

## **A maturational frequency discrimination deficit may explain developmental language disorder**

Jones, Sam; Stewart, Hannah J; Westermann, Gert

### **Psychological Review**

E-pub ahead of print: 27/07/2023

Peer reviewed version

[Cyswllt i'r cyhoeddiad / Link to publication](#)

*Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):*

Jones, S., Stewart, H. J., & Westermann, G. (2023). A maturational frequency discrimination deficit may explain developmental language disorder. *Psychological Review*.

#### **Hawliau Cyffredinol / General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

1 Running head: MATURATIONAL FREQUENCY DISCRIMINATION DEFICIT

2

3

4

5

6

7

8 **A maturational frequency discrimination deficit may explain developmental language**  
9 **disorder**

10

11 Samuel David Jones<sup>1, 2</sup>, Hannah Stewart<sup>2</sup>, and Gert Westermann<sup>2</sup>

12 <sup>1</sup>Department of Psychology, Bangor University

13 <sup>2</sup>Department of Psychology, Lancaster University

14

15 13448 words

16

17 **Author note**

18 Correspondence concerning this article should be addressed to Samuel Jones, Room

19 309, Brigantia, Bangor University, Bangor, LL57 2AS. Email: samuel.jones@bangor.ac.uk.

20 Gert Westermann was supported by Economic and Social Research Council (ESRC)

21 International Centre for Language and Communicative Development (LuCiD)

22 [ES/S007113/1 and ES/L008955/1]. The theoretical view described in this manuscript was

23 presented at the 28<sup>th</sup> Architectures and Mechanisms for Language Processing (AMLaP)

24 conference. All data and materials required to re-run the simulations and analyses presented

25 in this manuscript are available from the following public repository: <https://osf.io/x2h8k/>.

26 **Abstract**

27 Auditory perceptual deficits are widely observed among children with developmental  
28 language disorder (DLD). Yet the nature of these deficits and the extent to which they explain  
29 speech and language problems remain controversial. In this study, we hypothesise that  
30 disruption to the maturation of the basilar membrane may impede the optimisation of the  
31 auditory pathway from brainstem to cortex, curtailing high-resolution frequency sensitivity  
32 and the efficient spectral decomposition and encoding of natural speech. A series of  
33 computational simulations involving deep convolutional neural networks that were trained to  
34 encode, recognise, and retrieve naturalistic speech are presented to demonstrate the strength  
35 of this account. These neural networks were built on top of biologically truthful inner ear  
36 models developed to model human cochlea function, which – in the key innovation of the  
37 current study – were scheduled to mature at different rates over time. Delaying cochlea  
38 maturation qualitatively replicated the linguistic behaviour and neurophysiology of  
39 individuals with language learning difficulties in a number of ways, resulting in: (i) delayed  
40 language acquisition profiles; (ii) lower spoken word recognition accuracy; (iii) word finding  
41 and retrieval difficulties; (iv) ‘fuzzy’ and intersecting speech encodings and signatures of  
42 immature neural optimisation; and (v) emergent working memory and attentional deficits.  
43 These simulations illustrate the many negative cascading effects that a primary maturational  
44 frequency discrimination deficit may have on early language development, and generate  
45 precise and testable hypotheses for future research into the nature and cost of auditory  
46 processing deficits in children with language learning difficulties.

47 *Keywords:* developmental language disorder, auditory processing, spoken word  
48 recognition and retrieval, neural network, neural population geometry

49 **A maturational frequency discrimination deficit may explain developmental language**  
50 **disorder**

51 **Introduction**

52 There is astonishing variability in rates of early language development. Looking  
53 beyond population means, we see large windows of time in which language skills may  
54 emerge without any concern (Braginsky et al., 2018). Sometimes, however, a child's  
55 language is delayed enough to cause alarm among personal and professional caregivers. An  
56 estimated 7.5% of English-speaking children find acquiring and using language difficult  
57 enough to potentially interfere with their day-to-day emotional wellbeing and later with their  
58 educational outcomes (Norbury et al., 2016). Where such difficulties are evident in the  
59 absence of any obvious biomedical cause, such as Down's syndrome, the child may be  
60 diagnosed with developmental language disorder (DLD) and may undertake a tailored  
61 programme of language intervention targeting their specific areas of difficulty (Bishop et al.,  
62 2016).

63 Language disorder identification, assessment, and intervention are challenging  
64 because of the significant heterogeneity seen among affected children. Any aspect of  
65 language may be disrupted in DLD, from phonology through to syntax and pragmatics, and  
66 children often show concurrent developmental difficulties, for instance in motor control, or  
67 comorbidity with conditions such as developmental dyslexia or attention deficit hyperactivity  
68 disorder (ADHD) (Bishop et al., 2016). Furthermore, contrasting theoretical approaches have  
69 commonly centred on just one in a wide range of hypothesised cognitive faculties in  
70 accounting for discrete characteristics of this multifaceted profile. This approach has  
71 sometimes given the inaccurate impression that DLD is evidence of an isolated deficit in that  
72 faculty alone, for instance in working memory (Archibald & Gathercole, 2006), predictive

73 processing (Hestvik et al., 2022), lateral inhibition (McMurray et al., 2019), or statistical  
74 learning (Ullman & Pierpont, 2005).

75         The complex symptomology seen in DLD and overlap across associated diagnostic  
76 groups, at the level of both linguistic profile (i.e., from phonemes to pragmatics) and  
77 implicated cognitive faculties (e.g., working memory, statistical learning), has fostered a shift  
78 towards a ‘transdiagnostic’ mindset in neurodevelopmental disorder research (Astle et al.,  
79 2022). Here, focus is on what we might call *canonical* features of impairment – features  
80 sometimes termed ‘bridging symptoms’ – that hold widely not just within but often across  
81 diagnostic groups. Working memory deficits measured, for instance, in the nonword  
82 repetition and span tasks are widely considered one such canonical feature of developmental  
83 disorder, given that such deficits appear quite consistently across young children with a range  
84 of developmental difficulties (Archibald & Harder Griebeling, 2016; Gray et al., 2019; Henry  
85 & Botting, 2017).

86         Maintaining that there are canonical features of developmental disorder is, of course,  
87 very different from assuming there is a single cause of any given disorder. In general,  
88 contemporary research on early language disorder is averse to the notion that the varied  
89 profiles seen among children might have a single cause. This is perhaps a well-justified  
90 reaction to early research that held up DLD as evidence of an isolated deficit in an innately  
91 specified language acquisition device (a ‘grammar module’ of the brain encoded by the  
92 FOXP2 gene; Pinker, 1994) or similarly suggested that DLD was evidence of an discrete  
93 deficit in, for instance, working memory or statistical learning. We now know that the picture  
94 is considerably more complex. At the levels of genetics, neurobiology, and cognition, DLD  
95 appears to entail a constellation of causal mechanisms and risk factors (Bishop, 2006). A  
96 transdiagnostic, mechanism-centred approach fully appreciates this complexity and attempts  
97 to identify those dimensions of disorder that apply widely (though not *uniformly*) and which

98 may point us to better understanding and more effective intervention strategies (Fletcher-  
99 Watson, 2022). The careful, in-depth study of a specific and well-recognised canonical area  
100 of difficulty might show us how much we ‘get for free’ when we really explore the wide  
101 cascading effects implied by that area of difficulty.

102         The current study centres on one such canonical feature of developmental language  
103 disorder; auditory processing difficulties. While deficits in auditory perception are widely  
104 identified among children with neurodevelopmental disorder, most notably in DLD and  
105 dyslexia, the extent to which such deficits can explain early speech and language problems  
106 remains controversial (Bishop et al., 1999, 1999, 2012; Bishop & McArthur, 2005; Haake et  
107 al., 2013; McArthur & Bishop, 2004; Merzenich et al., 1996; Rosen, 2003; Tallal, 2013). In  
108 this study, we hypothesise that disruption to the maturation of the neural architecture  
109 underpinning high-resolution frequency discrimination from the prenatal period through the  
110 first two years of life (specifically, a disruption to basilar membrane maturation and resulting  
111 deficits in auditory brainstem optimization) may play a causal role in early speech and  
112 language disorder. Our account builds on prior work by McArthur and Bishop (2004) and  
113 Bishop and McArthur (2005), who first suggested that deficits in frequency discrimination  
114 may play an important role in the impairments observed among some children and  
115 adolescents with a diagnosis of DLD. In this study, we aim to substantially develop this  
116 account and to demonstrate its strength in a series of computational simulations that illustrate  
117 the varied consequences of a low-level frequency discrimination deficit within a controlled  
118 and transparent artificial learning environment. We aim to document the varied potential  
119 costs to early language development – i.e., the many cascading effects that we ‘get for free’ –  
120 as a result of a fundamental maturational deficit in frequency discrimination.

121         We begin this report by reviewing empirical research into the auditory processing  
122 skills of children with language disorder, highlighting an evolution from early theoretical

123 accounts centred around temporal processing, which relates to the speed at which the auditory  
124 system responds to acoustic input, to relatively recent accounts centred around frequency or  
125 *spectral* processing. We then review research into the maturation of the neural architecture  
126 supporting high-resolution frequency discrimination ability from the neonatal period through  
127 childhood, before considering how a disruption to this typical maturational trajectory might  
128 give rise to speech and language deficits. Subsequently, we present a computational model in  
129 which we simulate different rates of maturation in frequency discrimination ability while  
130 monitoring language acquisition rates, spoken word recognition accuracy, proxies for word  
131 finding latency, and neural speech representation integrity. We then discuss the implications  
132 of our results, the limitations of our computational approach, and directions for future  
133 investigation.

#### 134 **From temporal to spectral processing deficits in language disorder research**

135 A dominant view developed principally through the work of Tallal and colleagues is  
136 that children with language learning difficulties have a primary deficit affecting the  
137 perception of acoustic signals that change rapidly, something that these authors refer to as a  
138 temporal processing deficit<sup>1</sup> (e.g., Merzenich et al., 1996; Tallal et al., 1981). Much of the  
139 empirical research in this direction made use of the auditory repetition task, or ART, in which  
140 children press buttons to identify changes in frequency in a series of pure tones. In the ART,  
141 performance accuracy among children with DLD was regularly shown to decrease  
142 significantly when inter-stimulus interval (ISI; i.e., the gap between tones) was reduced to  
143 below approximately 250 milliseconds, lending apparent support to the hypothesis that these  
144 children's auditory processing systems were ill-equipped to accurately perceive and encode  
145 rapidly unfolding natural speech (Merzenich et al., 1996; Tallal et al., 1981). This line of

---

<sup>1</sup> We note that the term 'temporal processing deficit' has been objected to on the basis that this body of research shows no evidence that the awareness of temporal order is compromised among children with DLD. The assumed difficulty instead relates to rapid changes in frequency, and so the term 'rapid perception deficit' may be more appropriate (Bishop, 2014, p. 90).

146 argument has been pursued in a significant body of research and has motivated the  
147 development of the Fast ForWord programme of intervention, which claims to be able to  
148 train sensitivity to rapidly occurring auditory stimuli through the controlled manipulation of  
149 ISI and in doing so confer gains in speech and language abilities (Tallal, 2013).

150         Despite the initial dominance of the temporal processing deficit hypothesis, however,  
151 a series of failed replications, both of the basic research and of the Fast ForWord intervention  
152 (Strong et al., 2011; Bishop & McArthur, 2005; McArthur & Bishop, 2004; see Rosen, 2003,  
153 for review) has motivated the search for alternative characterisations of the auditory  
154 perceptual deficits that appear to affect many children with speech and language problems.  
155 One promising, though comparatively underexamined view is that such deficits are spectral  
156 rather than temporal in nature (Bishop & McArthur, 2005; McArthur & Bishop, 2004;  
157 Mengler et al., 2005). That is, that these children's difficulty relates principally to  
158 distinguishing discrete sounds of similar frequency rather than discrete sounds that rapidly  
159 follow one another. For instance, across two studies Bishop and McArthur presented children  
160 aged 10 to 19 with and without language disorder with a baseline tone of 600Hz and a  
161 distinct tone which was initialised at 700Hz, but which was raised or lowered by increments  
162 of 25Hz to determine the minimal frequency discrimination threshold, or limen, that  
163 participants could identify (Bishop & McArthur, 2005; McArthur & Bishop, 2004; see also  
164 Mengler et al., 2005). These authors found that the minimal frequency discrimination  
165 threshold among children with severe language disorder was 750Hz (i.e., a 150Hz disparity)  
166 during an initial assessment and 674Hz at follow up (i.e., a 74Hz disparity), compared to  
167 629Hz and 624Hz disparities respectively for control children. Readers may wish to visit one  
168 of the many freely available online pure tone generators to compare tones in this range  
169 themselves. For many, the average difference between the minimal threshold tones identified  
170 by children with DLD (i.e., 600Hz and 750Hz or 674Hz) will appear striking, attesting to the



171 difficulty such a deficit may cause during the analysis of the complex spectral profiles of  
172 natural speech (Nuttall et al., 2018; Sumner et al., 2018).

173         Crucially, Bishop and McArthur found that this deficit in frequency discrimination  
174 was observed regardless of the rate of stimulus presentation, providing compelling evidence  
175 that the auditory processing difficulties of some children affected by language disorder are  
176 spectral rather than temporal in nature, and perhaps explaining the failed replications of key  
177 studies in the temporal processing deficit literature (Bishop & McArthur, 2005; McArthur &  
178 Bishop, 2004; Mengler et al., 2005; Rosen, 2003; Strong et al., 2011). What is more, even  
179 those children with DLD who performed well in the behavioural tone discrimination task  
180 nevertheless showed immature waveforms during electroencephalography (EEG) monitoring,  
181 providing tentative support for the maturational account that Bishop and McArthur (2005)  
182 then offer to explain their findings.

### 183 **The maturation of frequency discrimination skills**

184         Bishop and McArthur (2005) explain their results in terms of a disruption to the  
185 typical maturation of high-resolution frequency discrimination. In order to situate this  
186 account, upon which we intend to elaborate, it is useful to review key research on the early  
187 maturation of frequency discrimination skills, and the neural basis of these skills. In younger  
188 children and infants, probing the maturation of frequency discrimination skills presents a  
189 significant challenge. Paradigms such as head turning and high-amplitude sucking have  
190 provided mixed results and are open to interpretation, not least that a failure to discriminate  
191 tones in such paradigms may be the result of immature motor skills or attention (see Burnham  
192 & Mattock, 2014, for review). In response, some researchers have advocated the use of  
193 neuroimaging methods such as EEG and magnetoencephalography when studying frequency  
194 discrimination in neonates and infants (e.g., Novitski et al., 2007). Despite their own

195 limitations, such neuroimaging methods are often considered to provide an index of neural  
196 activity that is relatively independent of motor and attentional factors (Novitski et al., 2007).

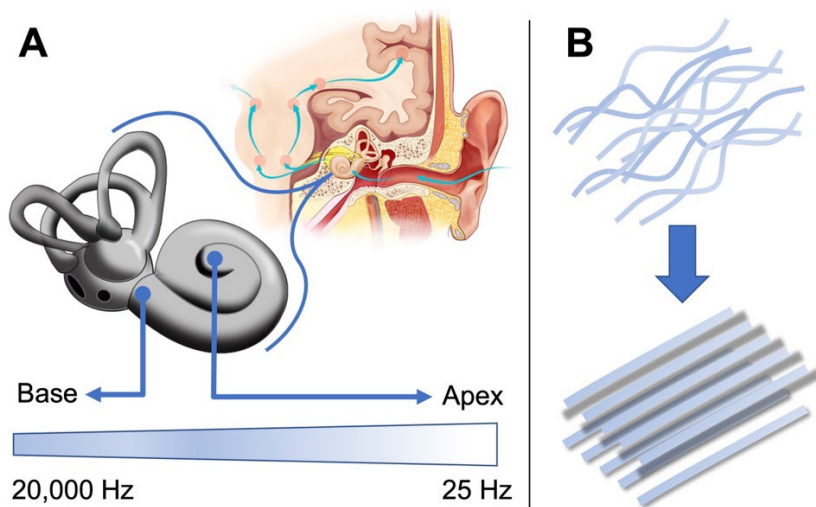
197       Neuroimaging involving neonates and infants corroborates indications from  
198 behavioural research of an early maturation in frequency discrimination ability (Jensen &  
199 Neff, 1993; Lopez-Poveda, 2014; Novitski et al., 2007; Shafer et al., 2000; Tharpe &  
200 Ashmead, 2001). This maturation is not uniform. High-frequency tone discrimination is  
201 approximately adult-like in apparently typically developing infants by six months of age. In  
202 contrast, low-frequency discrimination, in the range more regularly associated with speech  
203 signals (e.g., 400Hz), develops more slowly, with continued maturation apparent in children  
204 up to ages seven to nine (Jensen & Neff, 1993; Burnham & Mattock, 2014). While the  
205 empirical data vary somewhat, estimates from the ‘odd one out’ paradigm (also known as the  
206 ‘mismatch negativity paradigm’) suggest that newborns can detect a 20% though not a 5%  
207 change in frequency in a 250Hz-4000Hz window (Novitski et al., 2007; see Burnham &  
208 Mattock, 2014, for review). Such findings support the view that frequency resolution  
209 improves considerably from birth through childhood, making it increasingly easy to  
210 discriminate competing acoustic signals, and thus to perform the complex spectral analysis  
211 that accurate and efficient natural speech perception and encoding requires (Nuttall et al.,  
212 2018; Sumner et al., 2018).

