 776–791. https://doi.org/10.1121/1.3167842

Zahorik, Pavel, & Brandewie, E. J. (2016). Speech intelligibility in rooms: Effect of prior listening exposure interacts with room acoustics. *The Journal of the Acoustical Society of America*, *140*(1), 74–86. https://doi.org/10.1121/1.4954723

Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., & Schoonhoven, R. (2006). Top-down and bottom-up processes in speech comprehension. *NeuroImage*, *32*(4), 1826–1836. https://doi.org/10.1016/j.neuroimage.2006.04.199

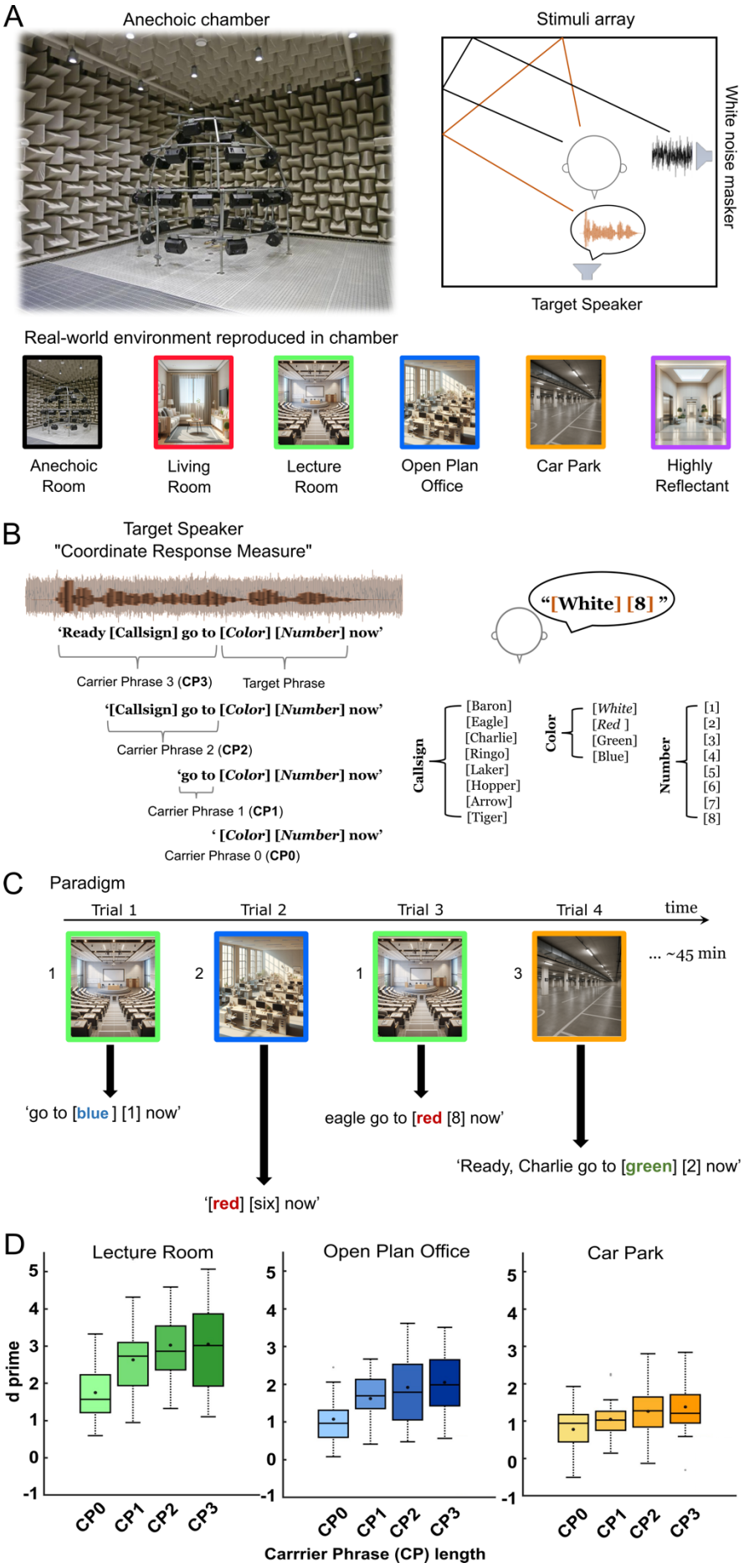
Zikopoulos, B., & Barbas, H. (2006). Prefrontal projections to the thalamic reticular nucleus form a unique circuit for attentional mechanisms. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *26*(28), 7348–7361. https://doi.org/10.1523/JNEUROSCI.5511-05.2006

Table 1. Real rooms acoustic characteristics.

	RT_{60iso} (ISO 3382) / $RT_{60Traer}$ (Traer & McDermott, 2016)	**Room Volume**	**Critical distance**	**Distance [m]**	**DRR [dB]**	**C50 [dB]**
Anechoic Room	0/0	n/a	n/a	1.3	n/a	n/a
Living room	0.33s/0.40s			1.3		
Lecture Room	0.46s/0.49s	164m³	2.78m	1.3	6.87	15.44
Open-Plan Office	0.96s/0.85s	446m³	2.49m	1.3	5.2	13.89
Small reflective room	1.55s/1.27s	134m³	1.18m	1.3	0.29	3.83
Underground Car Park	2.42s/1.32s	>5700m³	3.34m	1.3	7.21	11.53

