Speech Perception in Children with Reading Disabilities: Phonetic Processing is the Problem

Yashuo Wu*[1], Jie Lu*[1], Cynthia J. Johnson[2], Joseph C. Toscano[3], Jont B. Allen[1]

* First authors

[1] Dept. of Electrical and Computer Engineering and Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign

[2] Dept. of Speech and Hearing Science, University of Illinois at Urbana-Champaign

[3] Dept. of Psychological and Brain Sciences, Villanova University

September 8, 2019 @ 08:14; *Version 3.24.3*

Abstract

Purpose: Reading disability (RD) is a key obstacle in the development of literacy, and studies show that 15-20% of grade-school students have some degree of RD. The current study examines one potential source of RD in young children (8-12 years old): inadequate open set, non-categorical phone processing abilities, which can result from middle ear pathologies.

Method: We present data from two tasks: **(1)** A 3-interval forced choice procedure called the *Syllable Confusion Oddball (SCO)* task, which examines the listener's ability to identify different syllables (CV/VC) from a string of three such syllables, spoken by three different talkers, selected from a database of 18 adult mixed-gender talkers. **(2)** A single CV/VC presentation task with a verbal response, denoted the *Nonsense Syllable Confusion Matrix* (NSCM) task, where the listener labels individual utterances of CV and VC sounds. Ten children having well-documented RDs, normal hearing, and normal language function completed the tasks, and their performance was compared to that of six reading control (RC) children with no RD. The consonants and vowels each had 4.2 bits of entropy (19 vowels and 19 consonants), providing a sufficient range of responses to investigate perceptual confusions and differences in error rates.

Results: For the SCO task, the proportion of errors was statistically significantly higher for the RD listeners compared with the RC listeners; the RD listeners had, on average, three to five times as many errors as the RC listeners. These errors were also highly idiosyncratic, with differences between individual subjects in the errors they made as a function of phone type (consonant vs. vowel) and syllable position (initial vs. final). For the NSCM task, there were large individual differences in the confusion pattern errors across subjects, indicating idiosyncratic responses. These errors were consistent (had low variance), indicating the subjects were highly proficient in doing the task.

Conclusions: Given the fundamental role of phone encoding and decoding in human communication, it naturally follows that phone encoding must play a role in learning and its disorders. **(1)** RD children have a significant speech perception problem in identifying open set syllables, despite normal pure-tone hearing and language processing abilities. **(2)** Individual RD subjects were highly different in their confusions (idiosyncratic), yet were consistent in the task (low variance, high error) **(3)** For the NSCM task, the errors were somewhat reduced, indicating that the single-interval task was easier (lower error) than the SCO three-interval task. **(3)** These results are at odds with several previous studies which shows weak indication of phoneme identification impairment in RD children. These differences are likely due to (1) the differences in methods of the present study (SCO and NSCM), (2) with large numbers of trials per child (e.g., 1,500) and (3) the large numbers of talkers (e.g., 18).

Keywords: reading disability, speech perception, phone encoding

Introduction

In typically-developing (TD) children, speech perception happens naturally, even effortlessly. The ability to discriminate and identify speech sounds provides the foundation for learning to produce and comprehend spoken language. In contrast, learning to read requires considerable instruction and practice. A strong parallel exists between reading disability (RD)[1] and the hearing impaired (HI), the difference being there is no cochlear loss in the RD population. A comparison of the reading development in these two groups is interesting, as are developmental comparisons of the deaf, with and without a cochlear implant.

The HI population has consistent idiosyncratic consonant confusions. In this preliminary study, we shall show that the same is true of the RD subjects. The reasons for this parallel are presently unknown, but are consistent with poor performance on phonetic awareness profiles and some speculations of phoneme (aka phone) processing deficiencies (Torgesen, 2004; Tallal, 2000).

One of the first steps in learning to read is *decoding*, which involves translating printed words to sounds. Indeed, accurate decoding can be seen as fundamental to the reading process, as suggested by Hanford (2018): "The starting point for reading is sound. A child who can't decode will never become a reader."