213       The maturation of frequency discrimination skills reflects changes in neural  
214 architecture that, though many important questions remain, are now in a large part reasonably  
215 well understood. A key characteristic of the auditory perceptual system upon which speech  
216 representation and use is based is its tonotopic structure. That is, throughout the auditory  
217 pathway, from the inner ear to the auditory brainstem and on to the auditory cortex, we see  
218 selective responsivity to acoustic input of particular frequencies among sensory cells and  
219 neurons that constitutes the neural basis of frequency resolution and the decomposition of

220 auditory signals including speech (Echteler et al., 1989; Nuttall et al., 2018; Sumner et al.,  
 221 2018). The characteristic ‘tonotopic’ structure of the auditory pathway results predominantly  
 222 from the physical properties of the basilar membrane, a 35-mm coiled membrane within the  
 223 inner ear (Figure 1A).

224 **Figure 1**

225 *Schematic of Frequency Tuning and Structural Development in the Mammalian Cochlea*



227 *Note.* Panel A shows the location of the cochlea in the inner ear (coloured inset), its coiled  
 228 structure (in grey), and the mechanical frequency sensitivity gradient from base to apex of the  
 229 basilar membrane within the cochlea. Panel B illustrates the development of basilar  
 230 membrane micro-structure supporting high-resolution frequency tuning, from fibres that are  
 231 low-diameter, sparse, and ‘braided’, to fibres that are higher-diameter, dense, and regular.  
 232 The Panel A auditory system image (coloured inset) is in the public domain ([https://](https://commons.wikimedia.org/wiki/File:Hearing_mechanics.png)  
 233 [commons.wikimedia.org/wiki/File:Hearing\\_mechanics.png](https://commons.wikimedia.org/wiki/File:Hearing_mechanics.png)). The Panel A greyscale cochlea  
 234 image is available under a Creative Commons Attribution-Share Alike 4.0  
 235 International license ([https://commons.wikimedia.org/wiki/File:Inner\\_ear.png](https://commons.wikimedia.org/wiki/File:Inner_ear.png)).

236 The basilar membrane is narrow and firm at its base, and as a result of these physical  
 237 properties fibres in this basal region vibrate maximally to the high frequencies in auditory

238 input (Figure 1A; Sumner et al., 2018). The apex of the basilar membrane is, in contrast, wide  
239 and relatively slack, and as a result fibres in this apical region vibrate maximally to the low  
240 frequencies in auditory input (Figure 1A; Sumner et al., 2018). For instance, voiceless  
241 fricatives such as /f/, which contain relatively high-frequency components, may stimulate  
242 basal regions of the membrane, while vowels such as /a:/, which contain low-frequency  
243 components, may stimulate apical regions. Upon the basilar membrane sit a single row of  
244 approximately 3500 inner hair cells which become selectively responsive to specific  
245 frequencies – that is, they are ‘frequency-tuned’ – as a result of their position on the basilar  
246 membrane (Sumner et al., 2018; Tani et al., 2021). In turn, inner hair cells are innervated by  
247 spiral ganglion neurons which project to the cochlear nucleus, with this and subsequent  
248 innervation conserving tonotopic sensitivity and resulting in the emergence of frequency  
249 sensitive ‘maps’ throughout a complex array of subcortical structures of the auditory  
250 brainstem and on to the peripheral auditory cortex. The physical properties of the basilar  
251 membrane are, therefore, at the heart of frequency sensitivity and acoustic signal  
252 decomposition across the auditory pathway, and this itself underpins accurate and efficient  
253 speech processing and encoding (Burnham & Mattock, 2014; Echteler et al., 1989; Nuttall et  
254 al., 2018; Sumner et al., 2018; Tani et al., 2021). From the third trimester to 6 months of age  
255 structures from the auditory nerve throughout the auditory pathway to the auditory cortex  
256 undergo substantial changes in synaptic organisation, myelination, and dendritic arborisation,  
257 and this process of maturation continues through to two years of age during a typically rich  
258 period of language development (Chonchaiya et al., 2013). Work by Chonchaiya et al. (2013)  
259 indicates that, by nine months of age, auditory brainstem responses continuous with relatively  
260 mature brainstem organisation are predictive of better language outcomes.

261           Recent research has cast light on how the pre- and postnatal structural development of  
262 the basilar membrane underpins the emergence of high-resolution frequency tuning across the

263 auditory-linguistic pathway. Studies using electron microscopy and polarized light  
264 microscopy have shown that the basilar membrane is composed of collagenous filaments, or  
265 fibres, which are initially relatively low diameter, sparsely organised, and ‘braided’, but  
266 which increase in diameter, density, and linear regularity throughout early development  
267 (Figure 1B). Such studies have also determined an uneven time course in which structural  
268 maturation is slower in the membrane apex than it is in basal regions, a finding consistent  
269 with behavioural and neurophysiological evidence that low frequency component tuning  
270 comes online relatively slowly (Burnham & Mattock, 2014; Novitski et al., 2007; Tani et al.,  
271 2021). Animal models also provide mounting evidence that the protein coding gene *emilin*  
272 2 (elastin microfibril interfacier 2), which is part of the emilin family of glycoproteins that  
273 contribute in part to tissue elasticity, can seriously disrupt fibre development in the basilar  
274 membrane – i.e., can curtail typical increases in fibre diameter, density, and linear regularity  
275 – and can, therefore, disrupt the membrane’s capacity to propagate frequency sensitivity  
276 throughout posterior structures of the auditory pathway supporting accurate and efficient  
277 frequency decomposition (Amma et al., 2003; Russell et al., 2020; Tani et al., 2021). This  
278 literature demonstrates how a genetic abnormality can in principle disrupt the emergence of  
279 the mechanical gradient of the basilar membrane.

## 280 **Towards a maturational account of frequency resolution deficits and speech and** 281 **language difficulties**

282 Before stating our hypothesis, let us take stock of the key points reviewed so far:

- 283 1. Auditory processing deficits are widespread among children with DLD, and these  
284 deficits appear to be frequency-based rather than temporal in nature.
- 285 2. Evidence that deficits are related to frequency analysis points to specific cellular and  
286 neural structures of the auditory pathway. Specifically, the basilar membrane is at the  
287 heart of frequency tuning across the auditory pathway, with tonotopic maps emerging

288 throughout the auditory brainstem and cortex predominantly as a result of dynamic  
289 adaptation to the structural properties – i.e., the *mechanical gradient* – of the basilar  
290 membrane.

291 3. The basilar membrane undergoes crucial structural changes early in development,  
292 with the fibres from which the membrane is composed increasing in diameter, density,  
293 and regularity, in part as a result of *emilin 2* expression. This process of maturation is  
294 integral to the emergence of tonotopy across the auditory pathway.

295 Our hypothesis is, then, that:

296 *Early disruption to the maturation of the physical properties of the basilar membrane*  
297 *which underpin that membrane's mechanical gradient (i.e., increases in fibre density,*  
298 *diameter, and linear regularity) may disturb the optimisation of the posterior auditory*  
299 *pathway from the brainstem to the cortex, curtailing high-resolution tonotopic*  
300 *sensitivity and contributing to speech and language difficulties in some children.*

301 The auditory pathway is, of course, a highly complex system, which could be disrupted by  
302 any number of influences operating across any number of its subsystems. It is, for instance,  
303 possible that auditory brainstem and auditory cortex optimisation are disrupted despite a  
304 properly maturing basilar membrane. A range of such alternative possibilities are presented in  
305 our *Discussion* section. Nevertheless, we believe that the hypothesis above provides a strong  
306 starting point for investigation given that (i) the auditory processing deficits we see in DLD  
307 appear to be spectral in nature and (ii) that a fully matured basilar membrane sits at the heart  
308 of high-resolution frequency processing across the auditory pathway. Our hope is that this  
309 literature review has shown that – though more work is undoubtedly required – there already  
310 exists a great deal of empirical evidence bearing on typical and atypical auditory pathway  
311 maturation and the potential impact of a maturational delay in this area on the emergence of  
312 speech and language. In our view, what is currently required to direct future investigation is a

313 compelling theoretical account linking these fragmentary research strands, and this is what  
314 we attempt to provide in the current study. Our aim is emphatically not to suggest that  
315 frequency discrimination deficits wholly explain early language disorder. Instead, we aim to  
316 flesh out one candidate mechanistic pathway within a complex constellation of many.

317         In what follows we simulate and monitor the dynamic adaptation of an artificial  
318 auditory-linguistic pathway (broadly auditory brainstem to cortex) in response to biologically  
319 plausible representations of speech-elicited activation patterns in the developing cochlea,  
320 under (i) non-developmental, (ii) regular, and (iii) delayed maturational trajectories. We show  
321 how a disruption to the maturation of cochlea microarchitecture may result in the atypical  
322 optimisation of subsequent neural pathways, qualitatively accounting for several commonly  
323 recorded characteristics of atypical human linguistic behaviour and neurophysiology, namely:  
324 (i) delayed language acquisition profiles (e.g., Norbury et al., 2016); (ii) spoken word  
325 recognition deficits (Andreu et al., 2012; Evans et al., 2018; Rispens et al., 2015; Velez &  
326 Schwartz, 2010); (iii) word finding or retrieval problems (Kambanaros et al., 2015; Messer &  
327 Dockrell, 2006); (iv) ‘fuzzy’ long-term speech representations (Claessen et al., 2009); (v)  
328 atypical neural signatures of auditory signal processing (e.g., Bishop & McArthur, 2005); and  
329 (vi) apparent working memory deficits, attributable, we argue, to the imprecision of activated  
330 long-term speech representations (Henry & Botting, 2017; Jones & Westermann, 2022).

### 331 **Overview of simulations<sup>2</sup>**

#### 332 *Network and training and testing regimes*

333         The architecture used in these simulations is an artificial neural network known as a  
334 deep convolutional neural network. The work of McDermott and colleagues has been  
335 instrumental in demonstrating that despite obvious disparities between the biological auditory

---

<sup>2</sup> This paper is associated with a fully annotated Jupyter notebook (Kluyver et al., 2016), which is available from the following public repository and which can be used to replicate the simulations described or to experiment with alternative parameter configurations: <https://osf.io/x2h8k/>.

336 pathway and this artificial counterpart, including in general complexity and in learning  
337 procedures (see *Discussion*), close parallels are observed between convolutional neural  
338 network activity and human behavioural and neural responses across a wide range of tasks,  
339 such as speech localization, pitch perception, and hearing in noise (Francl & McDermott,  
340 2022; Kell et al., 2018; Saddler et al., 2021). Convolutional neural networks are not ‘circuit  
341 models’ of the brain. That is, these networks are not intended to explicitly model fine-grained  
342 physiology such as ion channel behaviour (e.g., see Higgins et al., 2017, for a circuit model  
343 of speech perception and category formation). Rather, convolutional neural networks can  
344 provide high-order ‘computational’ insight, in the sense of Marr (1982), into how a  
345 perceptual processing hierarchy dynamically adapts to a particular form of input to solve a  
346 certain problem under varying constraints.

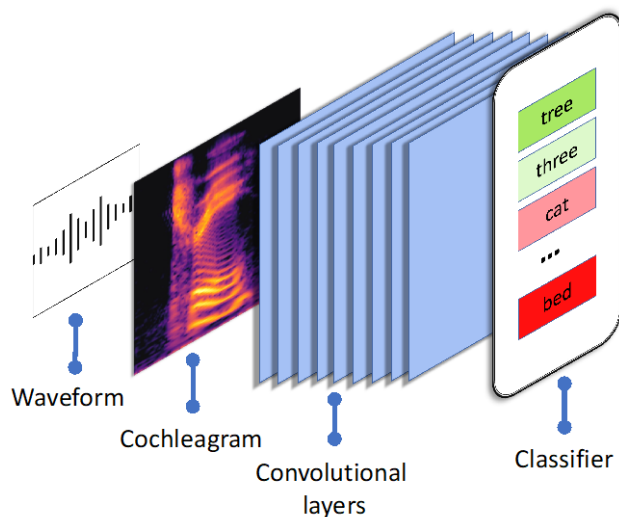
347         Our simulations made use of the ResNet-18 deep convolutional neural network (He et  
348 al., 2015), which we implemented using PyTorch (Paszke et al., 2019) in Python (Python  
349 Software Foundation, 2008). A full network description can be retrieved by running the  
350 Jupyter notebook associated with this project. Note that following the code examples  
351 associated with Stephenson et al. (2020; see  
352 [https://github.com/schung039/neural\\_manifolds\\_replicaMFT](https://github.com/schung039/neural_manifolds_replicaMFT)), many of our analyses centre  
353 around the networks’ 20 convolutional layers. For this reason, these layers are detailed in the  
354 Appendix alongside key hyperparameters. A total of nine convolutional neural networks  
355 ( $n=3$ ; conditions defined below) were trained and tested on spoken words from the speech  
356 commands dataset (Warden, 2018), which contains 105,829 one-second spoken word  
357 waveforms of 35 word types (Figure 2). The speech commands dataset was chosen for this  
358 project because it is free and openly available, and because it is perhaps unique in comprising  
359 such a large number of exemplars of natural speech. Limitations of the speech commands  
360 dataset are noted in our discussion.



361 Over ten cycles, or ‘epochs’, of training, networks were required to categorise each  
362 spoken word that they perceived by outputting a probability distribution over their 35-word  
363 lexicon. The word with the highest probability assigned was taken as the networks’ selection.  
364 Networks responded dynamically to error signals propagated upon an incorrect classification  
365 by updating their inner weight matrices using the backpropagation algorithm after each  
366 spoken word exposure (i.e., batch size = 1) in order to reduce the future error rate. This  
367 constitutes a broad computational analogy to fluctuation in synaptic connection strength due  
368 to long-term potentiation (Lillicrap et al., 2020). Throughout training, networks were  
369 presented with random samples of 4000 exemplars per-epoch from the speech commands  
370 dataset. Random samples were matched within epochs across the network groups we define  
371 below. For instance, network one in each experimental condition saw the same random  
372 samples of training data, which differed in each training epoch. This ensures that any later-  
373 observed group-level performance discrepancies are not a function of differences in the data  
374 that the network has been trained on. We note that there is nothing special about the word as  
375 a unit of representation here. Our choice of dataset principally reflects its scale and the fact  
376 that it contains authentic spoken words, and similar effects would be expected were we  
377 modelling phonemes or multi-word constructions.

378 **Figure 2**

379 *Neural Network Schematic*



380

381 *Note.* Authentic, raw spoken waveforms are first passed through a cochlea model, before  
382 being passed through the deep convolutional neural network and the 35-way classifier.

383 Later, at test, neural networks were presented with another random sample of 1000  
384 words from the speech commands dataset, a random sample which was again matched across  
385 conditions (defined below). We recorded a range of test performance metrics including  
386 speech recognition accuracy, proxies for response latency and word finding difficulties  
387 (namely, predictive distribution entropy or spread), confusion matrixes, and item specific  
388 effects (i.e., fitting a Bayesian model of what lexical features contributed to a correct or  
389 incorrect spoken word classification). We also analysed what form the networks' internal  
390 speech representations took, using a statistical physics method known as mean-field theory  
391 based manifold analysis to measure the average degree of spread of a single neural  
392 representation, and its overlap with competitor representations. These techniques are  
393 described in more detail below.