DRR: Direct to Reverberant Ratio

C50: Speech Clarity Index


A
Anechoic chamber
Stimuli array
White noise masker
Target Speaker
Real-world environment reproduced in chamber
Anechoic Room
Living Room
Lecture Room
Open Plan Office
Car Park
Highly Reflectant
B
Target Speaker
"Coordinate Response Measure"
'Ready [Callsign] go to [Color] [Number] now'
Carrier Phrase 3 (CP3)
Target Phrase
'[Callsign] go to [Color] [Number] now'
Carrier Phrase 2 (CP2)
'go to [Color] [Number] now'
Carrier Phrase 1 (CP1)
' [Color] [Number] now'
Carrier Phrase 0 (CP0)
"[White] [8] "
Callsign
[Baron]
[Eagle]
[Charlie]
[Ringo]
[Laker]
[Hopper]
[Arrow]
[Tiger]
Color
[White]
[Red]
[Green]
[Blue]
Number
[1]
[2]
[3]
[4]
[5]
[6]
[7]
[8]
C
Paradigm
Trial 1
Trial 2
Trial 3
Trial 4
time
... ~45 min
'go to [blue] [1] now'
'[red] [six] now'
eagle go to [red [8] now'
'Ready, Charlie go to [green] [2] now'
D
Lecture Room
Open Plan Office
Car Park
d prime
CP0
CP1
CP2
CP3
Carrier Phrase (CP) length

1613

Fig 1. Experimental environment. **A.** Speech-in-noise immersed in real reverberant environments reproduced in an anechoic chamber over a 41-loudspeaker array. All speech material was rendered as if it was originating from directly in front of a listener and was masked by spatially separated (90° to the left) white noise at -15 dB signal-to-noise ratio (SNR). The noise masker level was fixed at 70 dB SPL. The speech material was convolved with the measured impulse response of a variety of different real rooms, from anechoic to highly reverberant real rooms such as an underground car park (see Table 1 methods for acoustic details). **B.** Structure of the speech material. Sentences from the Coordinated Response Measures (CRM) corpus were modified to vary their duration depending on how many words preceded the target phrase i.e., 'carrier phrase' (CP) length. Participants were asked to recognize and repeat target words: |Color| from a list of four and |Number| from a 1-8 list. **C.** Example of the experimental paradigm. Trials consisted of 'carrier phrases' spoken in a specific room i.e., 'go to blue 1 now' in the Lecture Room. Participants were never exposed to the same room consecutively and the task lasted no longer than 45 minutes. **D.** Sensitivity to |Color| and |Number| combined in the Lecture Room, Open-Plan Office and Underground Car Park. Mean d' (measure of accuracy calculated as: Z (correct responses) – Z (false alarm)] denoted as circles in the boxplot (n=22 for all rooms). The horizontal line denotes the median. Upper and lower limits of the boxplot represent ^{1s}t (q1) and ^{3r}d (q3) quartiles respectively, while whiskers denote the interquartile range (IQR =q3-q1). As previously reported (Brandewie & Zahorik, 2010, 2013), we observed an increase in performance for longer 'carrier phrases' in all environments. https://doi.org/10.25949/24295342.v1. All images in this figure were generated using artificial intelligence, with the exception of the Anechoic Room photograph, which was taken by the authors