Decoding is also central to current models of reading, in that it provides the learner with the basic knowledge needed to map letters onto speech sounds and eventually directly to lexical representations (Seidenberg and McClelland, 1989). The decoding process can be contrasted with *encoding*, which involves mapping speech sounds to syllables, letters, or words. Clearly, basic encoding, allowing the listener to accurately recognize speech sounds, must be mastered before decoding can be taught. Thus, before teaching children to read, it is assumed they have normal speech perception (encoding) skills. We hypothesize it is exactly the breakdown of this assumption that is the source of reading dysfunction.

Issues with decoding and (possibly) encoding become relevant when we consider that a more than 10% of children have difficulty learning to read (Torgesen, 2004), and 10% or more are diagnosed as having a *reading disability* (RD). Understanding the nature of RDs is critically important, as a lack of literacy skills is associated with a number of negative outcomes (Torgesen, 2004, p. 25). For instance, according to national statistics, based on the 1994 Washington Summit on Learning Disabilities (Ellis and Cramer, 1994; NICHD/NRP, 2000a,b), 50% of inmates cannot read.

Moreover, understanding of the precise source of any RD could impact the success rate of treatment. Given this, and assuming that the encoding process is central to learning to read, one might naturally ask: (1) To what extent is accurate phonetic encoding important for reading, and (2) Can disruptions in phonetic encoding during early childhood lead to RD? For example, middle ear infections are common in early childhood, and these frequently lead to a temporary (less than one year), undetected 50 dB hearing loss (Williams and Jacobs, 2009). This could lead to a substantial disruption in exposure to speech sounds during a critical period, thus impacting speech development, contributing to RD.

In the current study, we ask whether reading development depends on the seemingly-easy preschool task of understanding speech (i.e., encoding), and we will show that RD children do not have TD speech perception. As with any correlated variables, causality is always an issue. It is easy to make the case that encoding is necessary for decoding skills, but difficult to understand how decoding could either proceed, or negatively influence encoding skills.

Below we investigate these issues by examining the speech perception abilities of RD children and TD *reading control* (RC) children. Specifically, we examine how accurately children recognize and classify speech sounds in a quiet environment (i.e., with no added noise) using a large database of natural speech sounds. This provides a measure of the listener's sensitivity to fine phonetic detail in the speech signal, which is necessary for accurate encoding (Allen, 2005; Phatak and Allen, 2007; Toscano and Allen, 2014).

We will first consider previous work on speech perception in RD children, as well as current models of reading that provide the basis for the experiment presented here. Next we present results from two tasks designed to measure listeners' speech perception abilities. Finally, implications for models of reading and approaches for RD interventions are discussed.

[1] We use the more general term *reading disability* (RD) to refer to any disruption in reading ability.

Corresponding Author:
Jont Allen
Dept. of Electrical and Computer Engineering
Univ. of Illinois at Urbana-Champaign
306 N Wright St
Urbana, IL 61801
Email: jontalle@illinois.edu *Phone: (217) 369-7711*

Table 1
Table of abbreviations.

TD	typically-developing
RC	reading control
RD	reading disabled or reading disability
SCO	syllable confusion oddball task
NSCM	nonsense syllable confusion matrix task
CI/CF	consonant in initial/final syllable position
VI/VF	vowel in initial/final syllable position
WI	Word Identification subtest
WA	Word Attack subtest
WRMT-R	Woodcock Reading Mastery Tests-Revised
R-FLU	Reading Fluency score
DYS	Dyslexia or dyslexic subjects
CA	Chronological age matched subjects
RL	Reading level matched subjects
dB	decibel
SNR	signal to noise ratio
AI	Articulation index

Phone perception research

An up to date historical and conceptual overview of learning disability may be found in (Torgesen, 2004). This summary of more than 100 years of research, has been carefully summarized, and is an important starting point for those not familiar with this extensive literature.