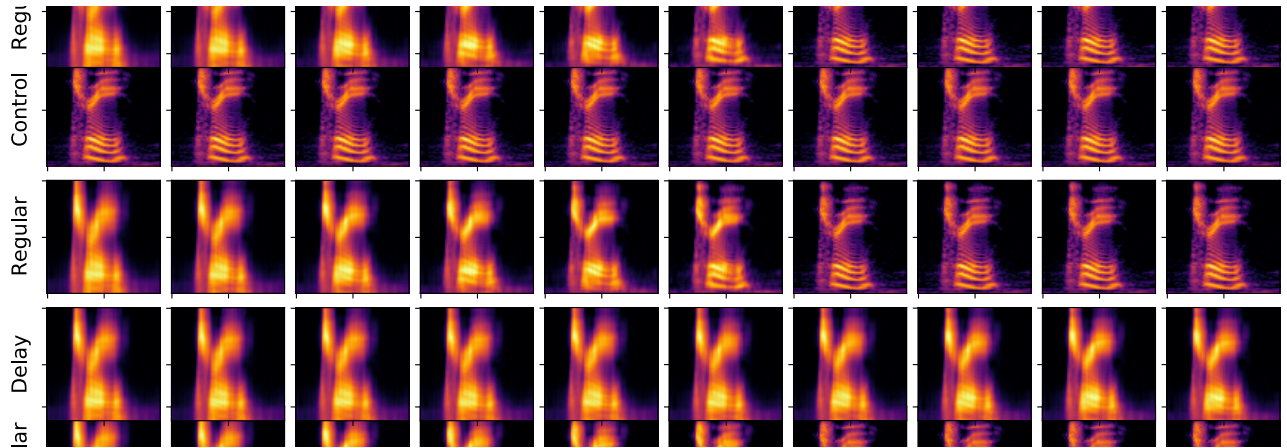
394 Convolutional neural networks are, in the vast majority of research, configured 'a-  
395 developmentally'. That is, parameters such as the number of layers or number of neurons per  
396 layer, etc. are fixed at the outset, and remain static during network training and testing (cf.  
397 Westermann et al., 2006; Westermann & Ruh, 2012; Westermann & Jones, 2021; these  
398 studies similarly involve neural networks that change structurally during learning, e.g., in

399 terms of the number of hidden units that they have). In contrast, one innovation of the current  
400 study was to model the maturation of high-resolution frequency discrimination skills using  
401 what is known as scheduled learning. That is, we ran distinct populations of neural networks  
402 in which frequency discrimination ability matured at different rates, according to different  
403 schedules across ten epochs of training. As can be seen in Figure 2, raw spoken word  
404 waveforms were initially passed through a cochleagram model developed specifically to  
405 replicate typical, human cochlea function (McDermott & Simoncelli, 2011). The resultant  
406  $100 \times 100$ -dimension cochleagram images were then passed through the deep convolutional  
407 neural network and later into a 35-way classifier. In three discrete conditions we manipulated  
408 the maturation of that initial cochleagram model in three neural networks ( $n = 3, N = 9$ ).  
409 Networks one, two, and three in each condition had identical weight initialisations. This  
410 ensured that any group-level performance discrepancies observed were not a function of the  
411 networks' starting states. Condition one was a-developmental – i.e., a baseline or control  
412 network – meaning that this network received high-resolution speech input from the outset  
413 and no changes to the network occurred during ten epochs of training (see Figure 3, row one).  
414 In contrast, the cochlea models of networks in conditions two and three matured according  
415 to a specific schedule. In condition two, frequency resolution started low, but improved  
416 rapidly, resulting in full-resolution processing (i.e., baseline-equivalent acuity) by epoch  
417 seven (Figure 3, row two). This can be seen in the increasing  $y$ -axis acuity (i.e., decreasing  
418 vertical blur) across the cochleagrams in row two of Figure 3. Networks in condition three, in  
419 contrast, started with precisely the same standard of frequency resolution as the networks in  
420 condition two – that is, frequency resolution is identical during training epoch one in the  
421 regular and delay conditions – but then followed a delayed maturation schedule, never  
422 reaching baseline acuity (Figure 3, row 1). In both the delay and regular conditions,  
423 frequency resolution was constrained using a normalised box filter with a kernel of shape

424 (1,  $y$ ), where  $y$  decreased at different rates over ten epochs: from 25 to 1 in the regular  
 425 condition and from 25 to 16 in the delay condition.

426 **Figure 3**

427 *Schedules of Simulated Basilar Membrane Maturation*



428  
 429 *Note.* Shown is a cochleagram of the word *tree* under varying rates of rates of maturation in  
 430 spectral (i.e.,  $y$ -axis) acuity within three conditions (control, regular, delay), and across ten  
 431 cycles (epochs) of training.

432 ***Methods of analysis***

433 All post-simulation analyses were conducted in R (RStudio Team, 2016). During  
 434 training and testing, networks were presented with cochleagrams and in response output  
 435 probability distributions over their 35-word lexicons. The word assigned the highest  
 436 probability was taken as a network's classification and where this corresponded to the true  
 437 target cochleagram a 'hit' was scored. The analysis of our training data involved measuring  
 438 spoken word classification accuracy by training epoch. At test, we measured classification  
 439 accuracy and the average maximum probability and probability distribution entropy output  
 440 when a classification was made. These metrics provide a proxy for a network's certainty in its  
 441 classifications. A high probability, low entropy (i.e., low spread) distribution signals high  
 442 certainty in a judgement, while a low probability, high entropy (i.e., high spread) distribution  
 443 signals low certainty in a judgement.

444           We then teased apart item-specific effects, looking for subsets of words on which  
445 regular or delayed networks performed better or worse. As part of this analysis into item-  
446 specific effects, we ran a Bayesian regression model (Burkner, 2017) in which the percentage  
447 of correct classifications per word was predicted by condition (i.e., regular, delayed) in  
448 interaction with two relevant independent variables that have generated considerable interest  
449 in developmental psycholinguistics: word frequency and word phonological neighbourhood  
450 density (e.g., Ambridge et al., 2015; Jones & Brandt, 2019; Rispens et al., 2015). Word  
451 frequency quantifies how common the word is in the exposure language, here the speech  
452 commands corpus from which training words were randomly sampled. Phonological  
453 neighbourhood density meanwhile quantifies the average distance, calculated on the basis of  
454 phonological transcriptions, between each word and the other 34 words in the training data.  
455 Relatively high input frequency is regularly associated with better language learning in  
456 children (Ambridge et al., 2015), while high phonological distance (i.e., phonemic  
457 dissimilarity) may improve speech classification accuracy among human listeners because  
458 potential between-item confusion is lower (Karimi & Diaz, 2020). As our modelling  
459 approach did not involve semantic representations it was not possible to include other  
460 variables of potential interest such as word concreteness, valence, or relevance to infants and  
461 babies (Braginsky et al., 2018; Jones & Brandt, 2019).

462           Artificial neural networks are sometimes criticised for being inscrutable ‘black  
463 boxes’. Yet, there exist numerous methods that enable the researcher to go beyond  
464 performance metrics such as accuracy alone to peer inside the network and understand how it  
465 is representing information in the service of completing a certain task. Exploiting such  
466 methods is vital to the current study because our interest is in how a processing hierarchy  
467 modelling the auditory pathway from brainstem to cortex optimises in the face of low-level  
468 constraints on frequency discrimination. Convolutional neural network activation patterns

469 have been shown to align broadly (i.e., not on a layer-to-structure level of granularity) with  
470 activation patterns in the biological brain (Kell et al., 2018; cf. Thompson, 2020).  
471 Furthermore, Bishop and MacArthur’s work in this direction shows that even when there is  
472 apparently no group difference in performance metrics such as accuracy, frequency resolution  
473 deficits may be associated with different neural signatures across groups with and without  
474 language disorder (Bishop & McArthur, 2005; McArthur & Bishop, 2004). Similarly,  
475 Chonchaiya et al. (2013) showed that auditory brainstem responses continuous with immature  
476 brainstem optimisation predict relatively poor language outcomes. We wondered whether a  
477 similar neural signature of auditory processing impairments within the context of language  
478 learning deficits would emerge within our computational framework.

479         To better understand how our neural networks dynamically optimised to cochlea  
480 representations with varying spectral acuity (Figure 3), we used a recently developed  
481 framework known as mean field theory based manifold analysis (MFTMA; Figure 4; Chung  
482 et al., 2018; Chung & Abbott, 2021; Cohen et al., 2020). Under this approach, each neuron in  
483 any given structure of the auditory pathway, for instance the inferior colliculus, is configured  
484 as a single axis against which the spiking activity in that neuron can be plotted. Collectively,  
485 neurons in a given neural structure then define a neural state space (Figure 4A; graphically, a  
486 collection of axes) in which patterns of activation can be plotted either as trajectories through  
487 time or averaged spikes-per-second vectors. Given neural noise and variability in speaker and  
488 communicative context, no two instances of any given speech string stimulate the same  
489 response vector within that neural state space, i.e., repeated spoken instances of a given  
490 linguistic structure never stimulate each neuron in the state space to the same degree.  
491 Repeated exposure to a range of exemplars from a single linguistic class, whether phoneme,  
492 word, or construction, therefore stimulates a unified population response known as a  
493 ‘manifold’, which is a quasi-continuous subspace of the neural state space that can be

494 considered the neural basis of the representation of that class (Cohen et al., 2020). Implicitly  
495 estimating the bounds of this neural manifold is considered integral to recognising and  
496 producing novel yet valid speech, as if recognising that instances of this class may regularly  
497 stimulate activation patterns within but not substantially outside this region of the state space  
498 (Cohen et al., 2020; DiCarlo & Cox, 2007; Stephenson et al., 2020; Yamins & DiCarlo,  
499 2016).

#### 500 **Figure 4**

#### 501 *Principles of Neural Population Geometry*



502

503 *Note.* Panel A shows the spoken words *tree*, *three*, and *two* as response vectors in high  
504 dimensional space, with axes  $N_1$  to  $N_n$  representing the response of a specific neuron within  
505 the population in spikes per second. The population here could be any structure within the  
506 auditory pathway (e.g., inferior colliculus, medial geniculate nucleus, etc.). Note that  
507 response vectors can also be shown as trajectories over time (e.g., see Chung & Abbott,  
508 2021). Exemplars of the same word, e.g., *tree*, reside in a different neural response vector as a  
509 function of neural noise and speaker and context effects, but collectively form a quasi-  
510 continuous manifold. (NB. In a deviation from the mathematical definition of a manifold,  
511 neural manifolds need not be smooth and continuous, but are instead held to comprise the  
512 convex hull of the distribution of neural responses elicited by a fixed class of stimulus.)

513 Panels B to D illustrate the neural basis of the well-studied transformation across the auditory  
514 system from noise sensitive to speech selective responses (e.g., Davis & Johnsruide, 2003;  
515 DeWitt & Rauschecker, 2012; Kaas et al., 1999; Okada et al., 2010). Early in the auditory

516 pathway manifolds of different speech strings intersect substantially due to cellular  
517 responsiveness to low-level auditory features. Intersecting manifolds are then incrementally  
518 untangled and reduced in dimensionality across the auditory pathway. Panel C shows an  
519 intermediate, ‘low-capacity’ system in which residual manifold tangling is evident. Panel D  
520 shows an optimal system with distributed speech representations that accommodate  
521 variability in the speech stream, but which are discrete and amenable to forming the focus of  
522 attention. The dotted line in panels C and D illustrates a simulated attentional mechanism  
523 (implicated in both recognition and retrieval) which is overwhelmed (Panel C) or effective  
524 (Panel D) as a function of the precision of activated long-term memories. Adapted from Jones  
525 and Westermann (2022).

526         The major contribution of the mean field theory based manifold analysis method is to  
527 enable us to treat distributed biological and artificial neural activation patterns as continuous  
528 geometric shapes that we can measure. Essentially, the convex hull of the collected response  
529 vectors (i.e., the points in Figure 4A) elicited by a fixed class of stimuli is treated as a single  
530 geometric object. In the current study, we are interested in two geometric quantities of neural  
531 representation that have received significant attention in the computational neuroscience  
532 literature. First, we are interested in the *dimensionality* of the pattern of activation (i.e., the  
533 manifold) underpinning responses to a certain class of spoken words (i.e., all instances of  
534 *tree*). That is, we are interested in how spread out through the neural state space speech  
535 representations are. Second, and relatedly, we are interested in the overlap between  
536 competitor neural representations, such as those underpinning the phonologically similar  
537 words *tree* and *three*. Within the MFTMA literature overlap is quantified in terms of  
538 *classification capacity*, which is derived by calculating the number of speech manifolds that  
539 can be linearly separated from all competitor representations and standardising the result by  
540 network layer size. In a low-capacity system representations are highly overlapping (i.e.,



541 discrete representations involve activity in shared neurons), and the system struggles to use a  
542 linear separator to recognise or retrieve any single representation given this overlap (Figure  
543 4B, Figure 4C). In a high-capacity system, representation dimensionality (and other highly  
544 correlated quantities including manifold radius) has been reduced to a level at which overlap  
545 is low and linear separation is more straightforward (Figure 4D).

546 With these properties in mind, Jones and Westermann (2022) drew a parallel between  
547 variance in a network's classification capacity and the demands placed on human working  
548 memory or attentional systems as a function of the precision of activated long-term  
549 memories. Activated low-precision long-term memories, i.e., memories with high  
550 dimensionality, place high demands on the system and compromise efficient processing,  
551 overwhelming working memory and attention (Figure 4C). On the other hand, activated high-  
552 precision long-term memories, i.e., memories with low dimensionality, place low demands on  
553 the processing system, because procedures including speech recognition and retrieval are  
554 facilitated if the target representation is relatively discrete (Figure 4D).

555 Research in this area, both computational work and work involving humans, points to  
556 potentially domain general transformations in representational structure from low-level  
557 structures such as the auditory nerve to high-level structures such as the peripheral auditory  
558 cortex. Broadly, low-level structures are *noise sensitive*, and so manifolds show extensive  
559 overlap (i.e., high dimensionality representations in a low-capacity system). However, within  
560 both biological and artificial neural processing hierarchies, architectural features such as  
561 pooling functions (where, for instance, a neuron fires if *any* antecedent neuron fires) mean  
562 that early noise sensitive representations become increasingly *speech selective* (Davis &  
563 Johnsruide, 2003; DeWitt & Rauschecker, 2012; Kaas et al., 1999; Okada et al., 2010; Yamins  
564 & DiCarlo, 2016). That is, we go from high-dimension representations in a low-capacity  
565 system early in the pathway, to low-dimension representations in a high-capacity system late

566 in the pathway. The neural population geometry view of this trajectory is illustrated in Figure  
567 4, panels B, C, and D. Jones and Westermann did not present a maturational account of  
568 frequency resolution and speech deficits. Instead, their interest was on explaining variance in  
569 working memory task performance. However, these authors did show that the trajectory  
570 shown in Figure 4 could be disrupted by the addition of broad Gaussian noise to input  
571 representations. Here we intend to build substantially on this work by (i) using cochleagrams  
572 developed expressly to simulate human auditory physiology, and (ii) manipulating  
573 cochleagrams during training in line with known trajectories in the maturation of frequency  
574 discrimination skills, something we believe to be unique to the current study.

575         It is worth noting that we are using a powerful neural network with a large number of  
576 training samples of a relatively small number of word types. In general, these are perfect  
577 conditions for training a highly robust neural network that copes well in the face of input  
578 noise. Our intention throughout this project was to keep our manipulation subtle in line with  
579 the notion of a possibly subtle derailment of a typical maturational trajectory. Indeed, looking  
580 at Figure 3 it is clear that the cochleagrams in epoch 10 retain something of a recognisable  
581 contour across conditions, and it might not be too challenging to visually identify this  
582 particular word, *tree*, from the cochleagrams of certain other words within the 35-word  
583 cohort. We did not, therefore, expect dramatic differential effects in the region of, for  
584 instance, 25% performance accuracy, which is the sort of disparity sometimes seen in  
585 empirical studies using so-called ‘extreme-group designs’, which compare quite severely  
586 language-impaired children to children with strong language skills (see West et al., 2017, for  
587 a criticism of this approach). Instead, we were looking for potentially subtle but consistent  
588 disparities in network optimisation indices and behaviour across conditions that align well  
589 with current behavioural and neurophysiological evidence from children with and without  
590 language learning difficulties.

591 This study was not preregistered. All data and materials required to re-run the  
592 simulations and analyses presented in this manuscript are available from the following  
593 repository: <https://osf.io/x2h8k/>.

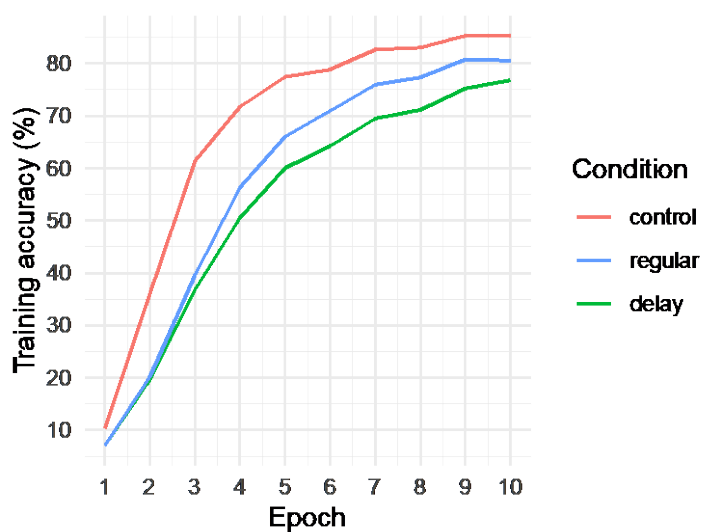
## 594 Results

### 595 *Classification accuracy, probability, and entropy*

596 In the analyses that follow, network performance is collapsed and reported as a  
597 condition mean. Spoken word classification accuracy by condition and training epoch is  
598 shown in Figure 5. Across epochs, networks in the optimal, a-developmental control  
599 condition outperformed the developmental networks in both regular and delay conditions.  
600 Constraining the maturation of high-resolution frequency discrimination according to the  
601 schedules shown in Figure 3 promoted a clear disparity between regular and delay networks,  
602 with the regular networks performing better after epoch two and this gap widening in line  
603 with the disparity in the resolution of spectral information generated by the networks' cochlea  
604 model (Figure 3).