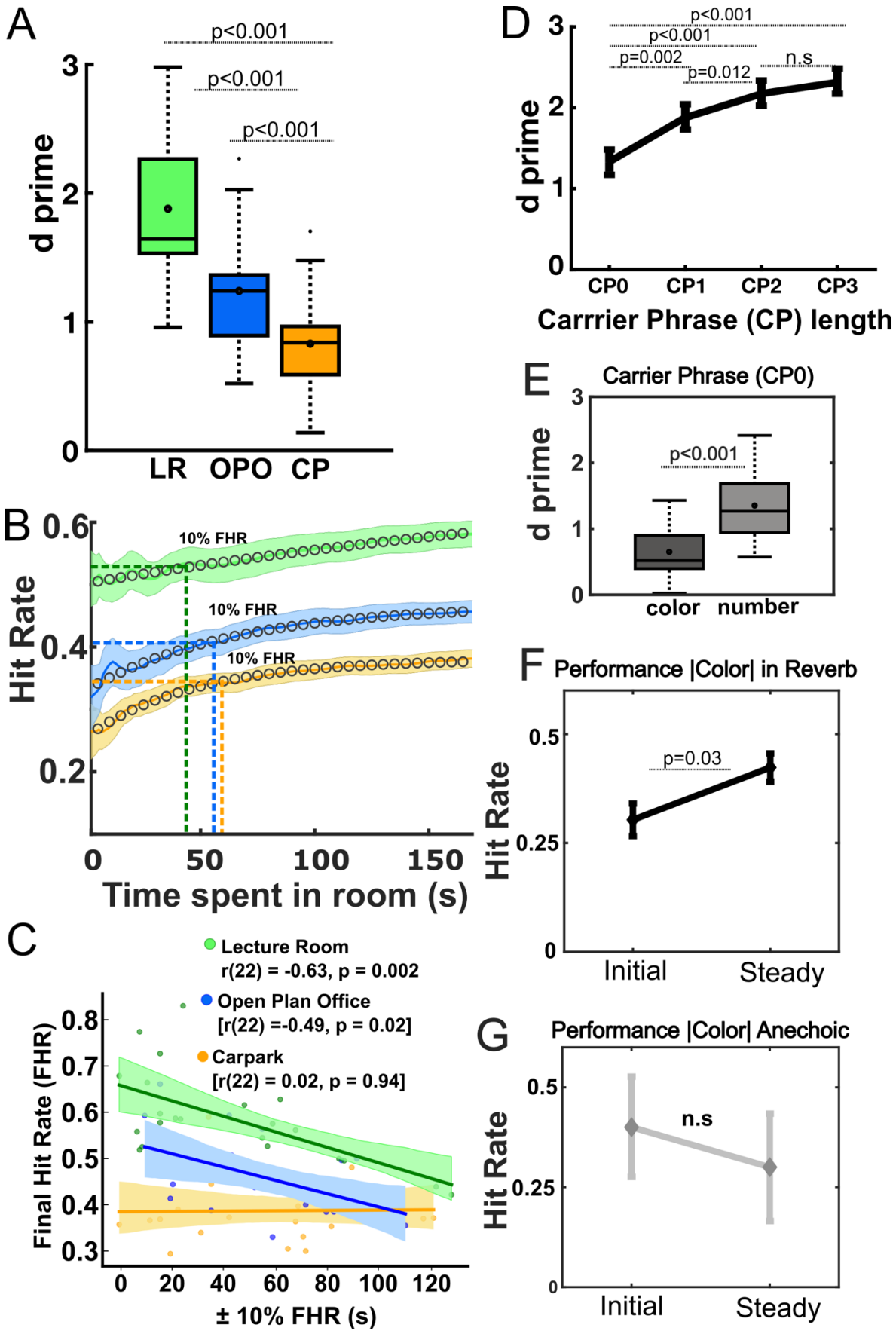
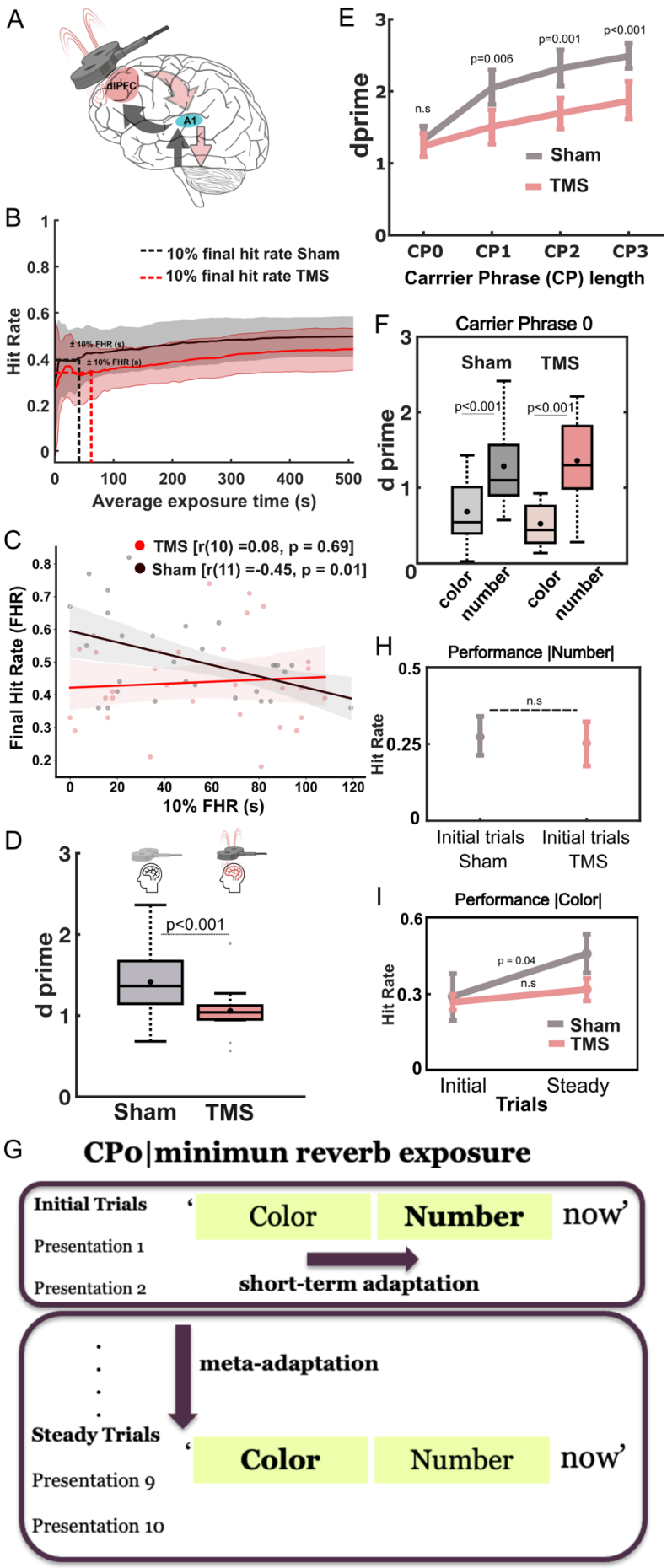


Figure 2. Learning effects in three real rooms. **A.** Overall Performance (d') (including correct and incorrect responses to all |*Colors*| and |*Numbers*|) in the Lecture Room (LR), Open-Plan Office (OPO)

and Car Park (CP). Mean d' denoted as circles in the boxplot (n=22 for all rooms). The horizontal line denotes the median. Upper and lower limits of the boxplot represent ^{1s}t (q1) and ^{3r}d (q3) quartiles respectively, while whiskers denote the interquartile range (IQR =q3-q1). **B.** Time course of performance in each room for 22 listeners. Correct responses (Hit Rate) across time spent in each room are shown, solid curves in color correspond to mean data after 5-time point moving averages, equivalent to ~7s, were applied). Optimal time course-fittings (using a two-phase association model) are plotted as color markers in each room. Associated shaded areas representing standard errors of the mean. Dashed coloured lines represent the time point at which participants reached a stable performance in each environment [i.e., ± 10% of the Final Hit Rate (FHR)], here 'global learning'. Notice that no statistical differences were found for 'global learning' among the different rooms. **C.** Shows a correlation analysis between FHR and the time where 10% of the FHR is achieved in each environment. Significant negative correlations (Pearson) were observed for Lecture Room, and Open Plan Office, suggesting the earlier participants reach a stable performance, the higher their FHR is. **D.** Performance for each carrier phrase length, d' was significantly better as the 'carrier phrases length' increased except for CP2 vs. CP3 (n.s), where a roll over effect was observed. **E.** Performance to |Color| and |Number| for CP0 i.e., our proxy for short-term adaptation the carrier phrase where exposure to an environment remained minimal. Here all the environments have been collapsed i.e., rANOVA, main effect of "target word". **F-G.** Hit Rate for Initial (1-2) and Steady (9-10) trials for keword |Color| in the CP0 condition, (F) in Reverberant conditions, indicating a significant improvement in performance for Steady trials, likely due to the accumulation of acoustic/environmental knowledge i.e., meta-adaptation (G) in anechoic (no echoes) showing a lack of improvement in performance i.e., lack of meta-adaptation in the absence of reverberation**.** https://doi.org/10.25949/24295342.v1