The importance of *phonological awareness* was identified early, however attempts to identify the source of learning, especially reading disabilities, have been unsuccessful. Many obvious possibilities have long since been ruled out, such as brain damage, low IQ and other higher brain dysfunction (Torgesen, 2004, p. 11). The persons with RD are normal in every measurable way, except that they cannot learn to read (Torgesen, 2004, p. 19). Thus identifying the source of this RD dysfunction is critical to future progress, and providing therapy. Many promising approaches have proven to be unsuccessful, the most notable is *whole word* learning (Torgesen, 2004, p. 12-13). This view is summarized by Torgesen

> . . . approaches to identifying and training deficient processes in children with LD were pressed into service when our understanding of mental processing operations, and their relationships to learning and performing academic tasks, were at only a rudimentary stage of development.

Namely, money was spend, but success was low, as summarized by Torgesen (2004) on page 20:

> However, whether any of these [RD] interventions will prove uniquely useful to children with LD, as opposed to other types of poor learners, remains to be demonstrated.

Studies on reading have primarily focused on using the categorical perception paradigm and involve varying stimuli (often synthetic speech sounds) along specific acoustic dimensions. While this provides a high degree of control over the stimuli, it does not capture the large variability present in natural speech. Moreover, these studies have often focused on only a small subset of speech sounds, often examining stop consonants, which may be recognized more accurately than other types of speech sounds. Other published work is aimed at measuring speech perception in children with RD using larger sets of speech sounds in other types of tasks that may allow us to better measure perception of sub-phonemic (i.e., phone-level) differences in speech.

Hazan and Adlard (1998) measured speech sound discrimination in 13 children with reading delays, 12 reading-age (RA) matched controls, and 12 chronological-age (CA) matched controls. Children were presented with several tasks, including a same-different speech sound discrimination task with stimuli presented in a VCV syllable structure, where the consonant was one of /b,d,g,k,v,ʒ,ʃ, s,f,m,n, ɹ,w,j,l/, all presented in the context of the vowel /ɑ/. There was no overall observed difference between the groups, but there was an interaction between group and consonant, such that children in the RD group made more errors for stop consonants. A subset of RD children were also found to make more errors overall compared with RC children across different speech sounds.

Messaoud-Galusi et al. (2011) carried out consonant identification and discrimination experiments using several different tasks with 62 DYS and 51 control children, examining perception of the contrast between /b/ and /p/ in quiet and in 20-talker babble noise. As in the studies described above, they found that the DYS subjects showed shallower identification functions than control subjects, but only for speech presented in quiet. In addition, DYS subjects showed poorer within-category phoneme discrimination; evidence of differences in across-category discrimination was mixed and depended on the specific task used. Overall, the authors conclude that there were not consistent speech perception deficits associated with dyslexia. This experiment was a follow-up of a similar study done with adults (Hazan et al., 2009) which used the same syllables and procedures and also suggested only weak support for a speech perception deficit in dyslexia.

Hazan et al. (2013) further investigated identification of consonants from the set /p,b,t,d,f,v,s,z,m,n,sp,st/ and discrimination of sounds varying in in place of articulation (/b/ vs. /d/) and voicing (/b/ vs. /p/). The study included 34 DYS subjects (mean age: 147.3 months) and 25 control subjects (mean age: 146.8 months). Children in the DYS group made more errors in identification but only for a subset of speech sounds, and they made more errors overall in the discrimination task.

In contrast to some of these studies, Ziegler et al. (2009) found speech perception deficits in a group of 19 DYS subjects compared with 18 CA control subjects and 19 RL control subjects. Listeners heard VCV syllables with the vowel /ɑ/ and one of 16 possible French consonants (/p,t,k,b,d,g,f,s,ʃ,m,n,r,l,v,z,j/). The results showed a difference between DYS and control subjects for speech recognition in noise, but no difference for speech recognition in quiet.

Finally, White-Schwoch et al. (2015) tested 112 children using an electrophysiological measure of phone processing (using the syllable /dɑ/ as testing material) with a group of 4-year-old children (N=37; mean age: 54.41 months) and a group of 3-year-old children (N=20; mean age: 43.35 months). They found that poor processing was related to differences in phonological awareness scores, suggesting a possible relationship between phone processing and reading ability. In addition, for a subset of children (N=34) who returned a year later, they found that the earlier neurophysiological measure predicted performance in measures of reading and literacy, again suggesting a link between phone processing and reading ability.