## 605 Figure 5

### 606 *Training Accuracy by Epoch and Condition*



607

608 By epoch ten accuracy averaged 85.3% in the control condition, 80.6% in the regular  
609 condition, and 76.8% in the delay condition. A similar pattern was observed at test, where  
610 speech classification accuracy averaged 85.1% in the control condition, 83.9% in the regular  
611 condition, and 79.6% in the delay condition. During training and at test, accuracy reflects the  
612 networks' ability to correctly classify spoken word cochleagrams. The difference between  
613 these analyses is that training-phase accuracy describes a learning trajectory, while test-phase  
614 accuracy reflects a cross-sectional analysis that is conducted when training is complete.

615         The accuracy data above represents a record of hits as a proportion of total exposures.  
616 However, it is also possible to get a picture of the networks' confidence in their predictions  
617 by analysing the maximum probability assigned to a prediction and the entropy (or spread, in  
618 bits) of the probability distribution output. This analysis indicated greater uncertainty in the  
619 predictions made by networks in the developmental conditions than in the optimal condition,  
620 and greatest uncertainty in networks in the delay condition. Mean maximum probability  
621 assignment stood at 86.7% in the control condition, 81.5% in the regular condition, and  
622 78.6% in the delay condition, while entropy or distribution spread in bits stood at 0.443  
623 control, 0.612 regular, and 0.693 delay (i.e., indicating increasingly spread-out predictive  
624 distributions). A similar pattern was observed when limiting our analysis to hits only: Mean  
625 maximum probability assignment = 91.4% control, 87.2% regular, and 85.3% delay; entropy  
626 in bits = 0.306 control, 0.449 regular, and 0.496 delay.

627         In summary, networks in the maturational delay condition not only performed  
628 significantly less accurately than comparison networks, but also output relatively broad,  
629 highly spread probability distributions over their lexicons, considering many competitor  
630 words and assigning the true target relatively low probability even when accurate. Therefore,  
631 neural networks with maturational deficits in frequency resolution take longer to encode  
632 speech information, and metrics of test performance (i.e., low max probability, high entropy)

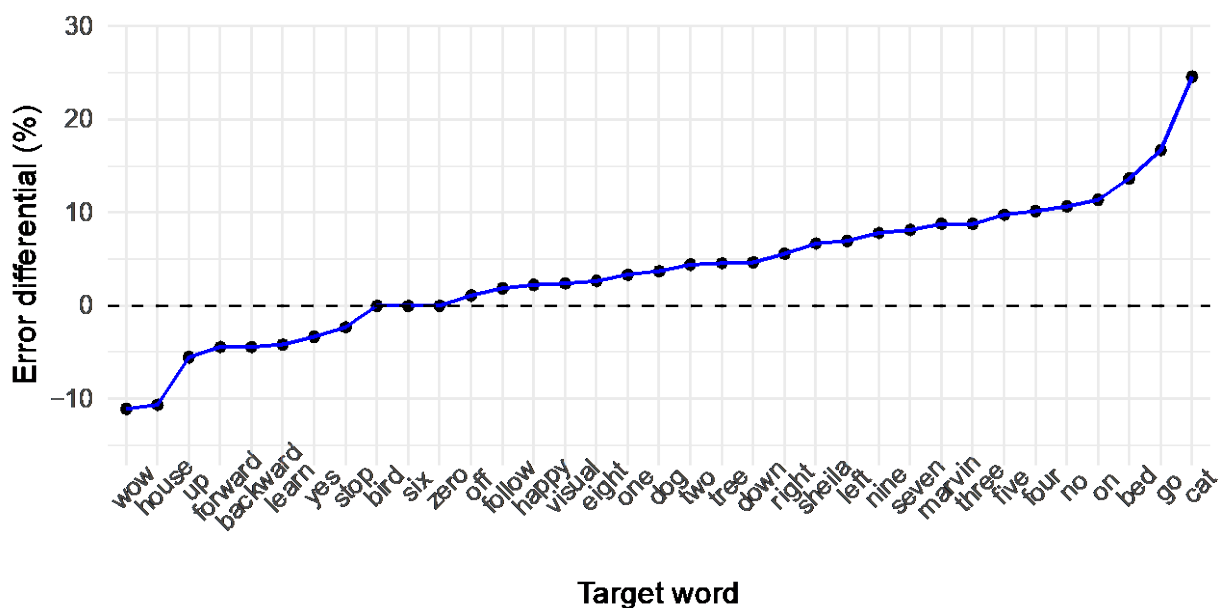
633 suggest that formed speech encodings are inefficiently organised. In response to speech input,  
634 more of what we might consider the networks' long-term memory (i.e., the fixed 35-word  
635 lexicon) becomes activated (i.e., we see high-spread predictive distributions), and the true  
636 target may be swamped in activated competitor representations. Qualitative analogies might  
637 be seen here between network performance and the DLD literature showing: (i) delayed  
638 acquisition profiles (Norbury et al., 2016; a parallel with the disparity in network accuracy  
639 over training epochs); (ii) lower spoken word recognition accuracy (Andreu et al., 2012;  
640 Evans et al., 2018; Rispens et al., 2015; Velez & Schwartz, 2010; a parallel with the network  
641 test-phase accuracy disparity), and; (iii) word finding difficulties and residual uncertainty  
642 even when performing accurately, as evidenced, for instance, in eye tracking paradigms  
643 (Kambanaros et al., 2015; McMurray et al., 2019; Messer & Dockrell, 2006; a parallel with  
644 high entropy, low probability activation patterns). Later, we examine the representational  
645 basis of these performance profiles. First, however, we aimed to determine the particular  
646 words that networks in the regular and delay conditions found difficult to encode and  
647 classify, as well as to understand why networks found these words difficult.

#### 648 *Item-specific effects*

649 We began our item-specific analyses by computing a by-item accuracy differential,  
650 calculated by subtracting the average percentage accurate at test for each word in the delay  
651 condition from the average percentage accurate for each word in the regular condition. The  
652 result is shown in Figure 6. Here, a positive value indicates a performance advantage, as a  
653 percentage, for the regular network, and a negative value indicates a performance advantage  
654 for the delay network. Zero differential indicates no performance difference between  
655 conditions with respect to a particular word.

#### 656 **Figure 6**

#### 657 *Item Accuracy Differential*



658

659 *Note.* All 35 words from the speech commands dataset are shown along the  $x$ -axis. The error  
 660 differential is shown on the  $y$ -axis. A positive differential value signals an advantage (as a  
 661 percentage accurate) for the networks in the regular maturation condition. A negative  
 662 differential value signals an advantage for the networks in the delay condition.

663 Networks in the regular condition outperformed networks in the delay condition with  
 664 respect to 24 out of 35 words, sometimes reaching a differential of 24.6% (for the word *cat*).  
 665 Networks in the delay condition, in contrast, performed better on eight words, with a  
 666 maximum differential of -11.11% for the word *wow*. Clearly, then, error rates vary as a  
 667 function of the target word. To better understand these effects, we looked at confusion  
 668 matrices for predictions made during speech classification in each condition. The top ten  
 669 most confused words in the regular and delay conditions are presented in Table 1 and Table 2  
 670 respectively. These tables show the true word, the total number of misclassifications of that  
 671 word, the most common misclassification of that word, the number of times that the most  
 672 common misclassification occurred, and most common misclassification as a proportion of  
 673 total misclassifications (%).

Table 1

*Top Ten Speech Classification Errors in the Regular Condition.*

Word	Total misclassifications	Most common misclassification	Number	Proportion of total misclassifications (%)
tree	66	three	17	25.76
no	141	go	26	18.44
follow	54	four	7	12.96
go	78	no	10	12.82
up	72	off	9	12.5
house	75	off	8	10.67
four	69	forward	7	10.14
five	123	on	10	8.13
one	90	nine	7	7.78
off	93	on	7	7.53

Table 2

*Top Ten Speech Classification Errors in the Delay Condition.*

Word	Total misclassifications	Most common misclassification	Number	Proportion of total misclassifications (%)
tree	66	three	17	25.76
no	141	go	30	21.28
go	78	no	13	16.67
four	69	forward	10	14.49
five	123	on	16	13.01

on	132	five	16	12.12
right	108	five	10	9.26
two	114	go	10	8.77
three	114	eight	9	7.89
no	141	down	9	6.38

---

674 In many cases, the phonological overlap likely responsible for the misclassification is clear,  
675 for instance with respect to *tree* and *three* or *no* and *go*, and it is noteworthy that networks  
676 struggled by some margin with respect to these particular competitor words. Similar patterns  
677 are discussed by Karimi and Diaz (2020), who review classification disadvantages for near  
678 neighbours under certain experimental conditions. At first glance, then, networks appear to be  
679 broadly sensitive to similar spectral features input as human listeners (e.g., struggling with  
680 items like *tree* and *three*). Yet, Table 1 and Table 2 also illustrate examples which apparently  
681 deviate from this pattern, for instance the apparently high rates of misclassification of the  
682 word *five* as the word *on*, or the misclassification of the word *house* as *off*. It is difficult to  
683 imagine this pattern performance in human participants, and this may attest to the fact that  
684 despite the many gross similarities between processing in artificial neural networks and the  
685 human brain, artificial neural networks may attend to different features of the input in the  
686 service of reducing error in a given task. We return to this question below.

687 To further understand the above disparities in item accuracy between conditions we  
688 fitted a Bayesian regression model in which test phase accuracy (as a percentage) was  
689 predicted by standardised frequency and phonological distance, both in interaction with  
690 condition (i.e., regular, delay). We centred on frequency and phonological distance as  
691 predictor variables given their importance in the child language literature. However,  
692 alternative predictor variables of interest (e.g., orthographic word length) can be  
693 experimented with using the Jupyter notebook and R script associated with this project.

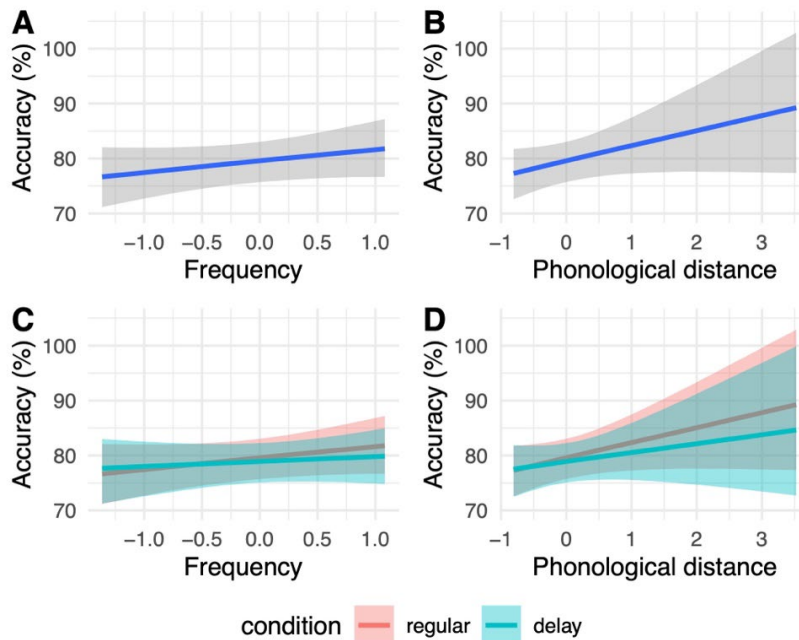


694 Frequency quantified the number of times that a word appeared in the randomly sampled  
695 training data. Meanwhile, phonological distance was computed as the mean optimal string  
696 alignment (OSA) distance between a phonological transcription of each target word and of all  
697 other words in the speech commands corpus.

698         A range of diagnostics showed that this simple regression model with a skew normal  
699 likelihood and weakly informative priors fitted well (i.e., rhats at 1.0, a large number of  
700 effective samples, and credible posterior predictive checks; see supplementary materials and  
701 the brms documentation for further details; Burkner, 2017). Figure 7 shows the estimates  
702 from our Bayesian model.

703 **Figure 7**

704 *Estimates from a Bayesian Model of the Influence of Frequency and Phonological Similarity*  
 705 *on Speech Classification Accuracy*



706

707 Figure 7, panels A and B show that across groups, classification accuracy was on average  
 708 higher for high frequency ( $\beta = 2.11$ ; 95% CI = -0.97 to 5.45), phonologically distinctive ( $\beta =$   
 709  $2.82$ ; 95% CI = -0.46 to 6.46) words. While the credible intervals (CIs) associated with these  
 710 estimates cross zero, indicating that zero may be the true effect, a substantial proportion of  
 711 probability mass is positively assigned, suggesting that a positive association is likely.

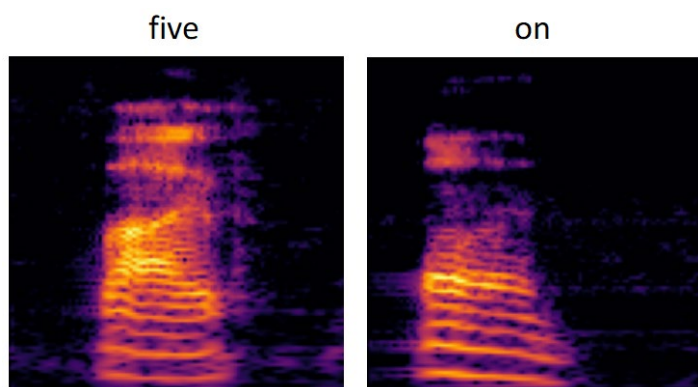
712 Meanwhile, Figure 7, panels C and D show that these effects interact slightly with condition  
 713 but tend in the same positive direction (see R code for full estimates: <https://osf.io/x2h8k/>). In  
 714 each case, networks with rapidly maturing high-resolution cochlea models benefitted slightly  
 715 more from high frequency and greater phonologically distinctiveness.

716 In summary, item specific analyses indicate that while networks struggled to different  
 717 degrees with different words, they nevertheless struggled with broadly similar features of the  
 718 dataset, misclassifying close competitor words such as *tree* and *three* most frequently and  
 719 performing best when words were highly frequent in the training data and phonologically

720 distinctive. Higher resolution low-level auditory representations enabled networks in the  
721 regular condition to better exploit these input statistics. These results may be expected given  
722 that at any particular period the regular and delay networks sit at different points on the same  
723 developmental trajectory. The resulting performance profiles are in agreement with the  
724 general observation that the language of children with DLD is delayed rather than deviant  
725 (Kan & Windsor, 2010; see also *Discussion*). That is, the language of children with DLD is  
726 often similar to that of younger children with typical language skills (though see Bishop,  
727 2014a). That said, our item-specific analysis also revealed potential discrepancies between  
728 artificial neural network and human performance. For instance, we observed a high rate of  
729 misclassification of exemplars of *five* and *on* (see also the *house* and *off* misclassification  
730 rate), which at face value would appear unlikely in human participants. If, however, we look  
731 at representative raw cochleagrams of the words *five* and *on*, for instance, these classification  
732 errors perhaps make more sense (all cochleagrams can be visualised using the associated  
733 scripts). The distributions of energy in the exemplars shown in Figure 8 are at least visually  
734 quite similar, and would of course be even more similar were we depreciate their acuity  
735 across the *y*-axis (for reference, compare the spectral profiles of *five* and *on* to the quite  
736 different profile shown for *tree* in Figure 3).

737 **Figure 8**

738 *Representative Cochleagrams of the Words 'Five' and 'On'*



740 Viewing Figure 8, it may appear reasonable that an artificial neural network would  
741 misclassify degraded instances of *five* and *on*. But how about a human? Of potential  
742 relevance when considering this question is a large research literature looking at so-called  
743 adversarial examples. These are stimuli which, when noise that is typically imperceptible to  
744 humans is added, result in the radical misclassification of those stimuli in an otherwise high-  
745 performing network (Goodfellow et al., 2014; of course, the *y*-axis blur in our study is  
746 perceptible to humans). For instance, an image of a panda with visually imperceptible noise  
747 added to it may be misclassified as a gibbon. Understanding adversarial examples is a vital  
748 part of research on human and machine learning alignment, because it throws light on the  
749 marginal disparities between biological and artificial systems that in many other ways appear  
750 to perform similarly. Intriguingly, there is limited evidence that the same adversarial  
751 examples that derail artificial neural network classification may also affect human  
752 performance, just to a lesser extent and emerging in metrics of classification confidence such  
753 as response time rather than in raw error rates (Elsayed et al., 2018). Two possibilities, then,  
754 are that either the *five* and *on* misclassification error and similar striking errors seen in the  
755 current simulations are evidence the inescapable disparity between artificial and biological  
756 auditory perceptual processing systems, or, on the other hand, that we might be able to elicit  
757 similar patterns of classification behaviour (e.g., extended response times) in humans using  
758 similar stimuli. There is a precedent for this type of work in the domain of visual processing  
759 (Elsayed et al., 2018) but a similar experiment in the domain of auditory processing was  
760 outside the scope of the current project.

