Atypical speech production of multisyllabic words by children with developmental dyslexia

Mahmoud Keshavarzi[1*], Giovanni M. Di Liberto[2,3], Fiona Gabrielczyk[1], Angela Wilson[1], Annabel Macfarlane[1], Usha Goswami[1]

[1]Centre for Neuroscience in Education, Department of Psychology, University of Cambridge, Cambridge, CB2 3EB, United Kingdom

[2]School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland.

[3]Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland.

* Email: mk919@cam.ac.uk.

Abstract

Cross-language data show that children with dyslexia are poor at recognizing syllable stress patterns, yet their speech production appears normal, suggesting an unexpected disconnect between speech input and output processes. Here we utilized a novel computerized speech copying task based on the speech amplitude envelope (AE) to investigate this disconnect. Seventy-five children with and without dyslexia copied familiar spoken targets like “Aladdin”. Children with dyslexia were significantly worse at producing the multi-syllabic targets as indexed by two similarity metrics for computing the AE. They did not differ from other children in producing the pitch contours of items. Accordingly, impaired AE perception in dyslexia is matched by impaired AE production. Children with dyslexia are heard to produce multi-syllabic words normally because their pitch contours are intact.

Developmental dyslexia is a disorder of learning that primarily affects accurate and fluent word reading and spelling[1]. Although there are multiple theories regarding etiology, the only theory with supporting data from infancy onwards is the phonological theory, whereby the core difficulty involves awareness of word sound structure[2]. In oral tasks, children with dyslexia have difficulty in perceiving whether words rhyme, difficulty identifying the mis-stressing of multi-syllabic words, difficulty counting syllables in words, and difficulty counting or manipulating phonemes (the smaller sound elements typically represented by the alphabet)[3]. These difficulties are found across spoken languages, whether oral syllable structure is simple (consonant-vowel, CV, as in Spanish) or complex (CCVCC, as in English)[4-6]. In the cross-language speech perception literature, impaired perception of phonology typically relates to impaired phonological production. For example, native speakers of Japanese have difficulty in perceiving the difference between the English speech sounds associated with the letters L and R, and also have difficulty in producing this phonological difference[7]. Yet despite their difficulties in perceiving the phonological structure of words, difficulties in *producing* accurate phonological structures have been difficult to demonstrate in children with dyslexia. This has led some researchers to propose that the phonological difficulties in dyslexia are primarily related to 'input phonological representations' – the representations required for speech perception rather than speech production[8].

However, for very young children there is evidence for speech production difficulties in dyslexia. In one longitudinal sample[9], toddlers who were genetically at-risk (AR) for dyslexia via family risk showed greater pausing and reduced production of syllables per second at 2 years of age compared to not-at-risk (NAR) toddlers, suggestive of difficulties in preparing multi-syllabic utterances. In earlier assessments between 8 – 19 months, the same AR toddlers had shown atypical babbling and proto-speech, producing less complex syllable structures, an early indicator of phonological sophistication[10]. In a separate line of research, word finding difficulties in tasks requiring picture naming of familiar

items (‘confrontation naming’ tasks) have frequently been documented in children with developmental dyslexia[11-17], particularly for multisyllabic items, whether of high frequency (e.g., *television*) or low frequency (e.g., *binoculars*). Producing the names for multi-syllabic items is significantly impaired when children with dyslexia are compared to reading-level-matched (RL) control children in addition to chronological-age-matched (CA) control children. The RL-matched children are younger than the children with dyslexia and consequently have less oral language learning and a lower mental age. When children with dyslexia perform more poorly than younger children despite being matched on word recognition and having a higher mental age and greater oral language experience, this is important in helping establish potential causal factors[18]. Reading acquisition is known to fundamentally change speech processing for all learners[19].

Dyslexic difficulties in accessing the phonological structure of multi-syllabic words would be predicted on the basis of the Temporal Sampling (TS) theory of developmental dyslexia[20], an auditory theory based on impaired perception of features of the speech amplitude envelope (AE). TS theory proposes that the automatic alignment of endogenous brain rhythms with rhythm patterns in speech is atypical in children with developmental dyslexia from birth, in part because the sensory cues (AE rise times, rates of change in amplitude onset at different AE frequencies) that trigger automatic neural alignment are perceived poorly by children with dyslexia. AE rise times are important sensory cues to stress placement in multi-syllabic words, as larger rise times at lower frequencies in the AE denote stressed syllables[21]. Children with dyslexia show impaired discrimination of amplitude rise times across languages and impaired neural representation of low frequency AE information[22,23]. If the AE is perceived poorly by children with developmental dyslexia, then it is logical to propose that the AE might also be produced poorly during natural speech.

To explore this hypothesis, we created a novel syllable stress interface for children and tested children with dyslexia and CA and RL controls. Children heard the computer pronounce a pictured known word (e.g. 'Aladdin') and simultaneously saw the target AE as a line on the computer screen (see Fig. 1). They were asked to repeat the target word 3 times, simultaneously viewing their own AE, which appeared in real time overlaying the target AE. Children were told to try to match their AE to the target AE. The online interface automatically computed the averaged child AE and correlated it to the target AE, simultaneously computing how well the pitch contour matched the target pitch contour (pitch contour was not depicted on-screen). There were 20 trials. This 'syllable stress game' was tested with 75 children aged between 9 and 11 years, 36 with dyslexia (DY1, N=19, dyslexic children who were receiving an oral rhythmic intervention based on poetry and music intended to support phonological awareness of syllable stress patterns; DY2, N=17, a waiting list group of dyslexic children due to receive the intervention later), 19 CA matched children, and 20 RL matched children.

Results

We focused our analyses on two different speech features of multi-syllabic words, the AE and the pitch contour. Two different metrics, correlation (r) and mutual information (MI), were utilized to compare AE and pitch contour similarity between the target stimuli and the child's responses. Prior to conducting the similarity analyses, we checked for repetition effects. It was expected that children would get closer to the target AE in each repetition. The results showed no evidence for repetition effects (all p's > 0.05).