A
dlPFC
A1
B
10% final hit rate Sham
10% final hit rate TMS
Hit Rate
Average exposure time (s)
C
TMS [r(10) =0.08, p = 0.69]
Sham [r(11) =-0.45, p = 0.01]
Final Hit Rate (FHR)
10% FHR (s)
D
p<0.001
d prime
Sham
TMS
E
dprime
n.s
p=0.006
p=0.001
p<0.001
Sham
TMS
CP0
CP1
CP2
CP3
Carrrier Phrase (CP) length
F
Carrier Phrase 0
Sham
TMS
p<0.001
p<0.001
d prime
color
number
color
number
H
Performance |Number|
n.s
Hit Rate
Initial trials Sham
Initial trials TMS
I
Performance |Color|
p = 0.04
n.s
Hit Rate
Sham
TMS
Initial
Steady
Trials
G
CPo|minimun reverb exposure
Initial Trials
Presentation 1
Presentation 2
'Color Number now'
short-term adaptation
meta-adaptation
Steady Trials
Presentation 9
Presentation 10
'Color Number now'

1661

Figure 3. Learning effects in 'sham' and 'TMS' conditions. **A.** Schematic representation of functional connections between the dorsolateral prefrontal cortex (dlPFC) and primary auditory cortex (A1) under transcranial magnetic stimulation (TMS)**. B.** Average exposure time (s) 'sham TMS' and 'real TMS' conditions. Correct responses across time spent in each room for 'sham' and 'TMS' conditions. Solid curves in pink and gray correspond to mean data from 'TMS' and 'Sham' conditions respectively (after 5-time point moving averages, equivalent to ~7s, were applied), with the associated shaded areas representing standard errors of the mean. Dashed black (sham TMS) and pink (real TMS) lines represent the time point at which participants reached a stable performance in each environment [i.e., ± 10% of the Final Hit Rate (FHR)], here 'global learning', no statistical differences were found for 'global learning' between TMS conditions. **C.** Shows a correlation analysis between FHR and the time where 10% of the FHR is achieved in under TMS (pink) and Sham (gray) conditions. Significant negative correlations (Pearson) were observed under Sham stimulation, suggesting the earlier participants reach a stable performance, the higher their FHR is. This relationship was disrupted (lack of significant correlation) under TMS conditions. **D. Overall** Performance in TMS and Sham exposed participants i.e., collapsed sensitivity (d') to |Color| and |Number| for all carrier phrases and environments (significant main effect of TMS conditions: Univariate ANOVA). d' values plotted in bright pink corresponds to the 'real TMS' condition whereas the 'sham TMS' accuracy is plotted in gray. Mean d' is denoted as circles in the boxplot (n=11 for 'sham TMS' conditions and n=10 for 'real TMS' conditions). **E.** Significant Interaction 'TMS condition' x 'carrier phrase length'. d' was significantly better as the 'carrier phrases length' increased for 'sham TMS' compared to 'real TMS' except for CPO (n.s), i.e., our proxy for short-term adaptation and minimal exposure to acoustic environments. **F.** Performance to |Color| and |Number| for CP0 i.e., our proxy for short-term adaptation and the carrier phrase where exposure to an environment remained minimal. Here all the environments have been collapsed as only a main effect of "target word", performance to number always significantly higher than for colour independent of TMS condition. **G.** Schematic for assessing short-term adaptation (performance to |Number| in the Initial Trials) and meta-adaptation (performance to |Color| between Initial and Steady Trials). **H.** Isolated short-term adaptation effects for performance to |Number| in CP0. Hit Rate for Initial (1-2) trials only for |Number| in the CP0 condition for both TMS (pink) and Sham (gray) conditions. Notice that no differences were observed between performance |Number| under Sham or TMS conditions, suggesting no disruption of short-term adaptation.**I.** Hit Rate for Initial (1-2) and Steady (9-10) trials only for |Color| in the CP0 condition for both TMS (pink) and Sham (gray) conditions. Notice that only a significant improvement in performance for from Intitial to Steady trials was observed in the 'sham TMS' condition. https://doi.org/10.25949/24295342.v1