### 761 ***Visualising internal representations – Mean field theory based manifold analyses***

762 The cochlea models that provide input to the deep convolutional neural networks used  
763 in these simulations were scheduled to mature according to one of two developmental time  
764 courses. In contrast, the neural networks into which cochleagrams were passed were provided

765 with a randomised initial weight matrix, which was matched across networks and conditions,  
766 but which then optimised freely to solve the specific problems of speech encoding,  
767 recognition, and retrieval. (Note that the control network presents an optimal system which is  
768 free to optimise in the absence of any significant low-level constraint.) The performance  
769 profiles detailed above – specifically the disparities in accuracy, probability, entropy, and  
770 item specific effects – point to systematic differences in dynamic optimization that, given  
771 matching across networks, can result only from these low-level maturational constraints in  
772 high-resolution frequency processing. We are, therefore, modelling discrepancies in optimal  
773 adaptation in the face of different low-level constraints. But what does optimisation in the  
774 face of a low-level frequency discrimination deficit look like? To better understand the  
775 optimisation profiles of networks in our three conditions, and therefore to unpick the  
776 representational basis of the performance discrepancies seen in networks across these  
777 conditions, we turned to mean field theory based manifold analyses.

778 Variables of primary interest were (i) manifold dimensionality and (ii) classification  
779 capacity. Manifold dimensionality quantifies how spread out through a neural state space  
780 long-term speech representations are – i.e., how many artificial neurons (as a proportion of  
781 the layer size) are implicated in the representation of that speech string. Classification  
782 capacity quantifies the number of speech manifolds that can be linearly separated from all  
783 competitor representations, again standardised by network layer size. Analysis of biological  
784 and artificial neural networks suggests that dimensionality decreases across the auditory and  
785 visual perceptual systems, and accordingly, that system capacity increases (Chung et al.,  
786 2018; Chung & Abbott, 2021; DiCarlo & Cox, 2007). This transformation reflects the gradual  
787 de-noising of neural representations in a perceptual hierarchy. Speech representations, for  
788 instance, are shown to become decreasingly noise sensitive and increasingly speech selective  
789 during transformation from the basilar membrane to the peripheral auditory cortex and

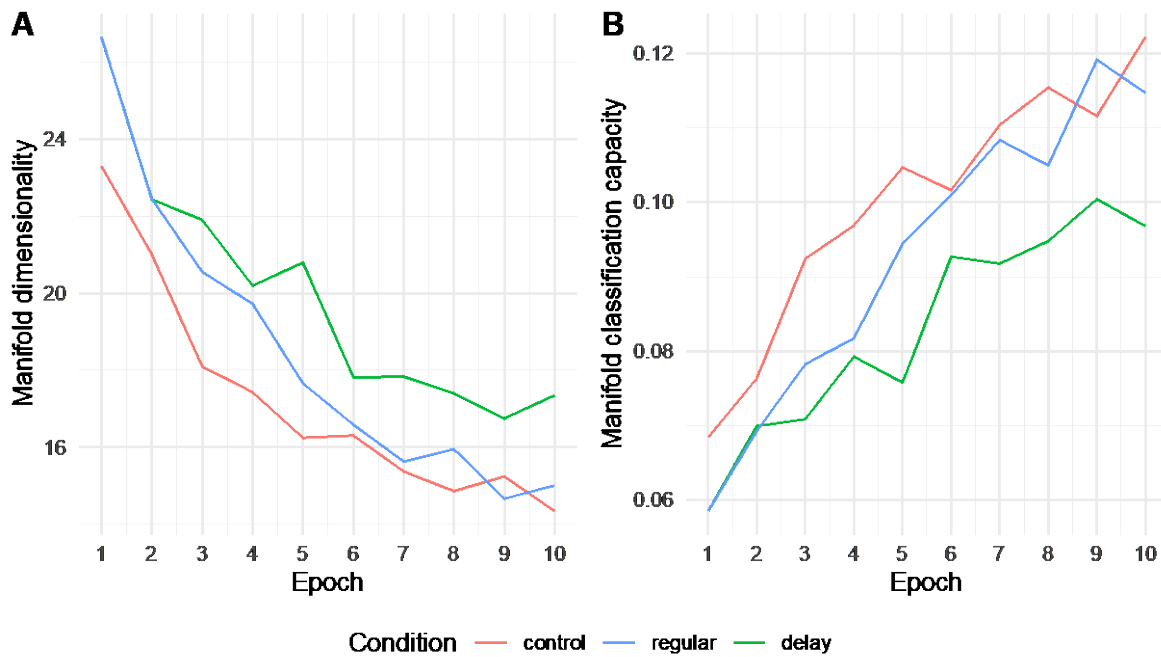
790 beyond (Davis & Johnsruide, 2003; DeWitt & Rauschecker, 2012; Kaas et al., 1999; Okada et  
791 al., 2010).

792 System classification capacity has been interpreted as a measure of not only  
793 representation overlap, but also of attention or working memory load, given that calculating  
794 classification capacity involves linearly discriminating discrete representations from the  
795 system's 'long-term memory' in a manner continuous with cognitive recognition and  
796 retrieval (Jones & Westermann, 2022). This view is in line with so-called state based  
797 frameworks in which working memory is understood as activated long-term memory that  
798 must be optimised to 'fit' within an attentional spotlight (Adams et al., 2018; Oberauer, 2013,  
799 2019). Importantly, reducing manifold dimensionality in order to boost system classification  
800 capacity is a product of training in a given task, here speech encoding and classification.  
801 Training with the same data in a different task, for instance speaker recognition, would result  
802 in an internal network structure optimised for this task (i.e., activation patterns forming  
803 manifolds of speaker voice characteristics; Stephenson et al., 2020). The result of this task-  
804 specific optimization process is presented in Figure 9, which shows the average manifold  
805 dimensionality and classification capacity in networks' penultimate layers antecedent to the  
806 classifier (see Figure 2) as a function of training epoch.

807 **Figure 9**

808 *Changes in Manifold Dimensionality and Classification Capacity During Training*

809



810

811 Figure 9 shows a clear disparity in the optimisation of internal speech representations across  
 812 conditions. Over ten epochs, networks following the regular cochlea maturation schedule  
 813 increasingly approached control standards of optimisation supporting low-dimensional  
 814 representation (Figure 9A). In contrast, despite an overall decrease across epochs, the average  
 815 dimensionality of internal spoken word representations formed in networks in the delay  
 816 condition remained significantly higher, i.e., these representations were substantially more  
 817 ‘spread out’ in a relatively poorly optimised neural state space (Figure 9A). Figure 9, panel B  
 818 shows that this inability to optimize efficiently and reduce manifold dimensionality had a  
 819 severe effect on the delay networks’ ability to retrieve any single representation from their  
 820 internal ‘long-term memory’ systems – what we interpret here as a form of simulated  
 821 working memory or attentional capacity deficit. In essence, the delay networks optimised to  
 822 noise, and this means that the artificial neural response patterns underpinning the long-term  
 823 representations of different spoken words intersect substantially, making efficient recognition  
 824 and retrieval difficult. Graphically, it is as though the delay networks remain in the  
 825 suboptimal state shown in Figure 4C, rather than approaching the relatively optimal state  
 826 shown in Figure 4D alongside networks in the regular and control conditions.

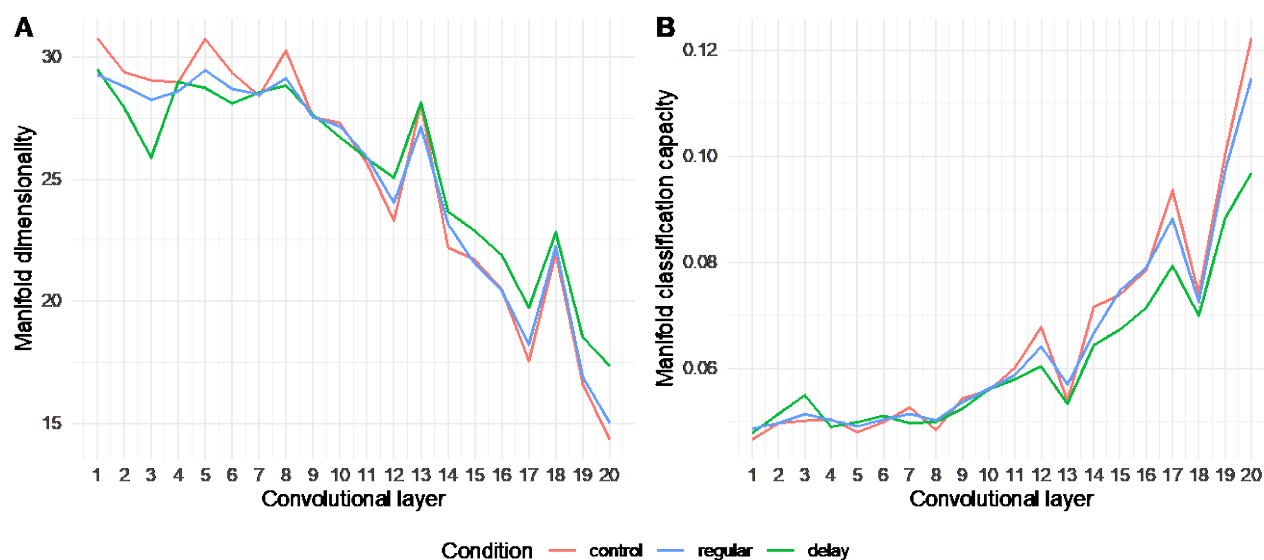
827           The same representational disparity can be seen post-training across the networks'  
828 layers. In Figure 10 we show the previously reported trajectory (e.g., Yamins & DiCarlo,  
829 2016) across the auditory processing hierarchy from high-dimensional manifolds in a low-  
830 capacity system to low-dimensional manifolds in a high-capacity system. Again, this reflects  
831 the system optimising to render initially noise sensitive representations (i.e., waveform  
832 representations containing speaker effects, etc.) increasingly speech selective (i.e., word type  
833 representations in the 35-word lexicon). There is, however, a clear optimisation disparity  
834 between networks in the delay condition and networks in the control and regular conditions in  
835 terms of both dimensionality and classification capacity at higher levels of the processing  
836 hierarchy. This again demonstrates that due to maturational constraints in the cochlea model,  
837 networks in the delay condition failed to learn those spectral features of the speech input that  
838 are essential to effective speech encoding, recognition, and retrieval, with noise permeating  
839 the system and attentional capacity overwhelmed accordingly (Figure 10B). This can be seen  
840 most clearly in Figure 10 with respect to layers 19 and 20, where delay networks deviate  
841 sharply from the regular and control networks with respect to both dimensionality and  
842 classification capacity.

843 **Figure 10**

844 *Manifold Dimensionality and Classification Capacity Across the Layers of Trained Networks*

845





846

847 Finally, we observed optimisation disparities even when regular and delay networks

848 were matched on performance accuracy. In Figure 11 we show manifold dimensionality and

849 manifold classification capacity as a function of training-phase accuracy, by network

850 condition. As in Figure 9, manifold dimensionality and classification capacity are computed

851 for the networks' final convolutional layer (see Appendix), which is antecedent to the 35-way

852 classifier. Despite very occasional overlap, manifold dimensionality is high and classification

853 capacity is low in the delayed networks relative to the regular networks even when networks

854 in these conditions perform with similar accuracy. This result demonstrates the importance of

855 scrutinising the internal representations that artificial neural networks form. Based on

856 accuracy alone we may have wrongly inferred that networks were achieving that level of

857 performance in the same task in the same way, overlooking important differences in the

858 standards of internal optimisation. The finding of representational deficits despite matched

859 levels of performance echoes Bishop and McArthur's reports of electrophysiological

860 discrepancies between children with and without DLD even when DLD-group performance is

861 at threshold (Bishop &amp; McArthur, 2005; McArthur &amp; Bishop, 2004; see also Mengler et al.,

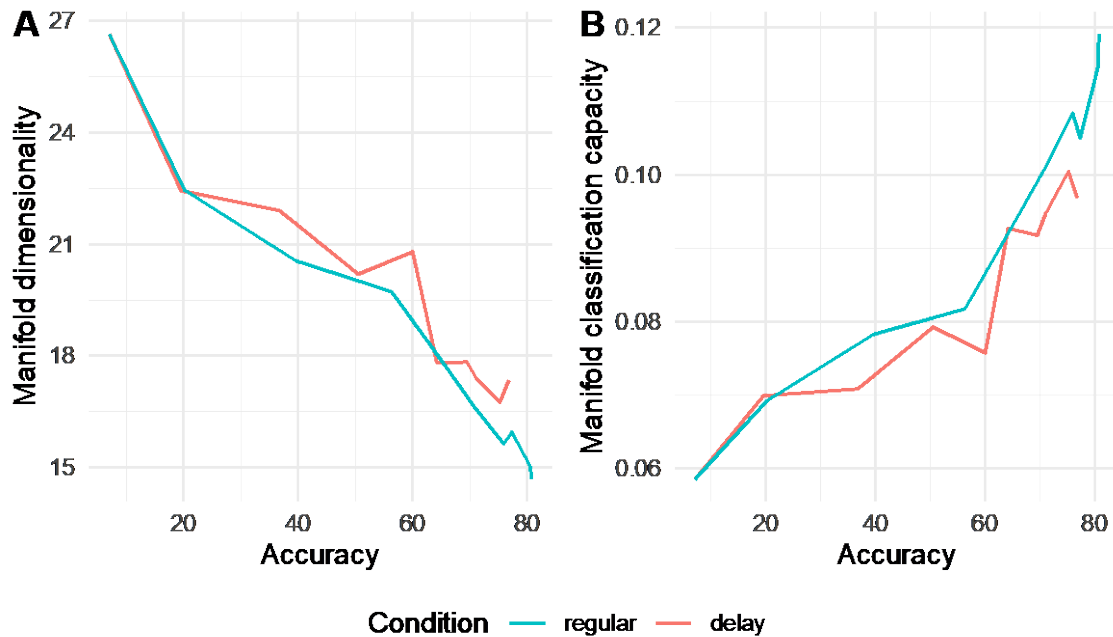
862 2005) and Chonchaiya et al.'s (2013) evidence that signatures of poor auditory brainstem

863 optimisation are predictive of language outcomes. This reaffirms the important point that

864 apparent successes in task performance may not be underpinned by similar qualities of  
 865 learning, a point also made by McMurray et al. (2012).

866 **Figure 11**

867 *Manifold Dimensionality and Classification Capacity by Performance Accuracy*



868

869 In summary, these simulations illustrate how dynamic adaptation to biologically  
 870 plausible models of cochlea function that mature at different rates results in different  
 871 optimization profiles, which underpin disparities in key performance metrics (i.e., accuracy,  
 872 max probability assignment, and entropy) and which are evident despite performance  
 873 accuracy matching (Figure 11). By constraining the development of high-resolution  
 874 frequency discrimination, we curtailed the systems' ability to optimise to encode the key  
 875 spectral features of the speech input that are integral to solving the task at hand, namely  
 876 speech recognition and retrieval. The performance of networks in the delayed condition in  
 877 this study makes the prediction that the optimization profile of a biological speech encoding  
 878 system with a low-level frequency discrimination deficit will show high dimensional speech  
 879 representations (i.e., relatively dispersed neural activation patterns on exposure to speech  
 880 stimuli) which intersect with competitor speech representations, and which are, therefore, not

881 amenable to forming an effective focus of attention. Apparent attention deficits then emerge  
882 as a result of being thinly spread rather than atypically capacity limited. Prior work involving  
883 typically developing adults has shown that this prediction regarding divergent neural  
884 activation patterns is in principle testable in language disordered populations (Davis &  
885 Johnsruide, 2003; DeWitt & Rauschecker, 2012; Kaas et al., 1999; Okada et al., 2010).  
886 Indeed, as described in our literature review, there is already some evidence from language  
887 disordered populations that is broadly continuous with this claim. For instance, low quality,  
888 ‘fuzzy’, speech representations are well documented in the behavioural literature looking at  
889 children with DLD (Claessen et al., 2009, 2013; Claessen & Leitão, 2012a, 2012b), and  
890 atypical neurophysiological signatures indicating suboptimal auditory pathway optimisation  
891 that is predictive of language impairment have been reported in a number of studies (Bishop  
892 & McArthur, 2005; Chonchaiya et al., 2013; McArthur & Bishop, 2004).

893

### Discussion

894 Frequency discrimination deficits are widely recognised among children with  
895 language learning difficulties (Bishop & McArthur, 2005; McArthur & Bishop, 2004;  
896 Mengler et al., 2005). Yet, the nature of these deficits and their relation to speech processing  
897 problems remain unclear. The neural microarchitecture supporting high resolution frequency  
898 discrimination matures from the prenatal period through to later childhood, and it is possible  
899 that the frequency discrimination deficits seen among some children with language learning  
900 difficulties stems from a disruption to this typical developmental trajectory (Bishop &  
901 McArthur, 2005; McArthur & Bishop, 2004). Given that frequency tuning throughout the  
902 auditory pathway is predominantly attributable to the structural properties of the basilar  
903 membrane (i.e., the membrane’s *mechanical* gradient, including fiber diameter, density, and  
904 regularity; Tani et al., 2021), we hypothesised that the protracted maturation of the structural  
905 properties of the basilar membrane may provide a good starting point for inquiry into the

906 source of frequency discrimination deficits in children with neurodevelopmental disorder.  
907 Disruption to the structure of the basilar membrane has been demonstrated empirically in  
908 animal models manipulating emilin 2 expression, which results in a deficient mechanical  
909 gradient and therefore suboptimal functioning of the auditory pathway not supporting high-  
910 resolution frequency processing (Amma et al., 2003; Russell et al., 2020).