The similarity of AE productions to the target pronunciation differed significantly between the groups for both similarity metrics. For the correlation value r, $\chi^2 = 14.96$, $p = 0.0018$, and post-hoc tests (Fisher's LSD) showed significant differences between the CA children and DY1 (the dyslexic children receiving the oral rhythmic intervention, $p = 0.01$), the CA children and DY2 (the untrained children with dyslexia, $p = 0.0003$), and between the CA and RL groups ($p = 0.005$). See Fig. 2A. For the MI

metric, χ^2 = 11.93, p = 0.0076, and post-hoc tests (Fisher's) showed significant differences between CA children and DY1 (p = 0.04), CA children and DY2 (p = 0.001), and significant differences between RL children and DY2 (the untrained dyslexic group, p = 0.02). See Fig. 2B. Removal of outliers did not change any of these significant effects, see Supplementary Information. To provide a visual instantiation, Fig. 3 shows both the AEs and the spectrograms for the stimulus "Aladdin", along with the corresponding responses from representative participants from each group (CA, RL, DY1, DY2).

The similarity of the pitch contours produced by the children to the target pronunciation showed no variation by group, for either similarity metric (r, χ^2 = 3.58, p = 0.31, see Fig. 4A; MI, χ^2 = 6.77, p = 0.0796, see Fig. 4B). Removal of outliers did not change any of these non-significant effects, see Supplementary Information. To provide a visual instantiation, Fig. 5 shows both the pitch contours and the spectrograms for the stimulus "Aladdin", along with the corresponding responses from representative participants from each group (CA, RL, DY1, DY2). The pitch estimates in Fig. 5 are shown only for the analysis windows with the harmonic component.

Discussion

Using a simple and novel copying task, we found robust difficulties in the accurate production of the AEs of familiar multi-syllabic words by children with developmental dyslexia. The visual instantiations (see Fig. 3) of their speech output suggest that children with dyslexia were not producing clear amplitude (intensity or loudness) differences between strong (stressed) and weak (unstressed) syllables. Children with dyslexia who were not yet receiving an oral rhythmic intervention designed to support their phonological awareness of multi-syllabic words (DY2 group) were particularly impaired, also performing significantly more poorly than younger RL-matched control children (*MI* measure). Although DY1 children were not significantly better than DY2 children in the copying task, the similar levels of performance of DY1 children to typically-developing RL controls at the group level suggests

that the oral intervention was helping them to perceive and produce speech rhythm patterns. Nevertheless, the DY1 group was still significantly less accurate than the CA controls in AE production, despite being matched to the typically-developing group for oral language development and IQ.

Of particular note, all groups performed at similar levels in terms of matching the pitch contours of the target items. This rules out any specific motoric or articulatory difficulties that could explain the impaired AE performance shown by the children with dyslexia. It may also explain why the speech production difficulties of children with dyslexia pass unnoticed by their listeners and teachers. The demonstration of impairments at the level of syllable stress and the AE is also important theoretically, supporting TS theory[20]. It is less consistent with auditory theories of dyslexia based on the phoneme[24]. The phoneme level of speech representation is typically thought to be learned via reading, and is seen by acousticians as a product of literacy acquisition[3,25]. Illiterate people lack phoneme awareness[26]. The data are also convergent with the pioneering longitudinal toddler work with AR samples, where differences in speech production at the syllable level were measurable long before a diagnosis of developmental dyslexia could be obtained[9,10].

The finding that toddler speech production differences were measurable at the syllable level rather than the phoneme level is linguistically important, because there are many null results in the dyslexia speech production literature. Most such studies have explored the phoneme level of speech representation, with negative results interpreted as showing no deficits. For example, when researchers used a range of tasks designed to measure ’phonological grammar’ (the typically unconscious and phonetic processes that are applied in speech production, such as voicing assimilation), French dyslexic adults performed as well as control adults[27]. In voicing assimilation in French, the voicing feature may spread backwards from obstruents or fricatives to the preceding consonant, so that for instance “cape grise” [ka**p**griz] becomes [ka**b**griz] (grey cloak). Yet French adults with dyslexia produced exactly the

same voicing assimilations as other adults. Similarly, French adults with dyslexia were able to produce non-native phonemes (Korean plosives) as accurately as non-dyslexic adults[28]. These data were interpreted as being inconsistent with the phonological theory of dyslexia[29]. The current data suggest instead that such studies have been relying on measures at the wrong linguistic level.

The current child data also have novel practical implications. First, it may be possible to identify risk for dyslexia early in development by adding multi-syllabic measures of speech production to current diagnostic tests. Although children show considerable individual variability in speech production measures, as indeed also demonstrated in our data, the addition of such measures could enable diagnosis and intervention long before children enter school and begin to experience literacy failure. Secondly, the data support the efficacy of dyslexia interventions that work at the level of syllable stress and speech rhythm[30-32]. It is likely that developmental effects will be optimized the younger that such interventions can occur[33]. Thirdly, acoustic interventions that address the sensory problems in perceiving AE structure in dyslexia are likely to be beneficial. For example, amplifying AE rise time cues in natural speech may help the dyslexic brain with speech processing and consequently facilitate development of a well-specified phonological system[34,35]. Finally, it is notable that language acquisition is thought to begin with speech rhythm, with stressed syllables providing clues to word onset in languages like English[36,37]. The demonstration that speech production in dyslexia is impaired at the level of syllable stress patterns suggests that the dyslexic experience of oral language is different from infancy onwards. This could suggest that early interventions should enhance the acoustic rhythm patterns that are naturally exaggerated in infant-directed speech[38]. Continuing to use infant-directed speech for longer in early life with AR infants may help to ameliorate the difficulties in perceiving and producing syllable stress patterns that characterize children with dyslexia.

Methods

Participant. Seventy-five children aged between 9 and 11 years participated: 19 chronological-age-matched control children (CA, average age 11 years 0 months), 20 reading-level-matched control children (RL, average age 9 years 4 months), 19 children with dyslexia who were receiving an oral rhythmic intervention comprising 18 sessions of musical and rhythmic language activities spread over the school term (DY1, average age 11 years 4 months), and 17 children with dyslexia who were awaiting intervention (DY2, average age 10 years 8 months). All participants were participating in an ongoing longitudinal study of auditory processing in dyslexia[39,40]. Due to participant drop-out related to the Covid pandemic, at the current test point the DY1 group were significantly older than the CA and DY2 groups, despite the groups being well-matched in age when the study began in 2017. Note that this age difference goes against the hypothesis, as DY1 should be at an advantage compared to CA.