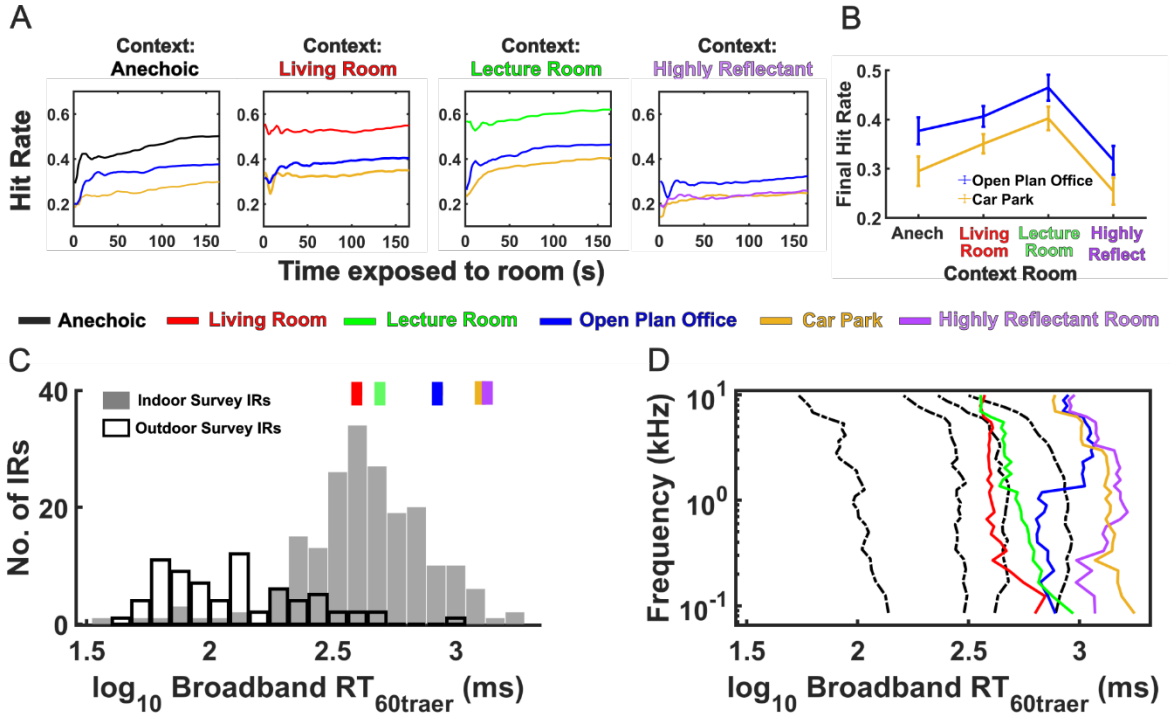


Figure 4. Performance in different 3-room combinations where acoustics of 3 rooms span an ecological range. **A.** Overall hit rate for different 3-room combinations. Mean data for four different 3-room combinations are plotted where Open-Plan Office (*dark blue*) and Underground Car Park (*yellow*) were always included and the context room was either Anechoic (*black*), Living Room (*red*), Lecture Room (*green*), Lecture Room (*green*) or Highly Reflective (*purple)*. **B.** Performance in Open-Plan Office and Car Park when paired with different context rooms. Final hit rates were highest for Office and Car Park when presented in conjunction with the Lecture Room, followed by the Living Room and Anechoic with poorest performance observed when the Highly Reflective space was the context. **C. Environments tested** compared to ecological range. The $RT_{60traer}$s for the five test rooms, calculated as the median RT_{60} value across 31 frequency sub-bands (Traer & McDermott, 2016), are plotted (*colored boxes*) above a histogram of the median $RT_{60traer}$s calculated for 199 indoor (*filled bars*) and 72 outdoor (*empty bars*) spaces recorded by Traer and McDermott (2016). **D**. Frequency dependence of reverberation time ($RT_{60traer}$) in test rooms compared to ecological range. RT_{60}s of the 5 test rooms are displayed for frequency sub-bands used to calculate $RT_{60traer}$ (*colored lines*). Quartiles for combined indoor and outdoor spaces (Traer & McDermott, 2016) are plotted as *black dashed lines*. Decreasing reverberation time above 0.5kHz is observed for Living Room, Lecture Room and Car Park as has been typically described for indoor spaces (Traer & McDermott, 2016); however $RT_{60traer}$ profiles for Open-Plan Office and Highly Reflective space were notable for their longer reverberation times at and above 1kHz. https://doi.org/10.25949/24295342.v1

Supplemental Table 1. Pairwise comparisons between performance (d') for carrier phrases length n 22 participants.

(I) Carrier Phrase	(J) Carrier Phrase	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
CP0	CP1	-.295*	.070	.002	-.498	-.091
	CP2	-.510*	.063	<.001	-.692	-.328
	CP3	-.586*	.089	<.001	-.846	-.325
CP1	CP2	-.215*	.045	<.001	-.345	-.085
	CP3	-.291*	.083	.012	-.532	-.050
CP2	CP3	-.076	.074	1.000	-.291	.139

Based on estimated marginal means
*. The mean difference is significant at the .05 level.
b. Adjustment for multiple comparisons: Bonferroni.

Supplemental Table 2. Pairwise comparisons among performance (d') for Interaction 'target word' x 'carrier phrase length' in 22 participants

Carrier Phrase	Mean Difference (Number-Color)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
				Lower Bound	Lower Bound
CP0	.701*	.104	<.001	.486	.917
CP1	.246*	.094	.016	.052	.441
CP2	.348*	.095	.001	.150	.546
CP3	.237	.128	.077	-.028	.503

Based on estimated marginal means
*. The mean difference is significant at the .05 level.
b. Adjustment for multiple comparisons: Bonferroni.

Supplemental Table 3. Pairwise comparisons among performance (Final Hit Rate) for the six talkers in 22 participants.

Talkers		Mean Difference	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	**2**	-.042	.016	.277	-.096	.012
	3	-.060*	.013	.002	-.101	-.018
	4	-.104*	.015	.000	-.154	-.055
	5	-.129*	.017	.000	-.185	-.073
	6	.015	.014	1.000	-.032	.063
2	**3**	-.018	.014	1.000	-.066	.030
	4	-.063*	.015	.007	-.113	-.012
	5	-.087*	.015	.000	-.137	-.037
	6	.057	.017	.054	-.001	.115
3	**4**	-.045	.015	.123	-.096	.006
	5	-.069*	.014	.001	-.116	-.022
	6	.075*	.013	.000	.032	.118
4	**5**	-.025	.017	1.000	-.080	.031
	6	.120*	.018	.000	.060	.179
5	**6**	-.144*	.015	.000	-.193	-.095

Based on estimated marginal means
*. The mean difference is significant at the .05 level.
b. Adjustment for multiple comparisons: Bonferroni.

Supplemental Table 4. Pairwise comparisons between performance (Final Hit Rate) for carrier phrases length in 22 participants anechoic conditions.