911 We developed this theoretical account through a series of computational simulations  
912 of speech encoding, recognition, and retrieval. The networks used in these simulations  
913 incorporated inner ear models developed to replicate human cochlea function (McDermott &  
914 Simoncelli, 2011) that were fed into deep convolutional neural networks. Despite many  
915 important differences, for instance in scale, complexity, and the use of undifferentiated cell  
916 types, deep convolutional neural networks have demonstrated significant correspondences  
917 with human behavioural and neural responses across a range of tests of audition including  
918 speech localization, pitch perception, and hearing in noise (Francel & McDermott, 2022; Kell  
919 et al., 2018; Saddler et al., 2021). Our own innovation was to configure the cochlea models  
920 that formed a fundamental component of our networks to mature according to different  
921 developmental trajectories (i.e., baseline or optimal, regular, and delayed), and to analyse  
922 how the subsequent auditory-linguistic pathway optimised in the service of speech encoding,  
923 recognition, and retrieval.

924 Our analysis of networks in the delayed cochlea maturation condition qualitatively  
925 replicated the linguistic behaviour and neurophysiology of individuals with language learning  
926 difficulties in a number of ways, showing: (i) delayed acquisition profiles (Norbury et al.,  
927 2016); (ii) lower spoken word recognition accuracy (Andreu et al., 2012; Evans et al., 2018;  
928 Rispens et al., 2015; Velez & Schwartz, 2010); (iii) word finding and retrieval difficulties and  
929 uncertainty even when performing accurately, as evidenced, for instance, in eye tracking  
930 paradigms (i.e., Kambanaros et al., 2015; McMurray et al., 2019; Messer & Dockrell, 2006);

931 (iv) ‘fuzzy’ long-term speech representations (Claessen et al., 2009, 2013; Claessen & Leitão,  
932 2012a, 2012b) and neurophysiological signatures of immature neural optimisation that are  
933 associated with speech and language difficulties (Bishop & McArthur, 2005; Chonchaiya et  
934 al., 2013; McArthur & Bishop, 2004); and (v) apparent working memory and attention  
935 deficits that are attributable, we believe, to the imprecision of long-term speech  
936 representations (Gray et al., 2019; Henry & Botting, 2017; Jones & Westermann, 2022). Our  
937 results illustrate that optimising to low-level, low-resolution spectral representations  
938 significantly curtails the capacity of the system to form speech representations supporting  
939 efficient recognition and retrieval.

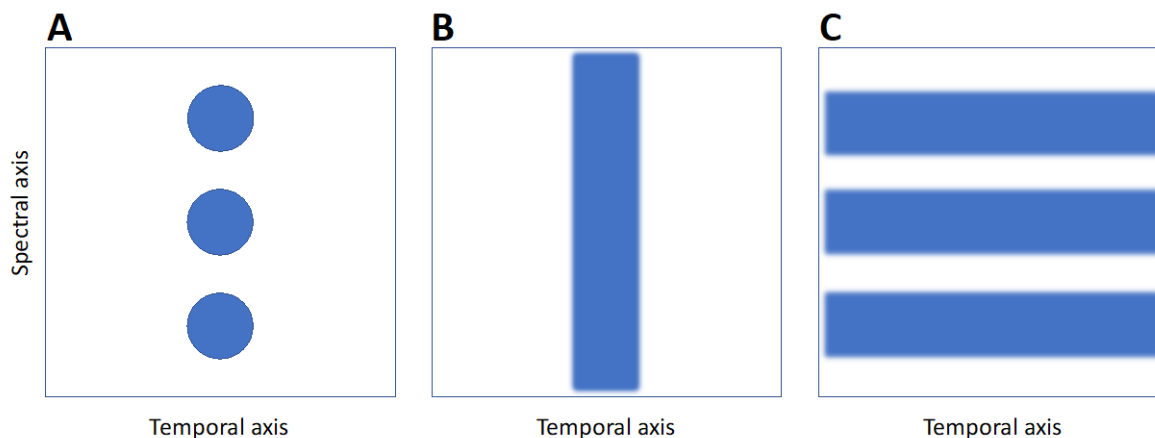
940         We see, then, that some of the mechanisms widely thought to play a causal role in  
941 speech and language disorder may ‘come for free’ if we assume a low-level frequency  
942 discrimination deficit. This includes not only the hypothesised working memory capacity  
943 bottleneck (Archibald & Gathercole, 2006), which dominates DLD research but which we  
944 have argued to be a possible epiphenomenon (see also Jones & Westermann, 2022), but also  
945 the so-called lateral inhibition deficit suggested by McMurray et al. (2019). McMurray et al.  
946 (2019) argue that a key feature of early language disorder may be an inability to inhibit  
947 activated competitor representations during speech recognition in retrieval. Our simulations  
948 suggest, however, that an apparent lateral inhibition deficit may be an emergent characteristic  
949 of a suboptimal auditory processing hierarchy. Networks in the delayed cochlea maturation  
950 condition of our simulations uniformly output predictive distributions with high spread (i.e.,  
951 high entropy) and low maximum probability assignment, signalling heightened uncertainty  
952 and broader activation of the lexicon in response to speech stimuli. As in the case of the  
953 hypothesised working memory capacity limitation, then, we believe that evidence offered in  
954 support of a deficit in a functionally discrete lateral inhibition mechanism may instead reflect

955 target isolation being overwhelmed due to the imprecision of activated long-term speech  
956 representations; a process illustrated in Figure 4C.

957         It may be argued that the results presented in the current study were inevitable. That  
958 is, that disrupting the quality of the cochlea representations that networks could form would  
959 necessarily lead to worse performance. But this is not the case. Indeed, data disruption, for  
960 instance blurring, skewing, re-colouring, or clipping the training data is regularly used in  
961 machine learning, where the process is termed ‘data augmentation’, to boost network  
962 performance by preventing overfitting and attenuating attention to consistent features  
963 (Chollet, 2021). The discrepancies in network performance seen in the current study are,  
964 therefore, attributable to the specific features that we degraded – i.e., frequency information  
965 distributed across the  $y$ -axis – being essential to the efficient encoding and therefore  
966 recognition and retrieval of natural speech. Feature importance is graphically illustrated in  
967 Figure 12. In Panel A we show three dots, exemplifying schematic features that may help us  
968 to classify a particular stimulus. In our case the dots in Figure 12 represent the distinctive  
969 frequency components of a speech string. If, as seen in Panel B, we were to degrade this  
970 stimulus across the  $y$ -axis (i.e., the frequency dimension) this would – as demonstrated in the  
971 current study – cause problems in determining the identity of that stimulus. On the other  
972 hand, degrading the same stimulus across the  $x$ -axis (i.e., the temporal dimension) preserves  
973 the stimulus’ critical features.

## 974 **Figure 12**

975 *Degrading Critical Features*



976

977 That is not to say that the  $x$ -axis degradation seen in Figure 12 Panel C, has no effect. Indeed,  
978 work by Saddler et al. (2021) and Saddler and McDermott (2022) has shown that  
979 manipulating auditory nerve firing rates to degrade temporal information has a significant  
980 negative effect on sound localisation and voice recognition. The point is, then, that when it  
981 comes to encoding speech efficiently specifically for the purposes of accurate recognition and  
982 retrieval, low-level auditory representations with high-resolution, discrete frequency  
983 components appear essential. And, as we have highlighted throughout this article, there is  
984 good evidence that high-resolution frequency discrimination is a core problem among some  
985 children with language learning difficulties.

986 The above discussion of the concept of feature importance may bring some light to  
987 the debate regarding whether the auditory processing deficits seen among some children with  
988 neurodevelopmental disorders are spectral (i.e., frequency-based) or temporal in nature. As  
989 discussed in our introduction, the early dominant view in DLD research was that the  
990 performance deficits seen are temporal in nature, but this view has weakened considerably in  
991 the face of failed replications (Strong et al., 2011; Bishop & McArthur, 2005; McArthur &  
992 Bishop, 2004; see Rosen, 2003, for review). In contrast, there is compelling evidence that the  
993 auditory processing deficits seen among some children with language problems are spectral  
994 in nature (Bishop & McArthur, 2005; McArthur & Bishop, 2004; Mengler et al., 2005).

995 Computational simulation indicates that both spectral and temporal information are crucial to

996 effective speech processing, but that the relative importance of these cues is differentially  
997 weighted as a function of the task. Temporal acuity is vital, for instance, in the context of  
998 voice recognition and sound localisation (Saddler et al., 2021; Saddler & McDermott, 2022).  
999 Yet when it comes to encoding speech for the purposes of recognition and retrieval, the  
1000 current simulations show that high frequency component acuity is key.

1001         It may also be argued that, had we allowed the cochlea models of our delayed  
1002 networks to continue maturing until they reach the same standard as the cochlea models of  
1003 our regular networks, network optimisation and therefore task performance may have  
1004 eventually normalised. This is true, and reflects the fact that artificial neural networks are not  
1005 bound by any hard and fast sensitive period or maturational constraints on physiology<sup>3</sup>.  
1006 Language problems are, in contrast, often evident across the lifespan, suggesting long-lasting  
1007 disparities in the organisation of neural substrates supporting audition and speech. If we take  
1008 a maturational view of frequency discrimination and speech and language deficits, then, the  
1009 critical questions are when and how the typical dynamic adaptation of the auditory pathway  
1010 becomes ‘frozen’ in a sub-optimal state. This appears particularly puzzling given that the  
1011 auditory pathway is, in general, highly plastic, for instance often adapting quickly to the  
1012 fitting of a cochlear implant (e.g., Wang et al., 2021). One possibility is that the mechanical  
1013 gradient of the basilar membrane (and, therefore, tonotopic sensitivity in membrane-posterior  
1014 structures) never reaches optimal differentiation, as in our delayed networks. However, the  
1015 locus of deficit may of course reside in any of the structures posterior to the cochlea that also  
1016 support tonotopic mapping. For instance, Bishop and McArthur (2005) note that while the  
1017 cochlea is typically fully developed by full-term birth, the auditory brainstem and subsequent  
1018 structures continue to adapt through childhood, with frequency discrimination skills

---

<sup>3</sup> That said, we note that sensitive periods may stem from entrenchment rather than biological ossification, and can therefore emerge in computational systems (Thomas & Johnson, 2006).



1019 improving accordingly. Bishop and McArthur (2005) hypothesise, therefore, that either (i) the  
1020 delayed optimisation of higher-level structures within the auditory pathway, including the  
1021 auditory cortex, may be protracted and then plateau with the onset of puberty, or (ii) that  
1022 structures of the auditory pathway that support high-resolution frequency tuning may develop  
1023 slowly but nevertheless fully, yet the cost of a protracted period of maturation during the  
1024 initial phases of language development may be long lasting.

1025         In this study, we have demonstrated how the auditory linguistic pathway may  
1026 optimise in the face of a cochlea maturation deficit. The basilar membrane remains in our  
1027 view a good starting point for future inquiry, because the deficits we see among children with  
1028 DLD are spectral in nature and because the basilar membrane is the seat of tonotopic  
1029 organisation throughout the auditory pathway. We also hypothesised that, given that emilin 2  
1030 plays a key role in the emergence of the development of the mechanical gradient of the  
1031 basilar membrane (Amma et al., 2003; Russell et al., 2020; Tani et al., 2021), potential  
1032 disruption to the expression of this gene might be considered (though we cite the emilin 2  
1033 literature primarily to emphasise how a genetic abnormality can in principle disrupt the  
1034 emergence of the mechanical gradient of the basilar membrane). Yet, given the enormous  
1035 complexity of the auditory pathway, numerous possibilities obviously remain. If, through  
1036 empirical testing, a maturational account is ruled out, it will be necessary to look beyond an  
1037 early ‘freezing’ of typical cochlea, auditory brainstem, and auditory cortex maturation, and to  
1038 instead identify deviances in auditory pathway develop that could give rise to low-resolution  
1039 frequency processing, for instance testing for mid-frequency sensorineural hearing loss (i.e.,  
1040 ‘cookie-bite’ hearing loss; see Ahmadmehrabi et al., 2022, for an adult study) that signals  
1041 problems with the cochlea or auditory nerve, or identifying cortical dysplasia in neural  
1042 substrates supporting audition and speech (Bishop, 2014b).

1043           An important feature of the current study was to let our networks develop over time,  
1044 using cochlea models that output representations of increasing spectral acuity according to  
1045 different maturational trajectories (Figure 3). This developmental approach to modelling with  
1046 neural networks is uncommon, though it is continuous with a limited number of connectionist  
1047 studies that have let their networks develop as a function of experience (e.g., Elman, 1993;  
1048 Westermann et al., 2006; Westermann & Ruh, 2012). We believe that such an approach is  
1049 integral to the study of the developing brain and mind. Similar work is being conducted by  
1050 Skelton (2022), who has developed a filter to simulate changes in the visual system during  
1051 the neonatal period and infancy, which can be used in both experimental stimulus design and  
1052 in computational models of neuro-cognitive development. This development-driven approach  
1053 to computational modelling is likely to provide us with a much richer understanding of the  
1054 emergence of human cognitive behaviour, relative to methods fundamentally aligned with a-  
1055 developmental adult norms.

1056           Like any method the use of artificial neural networks to understand human brain  
1057 function and behaviour has its limitations. Neural networks are, of course, a dramatic  
1058 simplification of the structure of the human brain, involving drastically fewer cells of  
1059 identical, undifferentiated types, with activation functions allowing the communication of  
1060 real numbers. What is more, biological and artificial neural networks learn differently. For  
1061 instance, biological neural networks appear not to need thousands of labelled exemplars in  
1062 order to learn spoken words (Lake et al., 2013; though see Lillicrap et al., 2020, for how the  
1063 brain might approximate the backpropagation algorithm used in our neural networks). These  
1064 architectural and algorithmic differences may underpin different performance profiles – the  
1065 high misclassification rates with respect to *five* and *on* in our data might be a case in point  
1066 here. Nevertheless, gross parallels between human performance and brain function and deep  
1067 neural network activation patterns and performance have been observed repeatedly (Kell et

1068 al., 2018; McDermott & Simoncelli, 2011; Saddler et al., 2021; Yamins & DiCarlo, 2016),  
1069 and a reasonable qualitative mapping with the empirical data in the current study further  
1070 supports this approach.

1071         Modelling of the form presented here of course constitutes a counterpart to, and not  
1072 replacement of, human assessment. Modelling forces us to be explicit about our assumptions,  
1073 and – as we have demonstrated – may provide computational insight into the nature of  
1074 representation, recognition, and retrieval within dynamic systems that have optimised to  
1075 different fundamental constraints. Of course, further analysis involving humans is vital. There  
1076 have already been important steps in this direction, with Chonchaiya et al. (2013) showing  
1077 that neural signatures of immature auditory brainstem organisation are indicative of poorer  
1078 language outcomes – a finding highly in agreement with the hypothesis developed in the  
1079 current paper. To date, however, many studies of children with a diagnosis of DLD have  
1080 included only rudimentary auditory assessments involving, for instance, backward masking,  
1081 mismatch negativity, or glide discrimination, which can show significant variability before  
1082 around eight years of age (Bishop et al., 2005; Bishop & McArthur, 2005; Sutcliffe et al.,  
1083 2006). One particularly elegant example of the inadequacy of such approaches comes from  
1084 research demonstrating that children diagnosed with attention deficit hyperactivity disorder  
1085 (ADHD) can complete pure tone discrimination tasks when taking their medication but not  
1086 when off their medication (Sutcliffe et al., 2006). This highlights the susceptibility of such  
1087 tasks to non-auditory perceptual influences, including attention. Given the ubiquity of  
1088 apparent auditory processing problems not only among children diagnosed with DLD but also  
1089 across other early neurodevelopmental disorders such as developmental dyslexia, there is  
1090 strong justification for a large-sample study involving rich early auditory assessments  
1091 (including, for instance, extended high-frequency audiometry), longitudinal neuroimaging,  
1092 and the assessment of later language outcomes.

1093           The speech commands dataset was chosen for this project because it is free and  
1094 openly available, and because it is unique in comprising such a large number of natural  
1095 speech exemplars. One limitation of this resource, however, is that it comprises only 35 word  
1096 types, meaning that only limited insight can be drawn from our item-specific analyses. While  
1097 we believe that the use of the speech commands dataset in the current project is well justified,  
1098 going forward it would be useful to replicate our findings using a larger dataset. In particular,  
1099 it would be valuable to test children and artificial neural networks using the same speech  
1100 stimuli, which could be recorded specifically for this purpose. This would support a relatively  
1101 direct comparison between child and artificial neural network behaviour. Indeed, using this  
1102 approach it would be possible to simulate real-world language interventions and to determine  
1103 the computational basis of their efficacy.