All participants had previously received a short hearing test across the frequency range 0.25 – 8 kHz (0.25, 0.5, 1, 2, 4, 8 kHz) using an audiometer, and were sensitive to sounds within the 20 dB HL range. For DY1, the syllable stress production data were collected during sessions 7 and 9 of the 18-session oral rhythmic intervention. Participant characteristics by group are shown in Supplementary Table 1. Oral language as measured by the British Picture Vocabulary Scales[41] did not differ between groups. Intelligence (IQ) was pro-rated from 4 subtests of the Wechsler Intelligence Scale for Children (WISC[42]; similarities, vocabulary, block design and matrix reasoning) and did not differ between groups. However, performance in reading (British Ability Scales [BAS[43]] single word reading and Test of Word Reading Efficiency [TOWRE[44]] nonword reading), BAS spelling and phonology (Phonological Awareness Battery [PhAB] rhyme and phoneme awareness[45] was significantly lower for the DYS participants. All children and their parents gave informed consent for the study in accordance with the Declaration of Helsinki, and the study was approved by the Psychology Research Ethics Committee of the University of Cambridge.

Stimuli and materials. The auditory stimuli were 20 multi-syllabic words or phrases listed in Supplementary Table 2. The computerized speech stimuli were created from a native female speaker of standard Southern British English, and sampled at a 44.1-kHz rate with 24-bit resolution. Stimuli were presented at the same sampling rate of 44.1 kHz, and were chosen to ensure that they included a mixture of stressed and unstressed syllables in different positions within the words or phrases.

Experimental set-up. Participants were seated beside the experimenter in front of a laptop in a quiet room. They were presented with the 20 auditory stimuli through a MATLAB-based app. During presentation of each stimulus, the AE of the stimulus and a related picture were presented on the app's layout. The children were instructed to listen to the stimulus and to repeat three times what they heard, and their responses were recorded by a microphone. This resulted in 60 responses per child. Each response was recorded at a time interval of 3 seconds. After the stimulus was presented, the target picture (Fig. 1A) and the stimulus AE (Fig. 1B) were shown on the main layout. As the child repeated a target stimulus, the AEs of both the target stimulus and the child's response overlapped on the main layout (Fig. 1C).

Similarity measures. Two methods for computing similarity were used, the Pearson correlation (*r*) and Mutual Information (*MI*). The correlation between variables $X = \{x_1, x_2, \dots, x_T\}$ and $Y = \{y_1, y_2, \dots, y_T\}$ is calculated as:

$$r = \frac{\sum_t (x_t - \bar{X})(y_t - \bar{Y})}{\sqrt{\sum_t (x_t - \bar{X})^2}\sqrt{\sum_t (y_t - \bar{Y})^2}} \qquad (1)$$

where $\bar{X}$ and $\bar{Y}$ are average for variable X and Y, respectively. *MI* between two random variables is defined as a measure of information that one random variable gives about the other one, and is calculated as[46]:

$$MI(X,Y) = \sum_{x \in X} \sum_{y \in Y} P(x,y) log \frac{P(x,y)}{P(x)P(y)} \quad (2)$$

where $P(x)$ and $P(y)$ are the marginal distributions of variables X and Y, respectively, and $P(x,y)$ is the joint distribution of these variables. Here marginal and joint distributions were estimated using the Gaussian kernel estimator:

$$P(x) = \frac{1}{\sqrt{2\pi b^2 T}} \sum_{t=1}^{T} e^{\frac{-1}{2b^2}(x-x_t)^2} \quad (3)$$

$$P(y) = \frac{1}{\sqrt{2\pi b^2 T}} \sum_{t=1}^{T} e^{\frac{-1}{2b^2}(y-y_t)^2} \quad (4)$$

$$P(x,y) = \frac{1}{\sqrt{2\pi b^2 T}} \sum_{t=1}^{T} e^{\frac{-1}{2b^2}[(x-x_t)^2+(y-y_t)^2]} \quad (5)$$

where *b* refers to the bandwidth parameter that tunes the kernel function[47].

Statistical analysis

Repetition effects. To check for repetition effects, the mean scores across all 20 trials were computed separately for each of the three responses and for each participant. This left three scores (corresponding to the 1st, 2nd, and 3rd responses) for each participant. We then averaged the scores and compared them across the participants in each group (CA, RL, DY1, DY2) using parametric One-way ANOVA tests. This was done separately for each feature (AE and pitch contour) and each metric (*r* and *MI*).

Statistical comparison. An identical procedure was followed for comparing each speech feature (AE, pitch contour) and each similarity metric (*r* and *MI*) across groups. The mean AE or pitch contour per item was computed per child by averaging the 3 scores per item and was then averaged by group. Kruskal-Wallis One-way ANOVAs by group, *df* = 71, were used to compare the *r* and *MI* values obtained (CA, RL, DY1, DY2).

Data and materials availability

The analyses were performed using MATLAB. Any data (including costume codes) with the exception of the children's original voice recordings can be shared upon request.

References

1 Lyon, G. R., Shaywitz, S. E. & Shaywitz, B. A. A definition of dyslexia. *Annals of dyslexia* **53**, 1-14 (2003).

2 Goswami, U. *Theories of Developmental Dyslexia*. *In M. Skeide (Ed.), The Cambridge Handbook of Dyslexia and Dyscalculia. Cambridge University Press*, 5-24 (2022).

3 Ziegler, J. C. & Goswami, U. Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychol Bull* **131**, 3-29 (2005).

4 Elbro, C., Borstrøm, I. & Petersen, D. K. Predicting dyslexia from kindergarten: The importance of distinctness of phonological representations of lexical items. *Reading research quarterly* **33**, 36-60 (1998).

5 Ziegler, J. C. *et al.* Orthographic depth and its impact on universal predictors of reading: a cross-language investigation. *Psychol Sci* **21**, 551-559 (2010).

6 Law, J. M., Wouters, J. & Ghesquiere, P. The influences and outcomes of phonological awareness: a study of MA, PA and auditory processing in pre-readers with a family risk of dyslexia. *Dev Sci* **20** (2017).

7 Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R. & Tohkura, Y. Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *J Acoust Soc Am* **101**, 2299-2310 (1997).