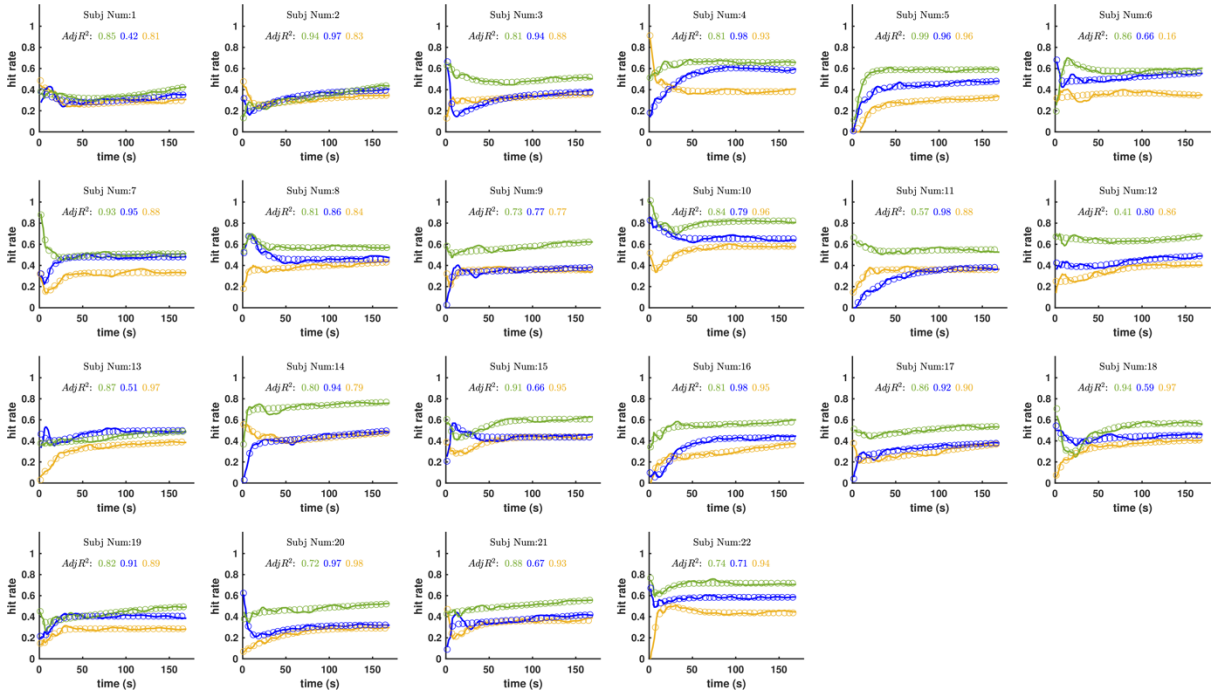
(I) Carrier Phrase	(J) Carrier Phrase	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
CP0	CP1	-6.000	3.602	.781	-18.118	6.118
	CP2	-15.667*	3.938	.019	-28.914	-2.419
	CP3	-13.167	4.314	.082	-27.679	1.346
CP1	CP2	-9.667	2.885	.051	-19.371	.038
	CP3	-7.167	2.972	.235	-17.164	2.830
CP2	CP3	-2.500	1.537	.829	-7.669	2.669

Based on estimated marginal means
*. The mean difference is significant at the .05 level.
b. Adjustment for multiple comparisons: Bonferroni.

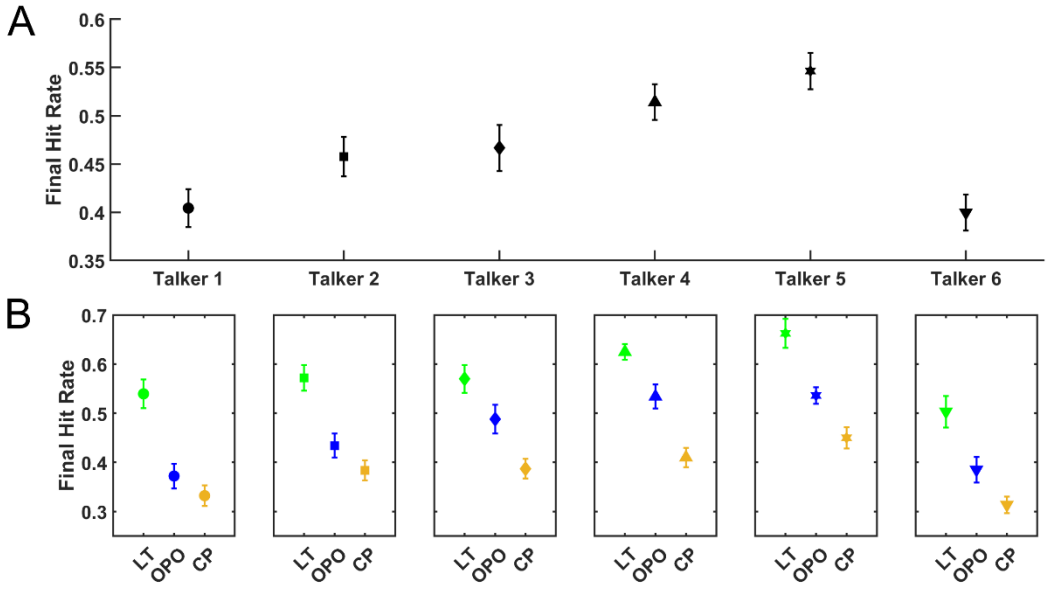
Supplemental Table 5. Pairwise comparisons for interaction ‘TMS condition’ x ‘carrier phrase length’.

Carrier Phrase	(I) TMS	(J) Sham	Mean Difference (I-J)	Std. Error	Sig.b	95% Confidence Interval for Difference b	
						Lower Bound	Upper Bound
CP0	1.00	2.00	-.022	.124	.859	-.265	.221
CP1	1.00	2.00	-.341*	.124	.006	-.585	-.098
CP2	1.00	2.00	-.409*	.124	.001	-.653	-.165
CP3	1.00	2.00	-.501*	.124	<.001	-.745	-.256

Based on estimated marginal means
*. The mean difference is significant at the .05 level.
b. Adjustment for multiple comparisons: Bonferroni.