Summary and Research Questions

To summarize, work investigating the relationship between speech perception and reading development, while variable, has not found consistent differences between RD and TD children. However, many of these studies have used paradigms such as categorical perception tasks that do not accurately capture sensitivity to low-level (sub-phonemic) differences in speech, and many have only investigated a small subset of speech sounds, such as stop consonants. Therefore, a more thorough investigation is in order, to clarify the roles of the various processing channels and evaluate whether the classic Fletcher model of speech recognition (Fig. 11) can provide insights into the nature of RDs.

As discussed above and as documented by Miller et al. (1951), it is difficult to analyze a speech perception problem using meaningful speech as testing material due to the context channel, which plays a critical role in the process. The importance of the context channel was obvious to Fletcher (1995) in his development of the 1921 AI model, which is why he broke the problem into a cascade of processing elements. In Fletcher's model, contextual processing is the last stage in the speech perception process. Therefore, relatively speaking, phone recognition (encoding) is the more basic (earlier) layer of speech perception. In turn, accurate phone recognition is important for decoding and other reading skills. This suggests that phone perception may determine the success of reading and comprehension.

Torgesen (2004, p. 32) provides an insight that seems relevant to the present research

> …the primary cause of this [RD] problem is deficiencies in the ability to process the phonological features of language.

In our opinion, this is exactly what we have shown with our RD population, as discussed below.

Although phonological awareness and decoding—both of which assume accurate speech perception—are viewed as causal factors in RD, there are still studies showing that speech perception deficits do not seem to be involved. However few studies (if any) specifically map out the early phonetic encoding abilities of children with reading difficulties. Hence, in the present study, the following questions were addressed:

1. Without access to visual (i.e., letter) and contextual (i.e., word) information, do children with RD show a phonetic encoding deficit?
2. If so, what type of task is most effective in diagnosing RD? What task is most informative about RD regarding perception of specific speech sounds?
3. Do RD and TD children have common or unique phonetic perception patterns (i.e., do they have similar or dissimilar patterns of speech sound confusions)? Alternatively are they idiosyncratic? If so, to what extent? Are they consistent, and if so, to what extent?

Two experimental tasks were used to address these questions: (1) the *Syllable Confusion Oddball* (SCO) task, which involves speech sound discrimination, and (2) the *Nonsense Syllable Confusion Matrix* (NSCM) task, which involves speech sound identification, reported orally. This report mainly details the results of the SCO task, but with supplementary data from the NSCM task briefly presented to illustrate key differences between RD and RC children. These two tasks were designed based on modern (computer) extensions of Harvey Fletcher's and George A. Miller's experimental designs, as discussed in (Allen, 2005), using Matlab©.

SCO Task. For the SCO task, on each trial, subjects hear three naturally-produced speech syllables (CV or VC), where two sounds are the same C (or V) and one is different. Only a C or V was modified on a given trial, not both. For example, listeners might hear /ka, ʒa, ka/. They were asked to identify the position of the oddball syllable. The oddball was always chosen randomly to occur in one of the three positions. The three sounds were always either CV or VC, and the two syllable structures were never mixed within a trial. Critically, the three syllables were always spoken by three different talkers, chosen randomly from a set of 18 mixed gender talkers. Thus, the three CV/VC tokens were *always* different, due to the talker differences. The children understood the task was to identify the oddball syllable based on the C or V difference and that they should ignore talker and gender differences.

Similar oddball tasks have been used in previous work on RD (e.g., Bradley and Bryant, 1978, 1983), but these studies used meaningful words. As discussed above, the use of real words as testing materials can lead to influence from the context channel (i.e., word meaning) in addition to the auditory channel (what we wish to measure). In the present study, the testing materials are all maximum entropy (*MaxEnt*) in the SCO task, defined as syllables consisting of all possible combinations in English, with equal probability. MaxEnt syllables represent the full range of phonological differences that may occur. Using materials with no meaningful linguistic content allows us to focus on the contribution of the auditory channel to estimate children's phone identification and discrimination abilities.