8 Szenkovits, G. & Ramus, F. Exploring dyslexics' phonological deficit I: lexical vs sub□lexical and input vs output processes. *Dyslexia* **11**, 253-268 (2005).

9 Smith, A. B., Smith, S. L., Locke, J. L. & Bennett, J. A longitudinal study of speech timing in young children later found to have reading disability. *J Speech Lang Hear*, 1300–1314 (2008).

10 Lambrecht Smith, S., Roberts, J. A., Locke, J. L. & Tozer, R. An exploratory study of the development of early syllable structure in reading-impaired children. *J Learn Disabil* **43**, 294-307 (2010).

11 Denckla, M. B. & Rudel, R. G. Naming of object-drawings by dyslexic and other learning disabled children. *Brain Lang* **3**, 1-15 (1976).

12 Katz, R. B. Phonological deficiencies in children with reading disability: evidence from an object-naming task. *Cognition* **22**, 225-257 (1986).

13 Wolf, M. & Goodglass, H. Dyslexia, dysnomia, and lexical retrieval: A longitudinal investigation. *Brain Lang* **28**, 154-168 (1986).

14 Snowling, M., Van Wagtendonk, B. & Stafford, C. Object naming deficits in developmental dyslexia. *J Res Read* **11**, 67-85 (1988).

15 Constable, A., Stackhouse, J. & Wells, B. Developmental word-finding difficulties and phonological processing: The case of the missing handcuffs. *Appl Psycholinguist* **18**, 507-536 (1997).

16 Swan, D. & Goswami, U. Picture naming deficits in developmental dyslexia: The phonological representations hypothesis. *Brain Lang* **56**, 334-353 (1997).

17 Goswami, U., Schneider, W. & Scheurich, B. Picture naming deficits in developmental dyslexia in German. *Dev Sci* **2**, 53-58 (1999).

18 Bryant, P. & Goswami, U. Strengths and weaknesses of the reading level design: A comment on Backman, Mamen, and Ferguson. *Psychol Bull* **100**, 101–103 (1986).

19 Ziegler, J. C. & Ferrand, L. Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychol Bull Rev* **5**, 683-689 (1998).

20 Goswami, U. A temporal sampling framework for developmental dyslexia. *Trends Cogn Sci* **15**, 3-10 (2011).

21 Greenberg, S. *A Multi-tier Framework for Understanding Spoken Language*. *Listening to Speech* 411-433 (Psychology Press, 2012).

22 Goswami, U. *et al*. Language-universal sensory deficits in developmental dyslexia: English, Spanish, and Chinese. *J Cogn Neurosci* **23**, 325-337 (2011).

23 Power, A. J., Colling, L. J., Mead, N., Barnes, L. & Goswami, U. Neural encoding of the speech envelope by children with developmental dyslexia. *Brain Lang* **160**, 1-10 (2016).

24 Tallal, P. Improving language and literacy is a matter of time. *Nat Rev Neurosci* **5**, 721-728 (2004).

25 Port, R. How are words stored in memory? Beyond phones and phonemes. *New ideas in psychology* **25**, 143-170 (2007).

26 Morais, J., Cary, L., Alegria, J. & Bertelson, P. Does awareness of speech as a sequence of phones arise spontaneously? *Cognition* **7**, 323-331 (1979).

27 Szenkovits, G., Darma, Q., Darcy, I. & Ramus, F. Exploring dyslexics' phonological deficit II: Phonological grammar. *First language* **36**, 316-337 (2016).

28 Soroli, E., Szenkovits, G. & Ramus, F. Exploring dyslexics' phonological deficit III: foreign speech perception and production. *Dyslexia* **16**, 318-340 (2010).

29 Ramus, F. & Szenkovits, G. What phonological deficit? *Q J Exp Physiol* **61**, 129-141 (2008).

30 Bhide, A., Power, A. & Goswami, U. A rhythmic musical intervention for poor readers: A comparison of efficacy with a letter based intervention. *Mind, Brain, and Education* **7**, 113-123 (2013).

31 Flaugnacco, E. *et al.* Music training increases phonological awareness and reading skills in developmental dyslexia: A randomized control trial. *PLoS One* **10**, e0138715 (2015).

32 Cancer, A. *et al.* The Effectiveness of Interventions for Developmental Dyslexia: Rhythmic Reading Training Compared With Hemisphere-Specific Stimulation and Action Video Games. *Front Psychol* **11**, 1158 (2020).

33 Kalashnikova, M., Burnham, D. & Goswami, U. Rhythm discrimination and metronome tapping in 4-year-old children at risk for developmental dyslexia. *Cogn Dev* **60**, 101129 (2021).

34 Van Hirtum, T., Moncada-Torres, A., Ghesquiere, P. & Wouters, J. Speech envelope enhancement instantaneously effaces atypical speech perception in dyslexia. *Ear Hear* **40**, 1242-1252 (2019).

35 Van Hirtum, T., Ghesquiere, P. & Wouters, J. A bridge over troubled listening: improving speech-in-noise perception by children with dyslexia. *J Assoc Res Otolaryngol* **22**, 465-480 (2021).

36 Cutler, A. & Carter, D. M. The predominance of strong initial syllables in the English vocabulary. *Comput Speech Lang* **2**, 133-142 (1987).

37 Mehler, J. *et al.* A precursor of language acquisition in young infants. *Cognition* **29**, 143-178 (1988).

38 Leong, V., Kalashnikova, M., Burnham, D. & Goswami, U. The temporal modulation structure of infant-directed speech. *Open Mind* **1**, 78-90 (2017).

39 Keshavarzi, M. *et al.* Atypical delta-band phase consistency and atypical preferred phase in children with dyslexia during neural entrainment to rhythmic audio-visual speech. *NeuroImage: Clinical*, 103054 (2022).

40 Mandke, K. *et al.* Neural sampling of the speech signal at different timescales by children with dyslexia. *NeuroImage* **253**, 119077 (2022).