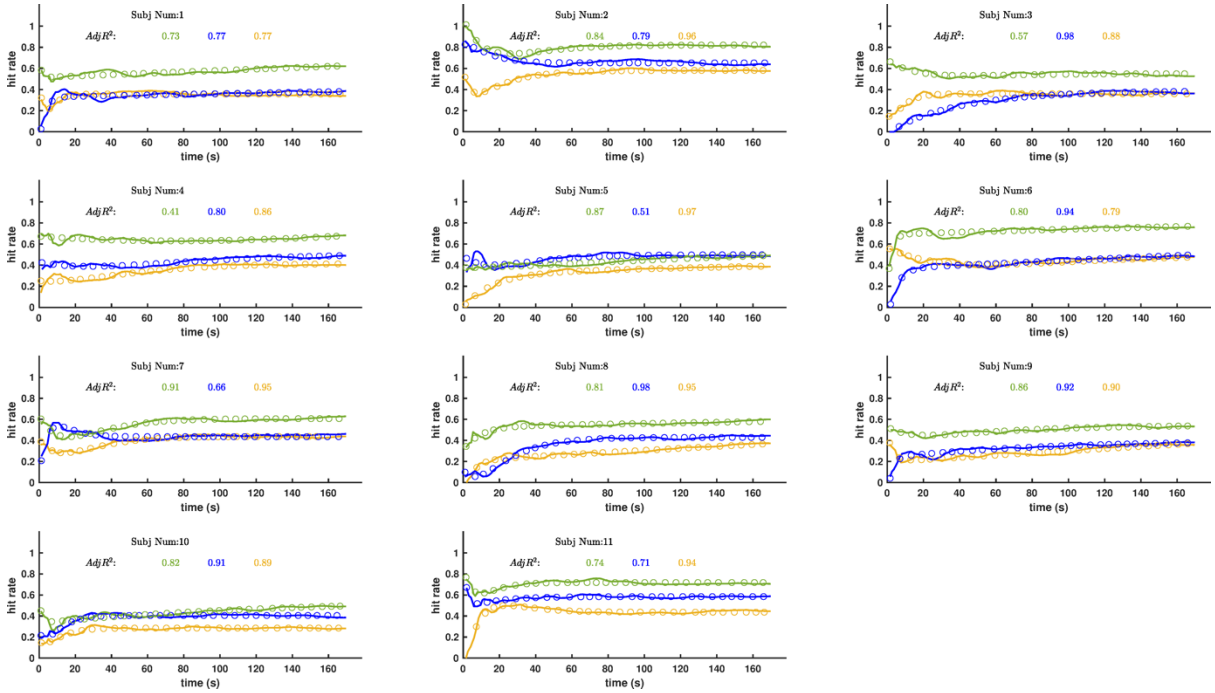


Supplemental Figure 1. Time courses of performance for 22 subjects in lecture theatre (green), open plan office (blue) and car park (yellow). Solid lines represent actual cumulative hit rate data for each room that were calculated from raw data using a 5-point moving average (equivalent to ~7s). Open circles represents best fit of double exponential with adjusted R2 values shown above for each individual. https://doi.org/10.25949/24295342.v1

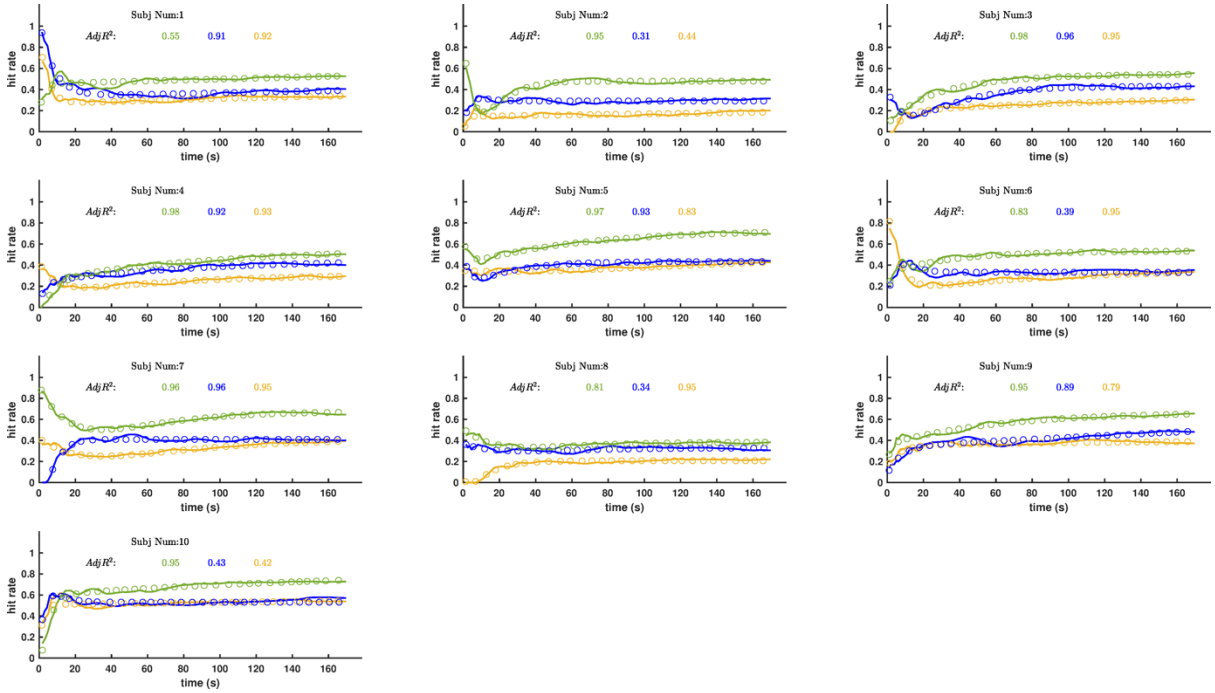


Supplemental Figure 2. Comparison of Final Hit Rate (FHR) performances for different talkers. **A.** Performance, calculated as FHR in all rooms, is shown for the 6 different talkers (1-3 male and 4-6 female) presented to 22 individuals who did not undergo experimental TMS during the task involving the Lecture Room/Open-Plan Office and Car Park RIRs. **B.** Performance was also similarly calculated for the different talkers but is now separated by the room RIR associated with the specific talker. https://doi.org/10.25949/24295342.v1