#### 1104 **Conclusion**

1105           Frequency discrimination is a core problem for many children with language  
1106 learning difficulties, and through computational simulation we have shown how this deficit  
1107 would propagate problems with the encoding, recognition, and retrieval of natural speech.  
1108 Our simulations provide proof of concept that the optimisation of the auditory-linguistic  
1109 pathway to low-resolution cochlea representations – part of a typical maturational trajectory  
1110 that may be protracted in DLD – result in patterns of linguistic behaviour that align  
1111 qualitatively with a range of empirical findings observed among children with DLD. Our  
1112 speculation that the locus of such deficits may be a disruption to the maturation of the basilar  
1113 membrane during a sensitive period of auditory pathway optimisation reflects the fact that the  
1114 mechanical gradient of the basilar membrane provides the basis for the emergence of  
1115 frequency sensitivity across the auditory-linguistic pathway. Yet, this hypothesis of course  
1116 requires empirical testing. The auditory-linguistic pathway is a highly complex system which  
1117 could be disrupted at any level. Also in need of further scrutiny is our speculation, given the

1118 contemporary animal model literature, that atypicalities in emelin 2 expression may be  
1119 implicated in the disruption of the emergence of the mechanical gradient of the basilar  
1120 membrane (i.e., the development of fibril microarchitecture supporting high resolution  
1121 processing, which promulgates the required tonotopic sensitivity through the auditory nerve,  
1122 brainstem, and cortex). We fully recognise these elements of our argument to be speculation,  
1123 albeit empirically driven speculation. Our view is simply that the weight of empirical  
1124 evidence with respect to structural changes in the basilar membrane suggests that this  
1125 hypothesis constitutes a strong starting point for further inquiry into the nature of auditory  
1126 processing deficits in children with language learning difficulties.

## References

- 1127  
1128 Adams, E. J., Nguyen, A. T., & Cowan, N. (2018). Theories of working memory: Differences  
1129 in definition, degree of modularity, role of attention, and purpose. *Language, Speech,*  
1130 *and Hearing Services in Schools, 49*(3), Article 3.  
1131 [https://doi.org/10.1044/2018\\_LSHSS-17-0114](https://doi.org/10.1044/2018_LSHSS-17-0114)
- 1132 Ahmadmehrabi, S., Li, B., Epstein, D. J., Ruckenstein, M. J., & Brant, J. A. (2022). How  
1133 Does the “Cookie-Bite” Audiogram Shape Perform in Discriminating Genetic  
1134 Hearing Loss in Adults? *Otolaryngology–Head and Neck Surgery, 166*(3), 537–539.  
1135 <https://doi.org/10.1177/01945998211015181>
- 1136 Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of  
1137 frequency effects in first language acquisition. In *Journal of Child Language.*  
1138 <https://doi.org/10.1017/S030500091400049X>
- 1139 Amma, L. L., Goodyear, R., Faris, J. S., Jones, I., Ng, L., Richardson, G., & Forrest, D.  
1140 (2003). An emilin family extracellular matrix protein identified in the cochlear basilar  
1141 membrane. *Molecular and Cellular Neuroscience, 23*(3), 460–472.  
1142 [https://doi.org/10.1016/S1044-7431\(03\)00075-7](https://doi.org/10.1016/S1044-7431(03)00075-7)
- 1143 Andreu, L., Sanz-Torrent, M., & Guàrdia-Olmos, J. (2012). Auditory word recognition of  
1144 nouns and verbs in children with specific language impairment (SLI). *Journal of*  
1145 *Communication Disorders, 45*(1), Article 1.  
1146 <https://doi.org/10.1016/j.jcomdis.2011.09.003>
- 1147 Archibald, L. M. D., & Gathercole, S. E. (2006). Short-term and working memory in specific  
1148 language impairment. *International Journal of Language and Communication*  
1149 *Disorders, 41*(6), Article 6. <https://doi.org/10.1080/13682820500442602>

- 1150 Archibald, L. M. D., & Harder Griebeling, K. (2016). Rethinking the connection between  
1151 working memory and language impairment. *International Journal of Language &*  
1152 *Communication Disorders*, 51(3), Article 3. <https://doi.org/10.1111/1460-6984.12202>
- 1153 Astle, D. E., Holmes, J., Kievit, R., & Gathercole, S. E. (2022). Annual Research Review:  
1154 The transdiagnostic revolution in neurodevelopmental disorders. *Journal of Child*  
1155 *Psychology and Psychiatry*, 63(4), Article 4. <https://doi.org/10.1111/jcpp.13481>
- 1156 Barman, A., Prabhu, P., Mekhala, V. G., Vijayan, K., & Narayanan, S. (2021).  
1157 Electrophysiological findings in specific language impairment: A scoping review.  
1158 *Hearing, Balance and Communication*, 19(1), 26–30.  
1159 <https://doi.org/10.1080/21695717.2020.1807277>
- 1160 Bishop, D. V. M. (2006). What causes specific language impairment in children? *Current*  
1161 *Directions in Psychological Science*, 15(5), Article 5. [https://doi.org/10.1111/j.1467-](https://doi.org/10.1111/j.1467-8721.2006.00439)  
1162 [8721.2006.00439](https://doi.org/10.1111/j.1467-8721.2006.00439)
- 1163 Bishop, D. V. M. (2014a). Problems with tense marking in children with specific language  
1164 impairment: Not how but when. *Philosophical Transactions of the Royal Society B:*  
1165 *Biological Sciences*, 369(1634), Article 1634. <https://doi.org/10.1098/rstb.2012.0401>
- 1166 Bishop, D. V. M. (2014b). *Uncommon Understanding (Classic Edition)*. Psychology Press.  
1167 <https://doi.org/10.4324/9780203381472>
- 1168 Bishop, D. V. M., Adams, C. V., Nation, K., & Rosen, S. (2005). Perception of transient  
1169 nonspeech stimuli is normal in specific language impairment: Evidence from glide  
1170 discrimination. *Applied Psycholinguistics*, 26, 175–194.  
1171 <https://doi.org/10.1017.S0142716405050137>
- 1172 Bishop, D. V. M., Bishop, S. J., Bright, P., James, C., Delaney, T., & Tallal, P. (1999).  
1173 Different origin of auditory and phonological processing problems in children with

- 1174 language impairment. *Journal of Speech, Language, and Hearing Research*, 42(1),  
1175 Article 1. <https://doi.org/10.1044/jslhr.4201.155>
- 1176 Bishop, D. V. M., Hardiman, M. J., & Barry, J. G. (2012). Auditory Deficit as a Consequence  
1177 Rather than Endophenotype of Specific Language Impairment: Electrophysiological  
1178 Evidence. *PLoS ONE*, 7(5), Article 5. <https://doi.org/10.1371/journal.pone.0035851>
- 1179 Bishop, D. V. M., & McArthur, G. M. (2005). Individual differences in auditory processing  
1180 in specific language impairment: A follow-up study using event-related potentials and  
1181 behavioural thresholds. *Cortex*, 41(3), Article 3. [https://doi.org/10.1016/S0010-](https://doi.org/10.1016/S0010-9452(08)70270-3)  
1182 [9452\(08\)70270-3](https://doi.org/10.1016/S0010-9452(08)70270-3)
- 1183 Bishop, D. V. M., Snowling, M. J., Thompson, P. A., & Greenhalgh, T. (2016). CATALISE:  
1184 A multinational and multidisciplinary delphi consensus study. Identifying language  
1185 impairments in children. *PLOS ONE*, 11(7), e0158753.  
1186 <https://doi.org/10.1371/journal.pone.0158753>
- 1187 Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2018). *Consistency and*  
1188 *variability in word learning across languages*. <https://doi.org/10.31234/osf.io/cg6ah>
- 1189 Burkner, P.-C. (2017). *brms: Bayesian Regression Models using 'Stan'*. 154.
- 1190 Burnham, D., & Mattock, K. (2014). Auditory development. In G. Bremner & T. Wachs  
1191 (Eds.), *The Wiley Blackwell Handbook of Infant Development* (2nd ed., pp. 83–121).
- 1192 Chollet, F. (2021). *Deep Learning with Python, Second Edition*.
- 1193 Chonchaiya, W., Tardif, T., Mai, X., Xu, L., Li, M., Kaciroti, N., Kileny, P. R., Shao, J., &  
1194 Lozoff, B. (2013). Developmental trends in auditory processing can provide early  
1195 predictions of language acquisition in young infants. *Developmental Science*, 16(2),  
1196 159–172. <https://doi.org/10.1111/desc.12012>



- 1197 Chung, S., & Abbott, L. F. (2021). Neural population geometry: An approach for  
1198 understanding biological and artificial neural networks. *Current Opinion in*  
1199 *Neurobiology*, 70, 137–144. <https://doi.org/10.1016/j.conb.2021.10.010>
- 1200 Chung, S., Lee, D. D., & Sompolinsky, H. (2018). Classification and Geometry of General  
1201 Perceptual Manifolds. *Physical Review X*, 8(3), Article 3.  
1202 <https://doi.org/10.1103/PhysRevX.8.031003>
- 1203 Claessen, M., Heath, S., Fletcher, J., Hogben, J., & Leitão, S. (2009). Quality of phonological  
1204 representations: A window into the lexicon? *International Journal of Language and*  
1205 *Communication Disorders*, 44(2), Article 2.  
1206 <https://doi.org/10.1080/13682820801966317>
- 1207 Claessen, M., & Leitão, S. (2012a). Phonological representations in children with SLI. *Child*  
1208 *Language Teaching and Therapy*, 28(2), Article 2.  
1209 <https://doi.org/10.1177/0265659012436851>
- 1210 Claessen, M., & Leitão, S. (2012b). The relationship between stored phonological  
1211 representations and speech output. *International Journal of Speech-Language*  
1212 *Pathology*, 14(3), Article 3. <https://doi.org/10.3109/17549507.2012.679312>
- 1213 Claessen, M., Leitão, S., Kane, R., & Williams, C. (2013). Phonological processing skills in  
1214 specific language impairment. *International Journal of Speech-Language Pathology*,  
1215 15(5), Article 5. <https://doi.org/10.3109/17549507.2012.753110>
- 1216 Cohen, U., Chung, S. Y., Lee, D. D., & Sompolinsky, H. (2020). Separability and geometry  
1217 of object manifolds in deep neural networks. *Nature Communications*, 11(1), Article  
1218 1. <https://doi.org/10.1038/s41467-020-14578-5>
- 1219 Davis, M. H., & Johnsruide, I. S. (2003). Hierarchical processing in spoken language  
1220 comprehension. *Journal of Neuroscience*, 23(8), 3423–3431.  
1221 <https://doi.org/10.1523/jneurosci.23-08-03423.2003>

- 1222 DeWitt, I., & Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory  
1223 ventral stream. *Proceedings of the National Academy of Sciences of the United States*  
1224 *of America*, 109(8), 505–514. <https://doi.org/10.1073/pnas.1113427109>
- 1225 DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in*  
1226 *Cognitive Sciences*, 11(8), Article 8. <https://doi.org/10.1016/j.tics.2007.06.010>
- 1227 Echteler, S. M., Arjmand, E., & Dallos, P. (1989). Developmental alterations in the frequency  
1228 map of the mammalian cochlea. *Nature*, 341(6238), 147–149.  
1229 <https://doi.org/10.1038/341147a0>
- 1230 Elmahallawi, T. H., Gabr, T. A., Darwish, M. E., & Seleem, F. M. (2021). Children with  
1231 developmental language disorder: A frequency following response in the noise study.  
1232 *Brazilian Journal of Otorhinolaryngology*. <https://doi.org/10.1016/j.bjorl.2021.01.008>
- 1233 Elman, J. L. (1993). Learning and development in neural networks: The importance of  
1234 starting small. *Cognition*, 48(1), Article 1.
- 1235 Elsayed, G. F., Shankar, S., Cheung, B., Papernot, N., Kurakin, A., Goodfellow, I., & Sohl-  
1236 Dickstein, J. (2018). *Adversarial Examples that Fool both Computer Vision and Time-*  
1237 *Limited Humans*. <https://doi.org/10.48550/ARXIV.1802.08195>
- 1238 Evans, J. L., Gillam, R. B., & Montgomery, J. W. (2018). Cognitive Predictors of Spoken  
1239 Word Recognition in Children With and Without Developmental Language Disorders.  
1240 *Journal of Speech, Language, and Hearing Research*, 61(6), Article 6.  
1241 [https://doi.org/10.1044/2018\\_JSLHR-L-17-0150](https://doi.org/10.1044/2018_JSLHR-L-17-0150)
- 1242 Fletcher-Watson, S. (2022). Transdiagnostic research and the neurodiversity paradigm:  
1243 Commentary on the transdiagnostic revolution in neurodevelopmental disorders by  
1244 Astle et al. *Journal of Child Psychology and Psychiatry*, 63(4), Article 4.  
1245 <https://doi.org/10.1111/jcpp.13589>

- 1246 Francl, A., & McDermott, J. H. (2022). Deep neural network models of sound localization  
1247 reveal how perception is adapted to real-world environments. *Nature Human*  
1248 *Behaviour*, 6(1), 111–133. <https://doi.org/10.1038/s41562-021-01244-z>
- 1249 Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). *Explaining and Harnessing Adversarial*  
1250 *Examples*. <https://doi.org/10.48550/ARXIV.1412.6572>
- 1251 Gray, S., Fox, A. B., Green, S., Alt, M., Hogan, T. P., Petscher, Y., & Cowan, N. (2019).  
1252 Working memory profiles of children with dyslexia, developmental language  
1253 disorder, or both. *Journal of Speech, Language, and Hearing Research*, 62(6), Article  
1254 6. [https://doi.org/10.1044/2019\\_JSLHR-L-18-0148](https://doi.org/10.1044/2019_JSLHR-L-18-0148)
- 1255 Haake, C., Kob, M., Willmes, K., & Domahs, F. (2013). Word stress processing in specific  
1256 language impairment: Auditory or representational deficits? *Clinical Linguistics and*  
1257 *Phonetics*, 27(8), Article 8. <https://doi.org/10.3109/02699206.2013.798034>
- 1258 He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Deep residual learning for image recognition*.
- 1259 Henry, L. A., & Botting, N. (2017). Working memory and developmental language  
1260 impairments. *Child Language Teaching and Therapy*, 33(1), Article 1.  
1261 <https://doi.org/10.1177/0265659016655378>
- 1262 Hestvik, A., Epstein, B., Schwartz, R. G., & Shafer, V. L. (2022). Developmental Language  
1263 Disorder as Syntactic Prediction Impairment. *Frontiers in Communication*, 6, 637585.  
1264 <https://doi.org/10.3389/fcomm.2021.637585>
- 1265 Higgins, I., Stringer, S., & Schnupp, J. (2017). Unsupervised learning of temporal features for  
1266 word categorization in a spiking neural network model of the auditory brain. *PLOS*  
1267 *ONE*, 12(8), e0180174. <https://doi.org/10.1371/journal.pone.0180174>
- 1268 Jensen, J. K., & Neff, D. L. (1993). Development of Basic Auditory Discrimination in  
1269 Preschool Children. *Psychological Science*, 4(2), 104–107.  
1270 <https://doi.org/10.1111/j.1467-9280.1993.tb00469.x>

- 1271 Jones, S. D., & Brandt, S. (2019). Do children really acquire dense neighbourhoods? *Journal*  
1272 *of Child Language*, 46(6), 1260–1273. <https://doi.org/10.1017/S0305000919000473>
- 1273 Jones, S. D., & Westermann, G. (2022). Under-resourced or overloaded? Rethinking working  
1274 memory and sentence comprehension deficits in developmental language disorder.  
1275 *Psychological Review*, Advance online publication.  
1276 <http://dx.doi.org/10.1037/rev0000338>
- 1277 Kaas, J. H., Hackett, T. A., & Tramo, M. J. (1999). Auditory processing in primate cerebral  
1278 cortex. *Current Opinion in Neurobiology*, 9(2), 164–170.  
1279 [https://doi.org/10.1016/S0959-4388\(99\)80022-1](https://doi.org/10.1016/S0959-4388(99)80022-1)
- 1280 Kambanaros, M., Michaelides, M., & Grohmann, K. K. (2015). Measuring word retrieval  
1281 deficits in a multilingual child with SLI: Is there a better language? *Journal of*  
1282 *Neurolinguistics*, 34, 112–130. <https://doi.org/10.1016/j.jneuroling.2014.09.006>
- 1283 Kan, P. F., & Windsor, J. (2010). Word Learning in Children With Primary Language  
1284 Impairment: A Meta-Analysis. *Journal of Speech, Language, and Hearing Research*,  
1285 53(3), Article 3. [https://doi.org/10.1044/1092-4388\(2009/08-0248\)](https://doi.org/10.1044/1092-4388(2009/08-0248))
- 1286 Karimi, H., & Diaz, M. (2020). When phonological neighborhood density both facilitates and  
1287 impedes: Age of acquisition and name agreement interact with phonological  
1288 neighborhood during word production. *Memory & Cognition*, 48(6), 1061–1072.  
1289 <https://doi.org/10.3758/s13421-020-01042-4>
- 1290 Kell, A. J. E., Yamins, D. L. K., Shook, E. N., Norman-Haignere, S. V., & McDermott, J. H.  
1291 (2018). A task-optimized neural network replicates human auditory behavior, predicts  
1292 brain responses, and reveals a cortical processing hierarchy. *Neuron*, 98(3), Article 3.  
1293 <https://doi.org/10.1016/j.neuron.2018.03.044>
- 1294 Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley,  
1295 K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., & Willing, C.