41 Dunn, L. M. & Dunn, D. M. *The British Picture Vocabulary Scale*. (GL Assessment Limited, 2009).

42 Wechsler, D. *Wechsler Intelligence Scale for Children*, 5th Edn. *London, UK: Pearson Assessment* (2016).

43 Elliot, C. D., Smith, P. & McCulloch, K. *British Ability Scales II. Windsor: NFER-Nelson* (1996).

44 Torgesen, J. K., Wagner, R. K. & Rashotte, C. A. *Test of Word Reading Efficiency (TOWRE)*. (Austin, TX: Pro-Ed, 1999).

45 Frederickson, N., Frith, U. & Reason, R. *Phonological Assessment Battery (manual and test materials)*. (nfer-Nelson, 1997).

46 Cover, T. M. *Elements of Information Theory*. (John Wiley & Sons, 1999).

47 Qiu, P., Gentles, A. J. & Plevritis, S. K. Fast calculation of pairwise mutual information for gene regulatory network reconstruction. *Comput Methods Programs Biomed* **94**, 177-180 (2009).

Acknowledgments

The authors would like to thank all the children, families and schools involved in the study. This research was funded by a grant awarded to UG by the Foundation Botnar (project number: 6064) and a donation to UG from the Yidan Prize Foundation. The sponsors played no role in the study design, data

interpretation nor writing of the report. GDL was supported by Science Foundation Ireland under the Career Development Award 15/CDA/3316 (EL) and under Grant Agreement No. 13/RC/2106_P2 at the ADAPT SFI Research Centre at Trinity College Dublin. ADAPT, the SFI Research Centre for AI-Driven Digital Content Technology, is funded by Science Foundation Ireland through the SFI Research Centres Programme.

Author contributions

M.K. contributed to conceptualization, methodology, data analyses, visualization, and writing – original draft; G.D.L. contributed to conceptualization, methodology, and writing – review and editing; F.G. contributed to investigation and writing – review and editing; A.W. contributed to investigation and writing – review and editing; A.M. contributed to investigation; U.G contributed to funding Acquisition, project administration, supervision, conceptualization, methodology, and writing – original draft.

Competing interests

The authors declare no competing interests.

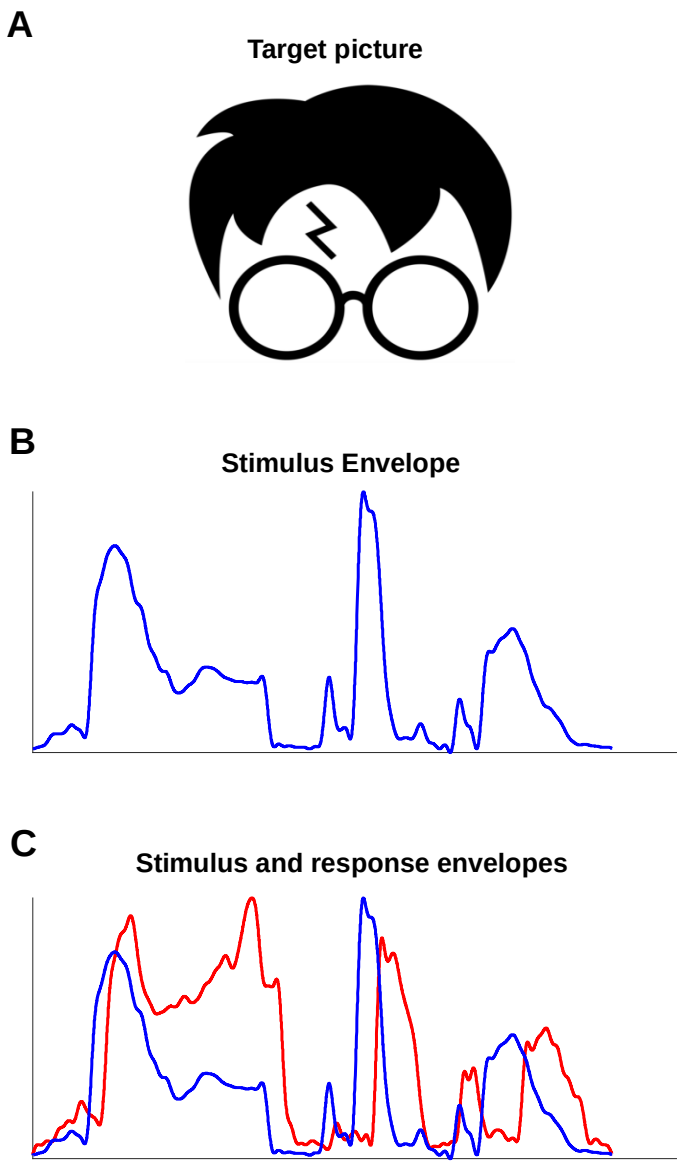

Fig 1. Schematic depiction of sequence of events during a single trial for the speech stimulus "Harry Potter". The child first saw a picture related to the target stimulus (panel A), then a depiction of the stimulus envelope as the word was pronounced (panel B), and then both the stimulus envelope (blue) and the participant's response envelope in real time (panel C). The child was instructed to repeat the target stimulus in a time duration of 3 seconds. As the child repeated the stimulus, the envelope of the target stimulus with the child's response envelope overlaying were shown on the main layout unfolding in real time (panel C).