Supplemental Figure 3. Hit Rate performances for |Color| and |Number| independent of environment presented or carrier phrase length. **A.** Performance, calculated for the 1262 trials where 22 participants heard in 2 consecutive trials the same talker, a Wilcoxon signed rank test revealed no significant differences in performance between trial 1 vs. trial 2 (n=1262, Z= -0.42, p=0.68). **B.** Performance was also similarly calculated when listeners heard the same talker in three consecutive trials, this happened in a total of 217 trials and a Wilcoxon signed rank test revealed no significant differences in performance between trial 1 vs. trial 3 (n=217, Z= -0.17, p=0.87). **C.** Performance calculated in 40 trials were listeners heard the same talker in four consecutive trials. No statistical differences were observed between Trial 1 and Trial 4 (n=40, Z= -0.159, p=0.11).



Supplemental Figure 4. Time courses of performance for 11 sham-TMS subjects in lecture theatre (green), open plan office (blue) and car park (yellow). Solid lines represent actual cumulative hit rate data for each room that were calculated from raw data using a 5-point moving average (equivalent to ~7s). Open circles represents best fit of double exponential with adjusted R_2 values shown above for each individual. https://doi.org/10.25949/24295342.v1



Supplemental Figure 5. Time courses of performance for 10 TMSin lecture theatre (green), open plan office (blue) and car park (yellow). Solid lines represent actual cumulative hit rate data for each room that were calculated from raw data using a 5-point moving average (equivalent to ~7s). Open circles represents best fit of double exponential with adjusted R_2 values shown above for each individual. https://doi.org/10.25949/24295342.v1

1774 Supplemental Information. Transcranial Magnetic Stimulation screening form

1775

TMS SCREENING FORM

Date of birth..... UNIQUE IDENTIFIER: .....

Transcranial Magnetic Stimulation (TMS) is a method for producing an electric current in a small part of the brain. During TMS, a current passes through a copper coil that is wound inside a plastic casing and held over the participant's head. The current in the coil produces a magnetic field, which passes safely through the scalp and causes electrical activity in brain tissue.

Before receiving TMS, please read the following questions carefully and provide answers. For a small number of individuals, TMS may carry an increased risk of causing a seizure. The purpose of these questions is to make sure that you are not such a person. You have the right to withdraw from the screening and subsequent scanning if you find the questions unacceptably intrusive. The information you provide will be treated as strictly confidential and will be held in secure conditions.

If you are unsure of the answer to any of the questions, please ask the person who gave you this form or the person who will be performing the study. Definitions of some of technical terms are given overleaf.

	Please tick
Have you ever had an adverse reaction to TMS?	☐Yes ☐No
Do you experience claustrophobia?	☐Yes ☐No
Have you fainted in the last 12 months?	☐Yes ☐No
Have you had a seizure?	☐Yes ☐No
Have you had a stroke?	☐Yes ☐No
Have you had a brain aneurysm, an arteriovenous malformation or any other brain related condition?	☐Yes ☐No
Have you had a serious head injury (including neurosurgery)?	☐Yes ☐No
Do you have any metal in your head (outside the mouth) such as shrapnel, surgical clips, or fragments from welding or metalwork?	☐Yes ☐No
Do you have any implanted devices such as cardiac pacemakers, aneurysm clips, cochlear implants, medical pumps, deep brain stimulators, or intracardiac lines?	☐Yes ☐No
Do you suffer from frequent or severe headaches?	☐Yes ☐No
Do you have diabetes?	☐Yes ☐No
Have you ever had any illness that caused brain injury?	☐Yes ☐No
Have you ever had a heart attack?	☐Yes ☐No
Have you ever had a heart condition such as atrial fibrillation, angina, congestive heart failure, arrhythmia, cardiomyopathy, congenital heart disease or arterial disease?	☐Yes ☐No
Are you taking any psychiatric or neuroactive medications, or do you have a history of drug abuse?	☐Yes ☐No
Have you been prescribed any medication? If yes, please list_____	☐Yes ☐No
Are you pregnant?	☐Yes ☐No
Do you, or does anyone in your family, have epilepsy?	☐Yes ☐No
Do you hold a heavy goods vehicle driving license or bus license?	☐Yes ☐No

I have read and understood the questions above and have answered them correctly.

SIGNED..... **DATE**.....

In the presence of ..... **(Name)**

..... **(Signature)**

PACEMAKER: **An electronic device that is surgically placed in the patient's body and connected to the heart to regulate the heartbeat.**

COCHLEAR IMPLANT: **An electronic medical device that bypasses damaged structures in the inner ear and directly stimulates the auditory nerve, allowing some deaf individuals to learn to hear and interpret sounds and speech.**

ANEURYSM CLIP: **A surgically implanted metal clip used to cut off blood flow through the neck of an aneurysm. An aneurysm is a deformity of a blood vessel in the body, which can swell and burst causing a haemorrhage.**

SHUNT: **A surgically implanted connector, which allows passage of fluid between two parts of the body. A common use of a shunt is to allow fluid to drain away from the brain, thus reducing pressure in the brain. May also describe a tube which allows blood to be moved from one part of the body to another.**

STENT: **A surgical implanted device that is inserted into a blood vessel to provide support, keep the vessel open and promote unblocked and enhanced blood flow. Sometimes used in other fluid carrying vessels in the body such as bile ducts etc.**