The SCO task has extensive coverage of consonants and vowels in English: 24 consonants spanning 15 vowel contexts. The sounds were produced by 18 talkers to better represent the natural variation encountered with speakers of English. Because the children only needed to respond based on the information in the specific tokens, there was no need for them to segment phones from the syllables, or to identify/label them. Hence, the task should be easier, and thus more accurate, than a phonological awareness task. Moreover, there is no influence of the visual channel, since printed materials are not used. Thus, in this experiment, the children had to rely exclusively on the auditory channel. We report results from this task in Figs. 1– 5.

NSCM Task. The NSCM task complements the SCO task in that it provides confusion data, which are unavailable from the SCO task (since it only identifies which sounds are confusable, not which other sounds they are confused with). In the NSCM task, listeners hear a single CV or VC and are instructed to orally repeat the syllable they heard. The downside of this test is that it requires two transcribers to code the verbal report. This is error prone since the children do not always clearly articulate their response, especially when they are unsure of the identity of the spoken token. Nonetheless, this task provides useful data, missing from the SCO task. In particular, he NSCM task adds value by providing confusion matrix information, which could be used during training sessions and as diagnostic feedback on any change in the child's status. Knowing the degree to which confusions occur could also prove useful to a speech therapist. We report preliminary results from this task in Figs. 6–10.

Methods

Participants

Two groups, RD and RC children, participated in the experiment. The RD group had 11 children (seven girls), aged 8 to 10 years. The RC group had six children (two girls), aged 8 to 11 years. This is a typical age range where deficits are discovered, but rarely overcome, during reading development. Interviews of at least one parent were conducted. All the children with RD were recruited from the Urbana *Reading Group*.[2] The study was approved by the University of Illinois at Urbana-Champaign IRB. Note that names used in the manuscript are pseudonyms.

Every child passed a pure-tone hearing screening (500Hz, 1kHz, 2kHz, and 4kHz in each ear, at 20 dB SPL), indicating normal hearing ability. This test was repeated at the beginning of each visit, to ensure that there was no temporary hearing loss. There were no known visual, neurological, cognitive, or emotional problems for these subjects. All the children had at least a normal IQ for their age.

To assess their reading abilities, a battery of reading tests were administered including the Woodcock Reading Mastery Tests-Revised (WRMT-R), specifically the Word Identification subtest (WI) and Word Attack subtest (WA), and the Grey Oral Reading Test, 4th edition (GORT-4), which included a Fluency score (R-FLU) and Comprehension score (R-COMP; Johnson et al., 2015). For each measure, the children in the RD group scored significantly lower than the children in the RC group (using Welch's t-test): WI ($t(9)=4.34$, $p=0.002$), WA ($t(11)=5.18$, $p<0.001$), R-FLU ($t(7)=4.03$, $p=0.005$), and R-COMP ($t(13)=4.64$, $p<0.001$). For all four reading measures, the RD group scored at least one standard deviation below the RC children, using the criterion of each measure.

Stimuli

Natural speech sounds have more subtle and realistic perceptual cues than synthetic sounds, thus are considered superior for human speech perception tests (Li, 2009). The set of natural sounds that were used for both tasks came from the commercial

[2] Reading Group: 3011 Village Office Pl, Champaign, IL 61822

Table 2
Definitions of related concepts.

Phone	speech sound token; smallest sound unit of speech (i.e., individual consonants and vowels)
Phoneme	set of phones of the same type; smallest unit that distinguishes meaning
Syllable	combination of any number of consonants (C) and a vowel (V), such as V, CV, and VC
Word	meaningful syllable sequence (spoken or written)
Maximum entropy syllable	syllable constructed by random selection of consonants and vowels; abbreviated *MaxEnt syllable*
Phonological awareness	the metalinguistic skill of attending to, judging, rhyming, blending, segmenting, and manipulating spoken words, syllables, phonemes, or phones (in the present study, use of the term is restricted to phones in non-words)
Phonetic encoding	the auditory ability of hearing, perceiving, and categorizing phones (speech sounds)
Articulation	recognition of syllables having no meaning; based on Fletcher's *Articulation Index* model
Intelligibility	recognition of syllable sequences having defined meaning (i.e., words)

Linguistic Data Consortium LDC-2005S22 database (Fousek et al., 2004). The database contains a set of all diphone syllables allowed in English (i.e., CV and VC syllables) formed from 24 consonants and 15 vowels, spoken by 18 talkers.[3] Necessarily, some subset of syllables are words, as required by the definition for MaxEnt syllables.