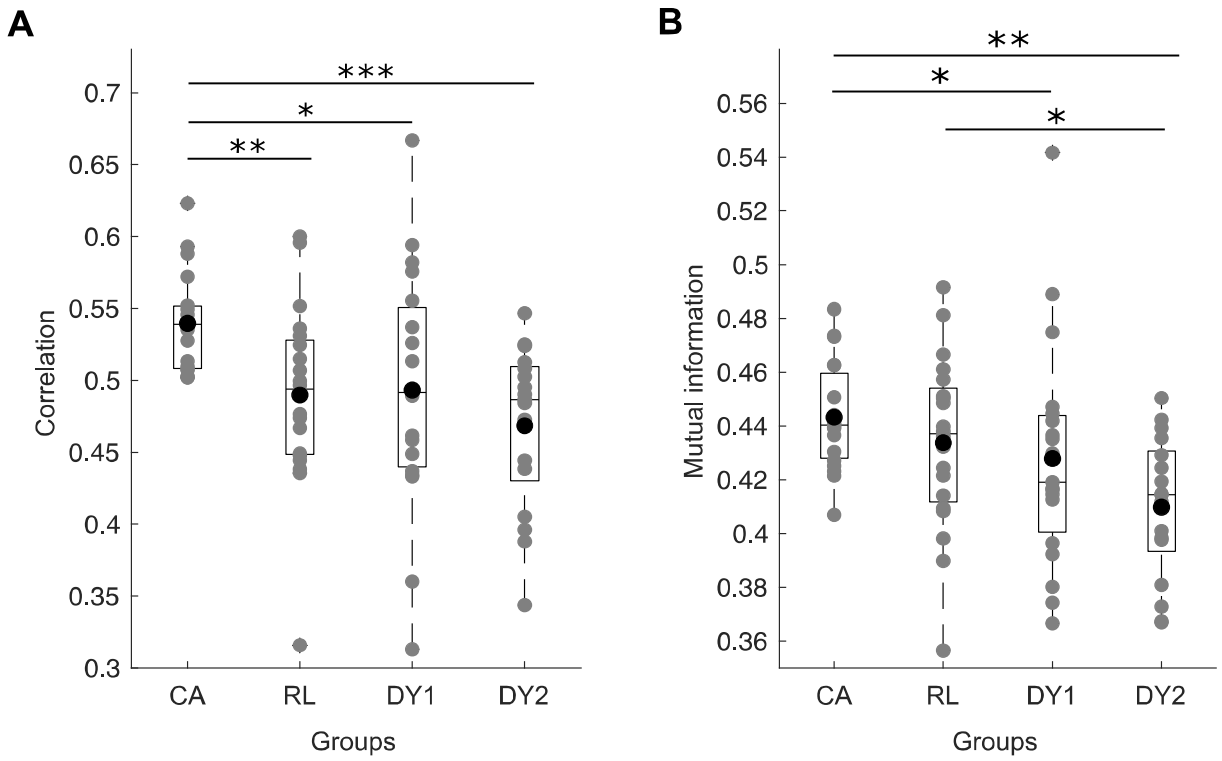

Fig 2. Performance for the four groups (CA, RL, DY1: Dyslexia receiving intervention, DY2: Dyslexia waiting for intervention) in amplitude envelope matching based on (A) the correlation between stimuli envelopes and the envelopes of corresponding responses, and (B) the mutual information between stimuli envelopes and the envelopes of corresponding responses. The grey disks denote individual participant scores and the black disks denote the mean values.

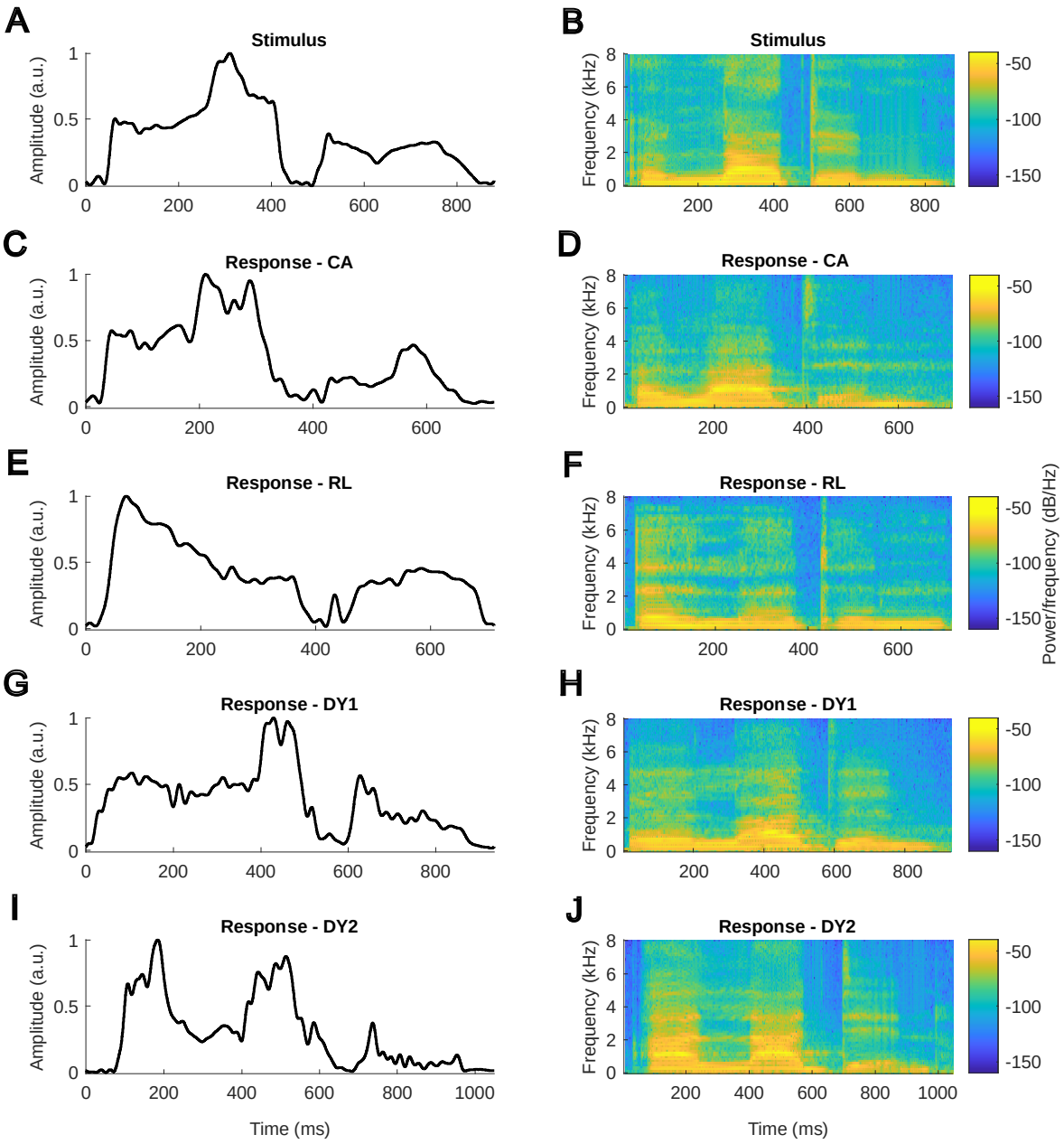

Fig 3. Amplitude envelopes (left) and spectrograms (right) of target stimulus "Aladdin" (panels A and B), corresponding response from a CA child (panels C and D), corresponding response from an RL child (panels E and F), corresponding response from a DY1 child (dyslexic receiving intervention, panels G and H), corresponding response from a DY2 child (dyslexic waiting for intervention, panels I and J). Note that for the correlation and mutual information analyses, the child's speech productions were resampled to have the same length as the corresponding target stimuli.