- 1296 (2016). Jupyter Notebooks—A publishing format for reproducible computational  
1297 workflows. *Positioning and Power in Academic Publishing: Players, Agents and*  
1298 *Agendas - Proceedings of the 20th International Conference on Electronic*  
1299 *Publishing, ELPUB 2016*, 87–90. <https://doi.org/10.3233/978-1-61499-649-1-87>
- 1300 Lake, B. M., Salakhutdinov, R. R., & Tenenbaum, J. (2013). One-shot learning by inverting a  
1301 compositional causal process. In C. J. Burges, L. Bottou, M. Welling, Z. Ghahramani,  
1302 & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems*  
1303 (Vol. 26). Curran Associates, Inc.  
1304 [https://proceedings.neurips.cc/paper/2013/file/52292e0c763fd027c6eba6b8f494d2eb-](https://proceedings.neurips.cc/paper/2013/file/52292e0c763fd027c6eba6b8f494d2eb-Paper.pdf)  
1305 [Paper.pdf](https://proceedings.neurips.cc/paper/2013/file/52292e0c763fd027c6eba6b8f494d2eb-Paper.pdf)
- 1306 Lillicrap, T. P., Santoro, A., Marris, L., Akerman, C. J., & Hinton, G. (2020).  
1307 Backpropagation and the brain. *Nature Reviews Neuroscience*, 21(6), 335–346.  
1308 <https://doi.org/10.1038/s41583-020-0277-3>
- 1309 Lopez-Poveda, E. A. (2014). Development of Fundamental Aspects of Human Auditory  
1310 Perception. In *Development of Auditory and Vestibular Systems* (pp. 287–314).  
1311 Elsevier. <https://doi.org/10.1016/B978-0-12-408088-1.00010-5>
- 1312 Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and*  
1313 *Processing of Visual Information*. Henry Holt and Co., Inc.
- 1314 McArthur, G. M., & Bishop, D. V. M. (2004). Which People with Specific Language  
1315 Impairment have Auditory Processing Deficits? *Cognitive Neuropsychology*, 21(1),  
1316 79–94. <https://doi.org/10.1080/02643290342000087>
- 1317 McArthur, G. M., & Bishop, D. V. M. (2005). Speech and non-speech processing in people  
1318 with specific language impairment: A behavioural and electrophysiological study.  
1319 *Brain and Language*, 94(3), Article 3. <https://doi.org/10.1016/j.bandl.2005.01.002>

- 1320 McDermott, J. H., & Simoncelli, E. P. (2011). Sound Texture Perception via Statistics of the  
1321 Auditory Periphery: Evidence from Sound Synthesis. *Neuron*, *71*(5), 926–940.  
1322 <https://doi.org/10.1016/j.neuron.2011.06.032>
- 1323 McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the  
1324 interaction of online referent selection and slow associative learning. *Psychological*  
1325 *Review*, *119*(4), Article 4. <https://doi.org/10.1037/a0029872>
- 1326 McMurray, B., Klein-Packard, J., & Tomblin, J. B. (2019). A real-time mechanism  
1327 underlying lexical deficits in developmental language disorder: Between-word  
1328 inhibition. *Cognition*, *191*, 104000. <https://doi.org/10.1016/j.cognition.2019.06.012>
- 1329 Mengler, E. D., Hogben, J. H., Michie, P., & Bishop, D. V. M. (2005). Poor frequency  
1330 discrimination is related to oral language disorder in children: A psychoacoustic  
1331 study. *Dyslexia*, *11*(3), 155–173. <https://doi.org/10.1002/dys.302>
- 1332 Merzenich, M. M., Jenkins, W. M., Johnston, P., Schreiner, C., Miller, S. L., & Tallal, P.  
1333 (1996). Temporal Processing Deficits of Language-Learning Impaired Children  
1334 Ameliorated by Training. *Science*, *271*(5245), Article 5245.  
1335 <https://doi.org/10.1126/science.271.5245.77>
- 1336 Messer, D., & Dockrell, J. E. (2006). Children’s naming and word-finding difficulties:  
1337 Descriptions and explanations. *Journal of Speech, Language, and Hearing Research*,  
1338 *49*(2), Article 2. [https://doi.org/10.1044/1092-4388\(2006/025\)](https://doi.org/10.1044/1092-4388(2006/025))
- 1339 Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., &  
1340 Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical  
1341 presentation of language disorder: Evidence from a population study. *Journal of Child*  
1342 *Psychology and Psychiatry*, *57*(11), Article 11. <https://doi.org/10.1111/jcpp.12573>
- 1343 Novitski, N., Huotilainen, M., Tervaniemi, M., Näätänen, R., & Fellman, V. (2007). Neonatal  
1344 frequency discrimination in 250–4000-Hz range: Electrophysiological evidence.

- 1345           *Clinical Neurophysiology*, 118(2), 412–419.
- 1346           <https://doi.org/10.1016/j.clinph.2006.10.008>
- 1347 Nuttall, A. L., Ricci, A. J., Burwood, G., Harte, J. M., Stenfelt, S., Cayé-Thomasen, P., Ren,  
1348 T., Ramamoorthy, S., Zhang, Y., Wilson, T., Lunner, T., Moore, B. C. J., &  
1349 Fridberger, A. (2018). A mechano-electrical mechanism for detection of sound  
1350 envelopes in the hearing organ. *Nature Communications*, 9(1), 4175.  
1351           <https://doi.org/10.1038/s41467-018-06725-w>
- 1352 Oberauer, K. (2013). The focus of attention in working memory—From metaphors to  
1353 mechanisms. *Frontiers in Human Neuroscience*, 7.  
1354           <https://doi.org/10.3389/fnhum.2013.00673>
- 1355 Oberauer, K. (2019). Working memory and attention – A conceptual analysis and review.  
1356           *Journal of Cognition*, 2(1), Article 1. <https://doi.org/10.5334/joc.58>
- 1357 Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I. H., Saberi, K., Serences, J. T., &  
1358 Hickok, G. (2010). Hierarchical organization of human auditory cortex: Evidence  
1359 from acoustic invariance in the response to intelligible speech. *Cerebral Cortex*,  
1360 20(10), 2486–2495. <https://doi.org/10.1093/cercor/bhp318>
- 1361 Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z.,  
1362 Gimselshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison,  
1363 M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., ... Chintala, S. (2019).  
1364           *PyTorch: An imperative style, high-performance deep learning library*.
- 1365 Pinker, S. (1994). *The language instinct* (1st ed). W. Morrow and Co.
- 1366 Python Software Foundation. (2008). Python. In *Python Language Reference* (No. 3).  
1367           <https://www.python.org/>
- 1368 Rispens, J., Baker, A., & Duinmeijer, I. (2015). Word recognition and nonword repetition in  
1369 children with language disorders: The effects of neighborhood density, lexical

- 1370 frequency, and phonotactic probability. *Journal of Speech, Language, and Hearing*  
1371 *Research*, 58(1), Article 1. [https://doi.org/10.1044/2014\\_JSLHR-L-12-0393](https://doi.org/10.1044/2014_JSLHR-L-12-0393)
- 1372 Rosen, S. (2003). Auditory processing in dyslexia and specific language impairment: Is there  
1373 a deficit? What is its nature? Does it explain anything? *Journal of Phonetics*, 31(3–4),  
1374 Article 3–4. [https://doi.org/10.1016/S0095-4470\(03\)00046-9](https://doi.org/10.1016/S0095-4470(03)00046-9)
- 1375 RStudio Team. (2016). RStudio: Integrated development for R. [Online] RStudio, Inc.,  
1376 Boston, MA URL <Http://Www.Rstudio.Com>. [https://doi.org/10.1007/978-81-322-](https://doi.org/10.1007/978-81-322-2340-5)  
1377 2340-5
- 1378 Russell, I. J., Lukashkina, V. A., Levic, S., Cho, Y.-W., Lukashkin, A. N., Ng, L., & Forrest,  
1379 D. (2020). Emilin 2 promotes the mechanical gradient of the cochlear basilar  
1380 membrane and resolution of frequencies in sound. *Science Advances*, 6(24),  
1381 eaba2634. <https://doi.org/10.1126/sciadv.aba2634>
- 1382 Saddler, M., Gonzalez, R., & McDermott, J. H. (2021). Deep neural network models reveal  
1383 interplay of peripheral coding and stimulus statistics in pitch perception. *Nature*  
1384 *Communications*, 12(1), 7278. <https://doi.org/10.1038/s41467-021-27366-6>
- 1385 Saddler, M., & McDermott, J. (2022, March 19). *The role of temporal coding in everyday*  
1386 *hearing: Evidence from deep neural networks* [Poster]. [https://www.world-](https://www.world-wide.org/cosyne-22/role-temporal-coding-everyday-hearing-ba39ae72/)  
1387 [wide.org/cosyne-22/role-temporal-coding-everyday-hearing-ba39ae72/](https://www.world-wide.org/cosyne-22/role-temporal-coding-everyday-hearing-ba39ae72/)
- 1388 Shafer, V. L., Morr, M. L., Kreuzer, J. A., & Kurtzberg, D. (2000). Maturation of Mismatch  
1389 Negativity in School-Age Children: *Ear and Hearing*, 21(3), 242–251.  
1390 <https://doi.org/10.1097/00003446-200006000-00008>
- 1391 Skelton, A. (2022, August 25). *A digital filter to simulate infant visual experience*. Lancaster  
1392 Conference on Infant and Early Child Development (LCICD), Lancaster, UK.
- 1393 Stephenson, C., Feather, J., Padhy, S., Elibol, O., Tang, H., McDermott, J., & Chung, S.  
1394 (2020). *Untangling in Invariant Speech Recognition*. <http://arxiv.org/abs/2003.01787>



- 1395 Strong, G. K., Torgerson, C. J., Torgerson, D., & Hulme, C. (2011). A systematic meta-  
1396 analytic review of evidence for the effectiveness of the ‘Fast ForWord’ language  
1397 intervention program. *Journal of Child Psychology and Psychiatry*, 52(3), Article 3.  
1398 <https://doi.org/10.1111/j.1469-7610.2010.02329.x>
- 1399 Sumner, C. J., Wells, T. T., Bergevin, C., Sollini, J., Kreft, H. A., Palmer, A. R., Oxenham,  
1400 A. J., & Shera, C. A. (2018). Mammalian behavior and physiology converge to  
1401 confirm sharper cochlear tuning in humans. *Proceedings of the National Academy of*  
1402 *Sciences*, 115(44), 11322–11326. <https://doi.org/10.1073/pnas.1810766115>
- 1403 Sutcliffe, P. A., Bishop, D. V. M., Houghton, S., & Taylor, M. (2006). Effect of Attentional  
1404 State on Frequency Discrimination: A Comparison of Children With ADHD On and  
1405 Off Medication. *Journal of Speech, Language, and Hearing Research*, 49(5), 1072–  
1406 1084. [https://doi.org/10.1044/1092-4388\(2006/076\)](https://doi.org/10.1044/1092-4388(2006/076))
- 1407 Tallal, P. (2013). Fast ForWord®. In *Progress in Brain Research* (Vol. 207, pp. 175–207).  
1408 Elsevier. <https://doi.org/10.1016/B978-0-444-63327-9.00006-0>
- 1409 Tallal, P., Stark, R., Kallman, C., & Mellits, D. (1981). A Reexamination of Some Nonverbal  
1410 Perceptual Abilities of Language-Impaired and Normal Children as a Function of Age  
1411 and Sensory Modality. *Journal of Speech, Language, and Hearing Research*, 24(3),  
1412 351–357. <https://doi.org/10.1044/jshr.2403.351>
- 1413 Tani, T., Koike-Tani, M., Tran, M. T., Shribak, M., & Levic, S. (2021). Postnatal structural  
1414 development of mammalian Basilar Membrane provides anatomical basis for the  
1415 maturation of tonotopic maps and frequency tuning. *Scientific Reports*, 11(1), 7581.  
1416 <https://doi.org/10.1038/s41598-021-87150-w>
- 1417 Tharpe, A. M., & Ashmead, D. H. (2001). A Longitudinal Investigation of Infant Auditory  
1418 Sensitivity. *American Journal of Audiology*, 10(2), 104–112.  
1419 [https://doi.org/10.1044/1059-0889\(2001/011\)](https://doi.org/10.1044/1059-0889(2001/011))

- 1420 Thomas, M. S. C., & Johnson, M. H. (2006). The computational modeling of sensitive  
1421 periods. *Developmental Psychobiology*, *48*(4), 337–344.  
1422 <https://doi.org/10.1002/dev.20134>
- 1423 Thompson, J. A. F. (2020). *Characterizing and comparing acoustic representations in*  
1424 *convolutional neural networks and the human auditory system* [PhD Thesis,  
1425 Université de Montréal].  
1426 [https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/24665/Thompson\\_Jessi](https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/24665/Thompson_Jessi)  
1427 [ca\\_2020\\_these.pdf?sequence=2](https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/24665/Thompson_Jessi_ca_2020_these.pdf?sequence=2)
- 1428 Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to  
1429 language: The procedural deficit hypothesis. *Cortex*, *41*(3), Article 3.  
1430 [https://doi.org/10.1016/S0010-9452\(08\)70276-4](https://doi.org/10.1016/S0010-9452(08)70276-4)
- 1431 Velez, M., & Schwartz, R. G. (2010). Spoken word recognition in Sschool-age children with  
1432 SLI: Semantic, phonological, and repetition priming. *Journal of Speech, Language,*  
1433 *and Hearing Research*, *53*(6), Article 6. [https://doi.org/10.1044/1092-4388\(2010/09-](https://doi.org/10.1044/1092-4388(2010/09-0042))  
1434 [0042\)](https://doi.org/10.1044/1092-4388(2010/09-0042))
- 1435 Wang, S., Lin, M., Sun, L., Chen, X., Fu, X., Yan, L., Li, C., & Zhang, X. (2021). Neural  
1436 Mechanisms of Hearing Recovery for Cochlear-Implanted Patients: An  
1437 Electroencephalogram Follow-Up Study. *Frontiers in Neuroscience*, *14*, 624484.  
1438 <https://doi.org/10.3389/fnins.2020.624484>
- 1439 Warden, P. (2018). *Speech commands: A dataset for limited-vocabulary speech recognition.*  
1440 <http://arxiv.org/abs/1804.03209>
- 1441 West, G., Vadillo, M. A., Shanks, D. R., & Hulme, C. (2017). The procedural learning deficit  
1442 hypothesis of language learning disorders: We see some problems. *Developmental*  
1443 *Science*, May 2016, Article May 2016. <https://doi.org/10.1111/desc.12552>

- 1444 Westermann, G., & Ruh, N. (2012). A neuroconstructivist model of past tense development  
1445 and processing. *Psychological Review*, *119*(3), 649–667.  
1446 <https://doi.org/10.1037/a0028258>
- 1447 Westermann, G., Sirois, S., Shultz, T. R., & Mareschal, D. (2006). Modeling developmental  
1448 cognitive neuroscience. *Trends in Cognitive Sciences*, *10*(5), 227–232.  
1449 <https://doi.org/10.1016/j.tics.2006.03.009>
- 1450 Yamins, D. L. K., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to  
1451 understand sensory cortex. *Nature Neuroscience*, *19*(3), Article 3.  
1452 <https://doi.org/10.1038/nn.4244>

1453

**Appendix**

1454

**ResNet-18 convolutional layer specification and hyperparameters**

Layer index	Layer name	Output size	Kernel size	Stride	Padding
1	Conv. 2D	1, 64	7, 7	2, 2	3, 3
2	Conv. 2D	64, 64	3, 3	1, 1	1, 1
3	Conv. 2D	64, 64	3, 3	1, 1	1, 1
4	Conv. 2D	64, 64	3, 3	1, 1	1, 1
5	Conv. 2D	64, 64	3, 3	1, 1	1, 1
6	Conv. 2D	64, 128	3, 3	2, 2	1, 1
7	Conv. 2D	128, 128	3, 3	1, 1	1, 1
8	Conv. 2D	64, 128	1, 1	2, 2	n/a
9	Conv. 2D	128, 128	3, 3	1, 1	1, 1
10	Conv. 2D	128, 128	3, 3	1, 1	1, 1
11	Conv. 2D	128, 256	3, 3	2, 2	1, 1
12	Conv. 2D	256, 256	3, 3	1, 1	1, 1
13	Conv. 2D	128, 256	1, 1	2, 2	n/a
14	Conv. 2D	256, 256	3, 3	1, 1	1, 1
15	Conv. 2D	256, 256	3, 3	1, 1	1, 1
16	Conv. 2D	256, 512	3, 3	2, 2	1, 1
17	Conv. 2D	512, 512	3, 3	1, 1	1, 1
18	Conv. 2D	256, 512	1, 1	2, 2	n/a
19	Conv. 2D	512, 512	3, 3	1, 1	1, 1
20	Conv. 2D	512, 512	3, 3	1, 1	1, 1

1455

*Note.* See Jupyter notebook for full network specification.

Hyperparameters

---

---

Optimizer: stochastic gradient descent

Learning rate: .001

Momentum: .9

Loss function: cross-entropy loss

---

1456