Stimuli were presented without background noise, in random sequence, at the listener's most comfortable loudness level, chosen by the subject at the beginning of each session. The sounds could be replayed as many times as desired, but very few sounds were requested for more than three presentations. Children listened through AKG K240 Monitor headphones (circumaural, 600 Ω) via the laptop's 24 bit sound card. Sounds were processed to remove artifacts (e.g., lip smacks) and loudness variations.

Table 3
Average number of trials and standard deviation for CI, CF, VI, and VF stimuli.

	Initial	Final
C	42(19)	44(17)
V	43(19)	45(15)

Procedure

The data were collected during 2005 and were formatted and cleaned up between 2006-2010. The clean-up process was labor intensive with a great deal of hand-processing and crossing checking was involved, which explains the lengthy processing time. The processing was mostly done by graduate students in the Speech and Hearing Sciences Department. Data collection was done by graduate students from both SHS and ECE.

Each subject (child) participated in the study for up to 10 weeks, for a maximum of two hours per week. They completed one hour of testing using the SCO task every other week. On average, each child completed 1600 trials of the SCO task, with a standard deviation of 548 trials. During each task, to avoid the possibility of fatigue and boredom, children were given five minute breaks for rest and treats, for every ten minutes of testing. Table 3 shows the average number of trials and standard deviation for consonants (row 1) and vowels (row 2) in initial position (column 1) and final position (column 2), averaged over the RD and RC children (e.g., for consonant initial stimuli, the average was 42 and the standard deviation was 19).

The SCO task consisted of a combination of three MaxEnt (i.e., random) syllables (either all CVs or all VCs). The children were asked to point at one of three wooden blocks labeled with the numbers 1, 2, or 3, corresponding to the oddball sound. The response was recorded by the research assistant for subsequent analysis. In the NSCM task, children heard a single syllable and repeated what they heard.

[3]Instead of International Phonetic Alphabet (IPA) symbols, some figures in this report were generated using Darpabet to represent the phones. Sounds used in this study are presented in Appendix A along with the conversion between Darpabet and IPA.

Results

Results for the SCO task are presented first, starting with the results of statistical analyses of the data, followed by a description of the results for individual listeners compared with average performance for the RD and RC groups. Data are organized based on each of the four experimental conditions (i.e., consonant initial [CI], consonant final [CF], vowel initial [VI], vowel final [VF]).

Table 4

Summary of fixed effects in logistic regression analysis. There are three fixed variables: Group: {RC, RD}; Phone Position: {Initial, Final}; Type (see Appendix A1): {Consonant, Vowel}. The parameter b is the regression coefficient (i.e., slope) with respect to each fixed effect, and SE is the standard error of the regression estimate. χ^2 values are from the likelihood ratio tests (see text), and p is the p-value for those tests.

Fixed Effects	b	SE	$\chi^2(1)$	p	Sig. Level
Group: RC, RD	1.123	0.304	9.918	0.002	$**$
Phone Position: Initial, Final	-0.279	0.096	7.706	0.006	$**$
Phone Type: C,V	0.057	0.113	0.320	0.572	
Phone Position × Group	-0.217	0.191	1.213	0.271	
Phone Position × Phone Type	-0.054	0.160	0.326	0.568	
Group × Phone Type	0.013	0.192	0.013	0.910	
Group × Phone Position × Phone Type	-0.344	0.312	1.178	0.278	
Significance codes	$p < 10^{-3}$	$***$;	$p < 10^{-2}$ $**$;	$p < 0.05$ $*$	

Table 5

Summary of random effects in logistic regression analysis. There are two random effects: Subject: {subjects within each group (RD, RC)}; Phone: {see A1}; along with their interactions with phone position, phone type, and group.