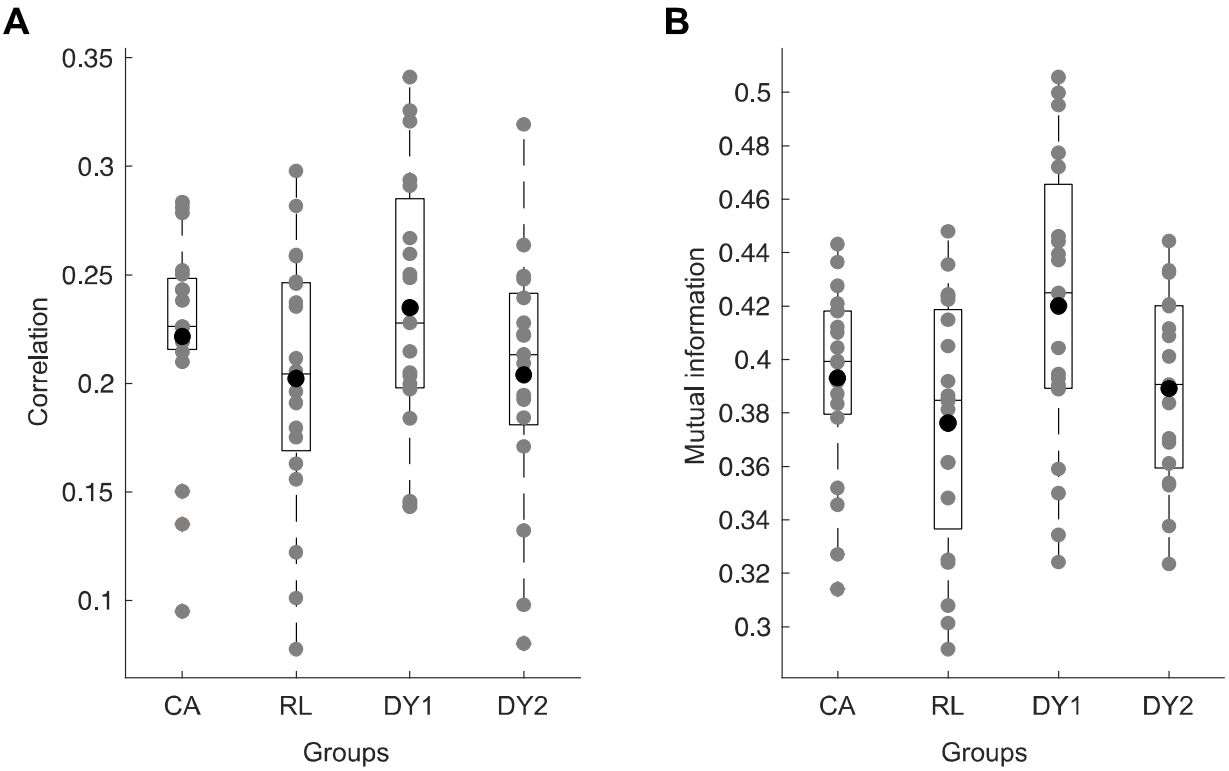

Fig 4. Performance for the four groups (CA, RL, DY1, DY2) in pitch contour matching based on (A) the correlation between pitch contours of the targets and the corresponding child responses, and (B) the mutual information between target pitch contours and the pitch contours of the corresponding child responses. The grey disks denote individual participant scores and the black disks denote the mean values.

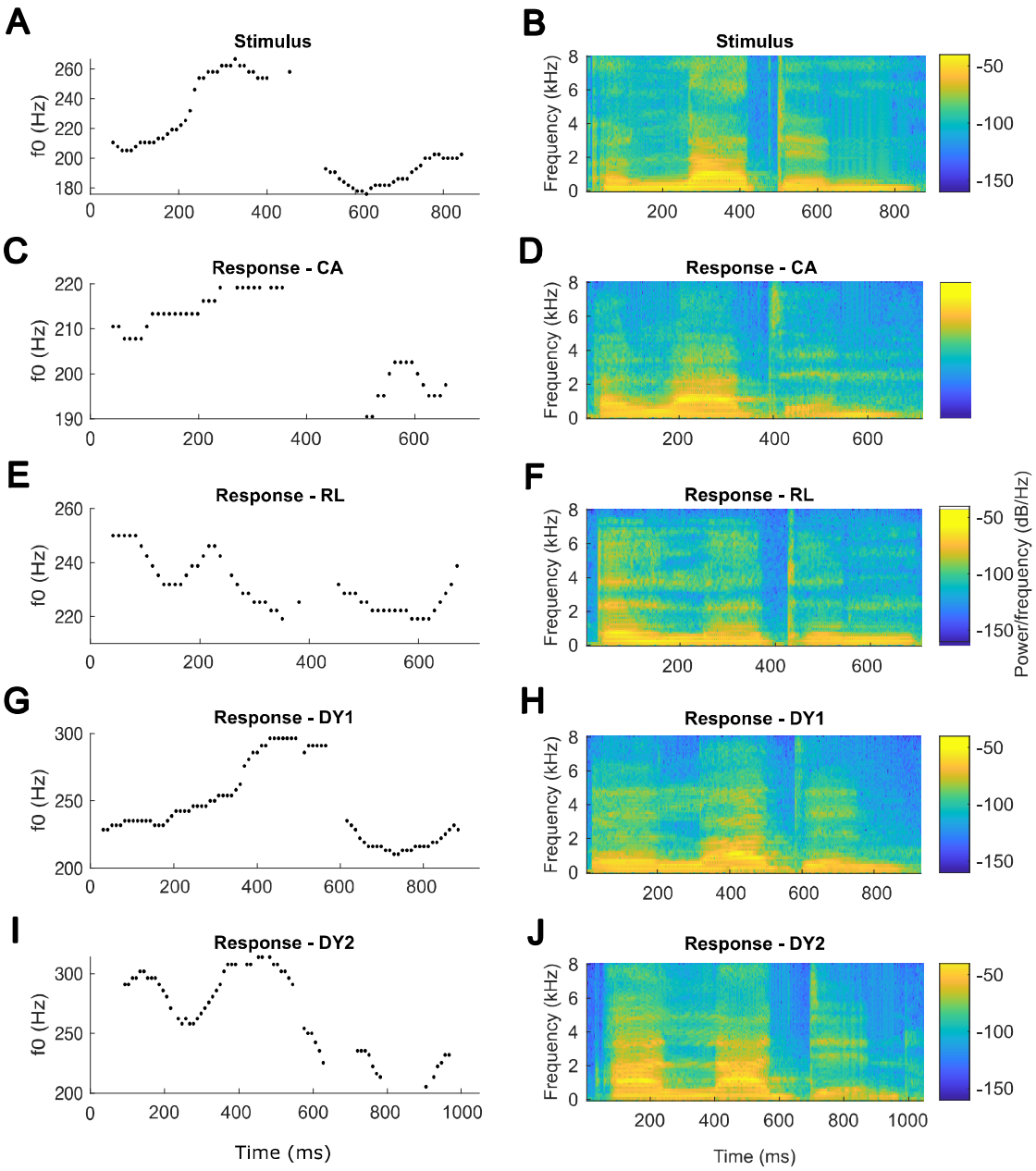

Fig 5. Pitch contours (left) and spectrograms (right) of target stimulus "Aladdin" (panels A and B), corresponding response from a CA child (panels C and D), corresponding response from an RL child (panels E and F), corresponding response from a DY1 child (dyslexic receiving intervention, panels G and H), corresponding response from a DY2 child (dyslexic waiting for intervention, panels I and J). The pitch estimates in the left column of the figure have been shown only for the analysis windows with the harmonic component. Note that for the correlation and mutual

information analyses, the child's speech productions were resampled to have the same length as the corresponding target stimuli.

Supplementary Information for

Atypical speech production of multisyllabic words by children with developmental dyslexia

Mahmoud Keshavarzi*, Giovanni M. Di Liberto, Fiona Gabrielczyk, Angela Wilson, Annabel Macfarlane, Usha Goswami

* Email: mk919@cam.ac.uk.

Results for Outliers Removal

We excluded individual data from the analyses as an outlier if the corresponding value was more than a 1.5 interquartile range above the upper quartile or below the lower quartile of the population data. We then repeated the statistical analyses described in the manuscript (see Methods – Statistical analysis).

AE – *r*. A statistically significant variation was found between the groups (Kruskal-Wallis One-way ANOVA, $df = 69$, $\chi^2 = 13.44$, $p = 0.0038$). Multiple comparison post-hoc tests (Fisher's least significant difference) showed significant differences between CA and RL groups ($p = 0.0176$), significant differences between CA children and DY1 ($p = 0.0137$), and significant differences between CA children and DY2 ($p = 0.0004$).

AE – *MI*. We found statistically significant variation between the groups (Kruskal-Wallis One-way ANOVA, $df = 70$, $\chi^2 = 12.94$, $p = 0.0048$). Multiple comparison post-hoc tests (Fisher's least significant difference) showed significant differences between CA children and DY1 ($p = 0.0201$), significant differences between CA children and DY2 ($p = 0.0008$), and significant differences between RL children and DY2 ($p = 0.0182$).

Pitch Contour – *r*. After the outlier removal, no statistically significant variation was found between the groups (Kruskal-Wallis One-way ANOVA, $df = 67$, $\chi^2 = 5.75$, $p = 0.1249$).

Pitch Contour – *MI*. After the outlier removal, we found no significant variation between the groups (Kruskal-Wallis One-way ANOVA, $df = 70$, $\chi^2 = 6.9$, $p = 0.0751$).

Supplementary Table 1. Details of the participating children.

	Dyslexic 1 (intervention group, DYS1)	Dyslexic 2 (wait list group, DYS2)	Age-Matched Control (CA)	Reading-Matched Control (RL)
N	19	17	19	20
Age (months) [A]	136.6 (6.9)	128.3 (3.2)	132.2 (6.8)	111.8 (11.2)
BPVS SS [B]	97.5 (13.2)	100.3 (11.4)	102.8 (13.6)	97.7 (9.1)
WISC FSIQ pro-rated [B]	97.3 (13.0)	100.1 (11.6)	100.8 (11.1)	95.8 (15.8)
BAS Reading SS [C]	80.0 (8.4)	82.1 (7.9)	95.8 (10.5)	97.1 (8.7)
BAS Reading Age in months [D]	103.6 (15.6)	100.2 (14.3)	127.9 (15.7)	109.9 (17.8)
BAS Spelling SS [C]	78.7 (9.3)	81.1 (6.3)	96.6 (9.4)	99.3 (9.3)
TOWRE PDE SS [C]	84.6 (14.7)	82.0 (9.2)	103.3 (8.9)	99.9 (10.8)
PhAB Rhyme SS [E]	92.8 (19.6)	89.7 (13.3)	102.6 (12.2)	101.5 (11.2)
PhAB Phoneme SS [C]	101.3 (14.1)	100.7 (11.6)	110.9 (9.5)	111.0 (10.8)

Note. A – CA, DYS1, DYS2 sig older than RL; B – no group differences in oral language nor IQ; C – DYS1, DYS2 sig worse than CA, RL; D – CA sig higher than DYS1=DYS2=RL; E – CA sig better than DYS1, DYS2, RL = DYS1, DYS2 (p= 0.053). BPVS = British Picture Vocabulary Scales; BAS = British Ability Scales, TOWRE PDE = Test of Word Reading Efficiency Nonword scale; PhAB = Phonological Awareness Battery.

Supplementary Table 2. Words and phrases used in the study.

Aladdin	Bambi	Bob The Builder	Dr Who	Flushed Away
Harry Potter	Horrid Henry	Madagascar	Over The Hedge	Scooby Doo
Shaun The Sheep	The Cat In The Hat	The Gingerbread Man	The Jungle Book	The Lion King
The Muppets	The Simpsons	The Smurfs	Wallace And Gromit	Winnie The Pooh