Random Effects	$\chi^2(1)$	p	Sig. Level	
Subject	983.918	$< 2 \times 10^{-16}$	$***$	
Phone × Phone Position	4.218	0.040	$*$	
Subject × Phone Position	33.472	7×10^{-9}	$***$	
Subject × Phone Type	33.545	7×10^{-9}	$***$	
Phone × Group	1.768	0.184		
Phone × Group × Phone Position	0.144	0.705		
Subject × Phone Position × Phone Type	15.275	9×10^{-5}	$***$	
Subject × Phone	0.403	0.526		
Significance codes	$p < 10^{-3}$	$***$;	$p < 10^{-2}$ $**$;	$p < 0.05$ $*$

Regression Analysis for the SCO task

Data were analyzed using logistic mixed-effects models fit using the lme4 package (Bates et al., 2015) using the R language (R Core Team, 2014) with proportion error as the dependent measure. Note there is a simple relation between log-odds ($\log(P_e/1 - P_e)$) and probability of error P_e, allowing one to easily convert between them. Baayen et al. (2008) discusses the mixed-effects models, while Jaeger (2008) discusses the log-odds method.

The model includes three *fixed effects* (Table 4), defined as discrete binary sets (having elements $\in \mathbb{N}$), which are: (1) Subject Group (RD vs. RC), allowing us to determine whether there is a difference in P_e between the two groups, versus the null hypothesis (i.e., no difference between the two groups); (2) Phone position (initial vs. final), and (3) Phone type (consonant vs. vowel). Fixed effects were numerically coded and centered with a mean of zero, with predictors for main effects of each factor and all two-way and three-way interactions between the factors.

The model also includes two *random effects* (Table 5) defined as continuous sets (having elements $\in \mathbb{R}$), such as factors sampled from the population: (1) individual listeners and (2) individual phones (i.e., the specific consonant or vowel that the listener heard). Main effects corresponding to these differences are entered as random intercepts in the model. We also included by-subject and by-phone random-effect slopes, which describe interactions with fixed effects in the model as shown in Table 4. These include by-subject slopes for phone position (i.e., do phone position effects vary depending on the individual listener?), phone type (do consonant vs. vowel effects vary depending on the individual listener?), and their interaction. By-phone random-effect slopes were included for subject group (do differences between RD and RC listeners depend on the specific phone?), phone position (do differences between initial and final position depend on the specific phone?), and their interactions.

Lastly, we included a term corresponding to the interaction between subject and phone, allowing us to see whether individual listeners differ in the specific phones for which they make errors. Note that these models do not test for a overall effect of phone, as we already know that errors vary considerably as a function of this factor (i.e., some phones produce more errors than others; Singh and Allen, 2012; Toscano and Allen, 2014).

Likelihood ratio (χ^2) tests were used to determine whether each of the factors above had a significant effect on P_e.[4] These tests were performed using nested model comparisons in which factors are added to the model sequentially and we measure whether including the factor significantly improves the model's fit to the data (based on the change in its log likelihood). This provides a way to test significance for both the fixed and random effects in the model. Two sets of likelihood ratio tests were performed, one for fixed effects (including models with all random effects terms) and one for random effects (including all fixed effects terms).

Results summary

A summary of results from the regression analysis for the fixed effects is presented in Table 4, where b is the regression coefficient and SE is the standard error. The largest effect is subject group (b=1.123, SE=0.304, $\chi^2(1)$=9.918, p=0.002), demonstrating that the RD listeners made significantly more errors than the RC listeners, with an average of 17.9% error for the RD group and 6.6% error for the RC group.[5] There was also a significant effect of phone position (b=-0.279, SE=0.096, $\chi^2(1)$=7.706, p=0.006), with listeners making more errors for sounds in final position (mean error: 14.4%) than initial position (mean error: 12.1%). No other fixed effects or interactions between the fixed effects were significant.

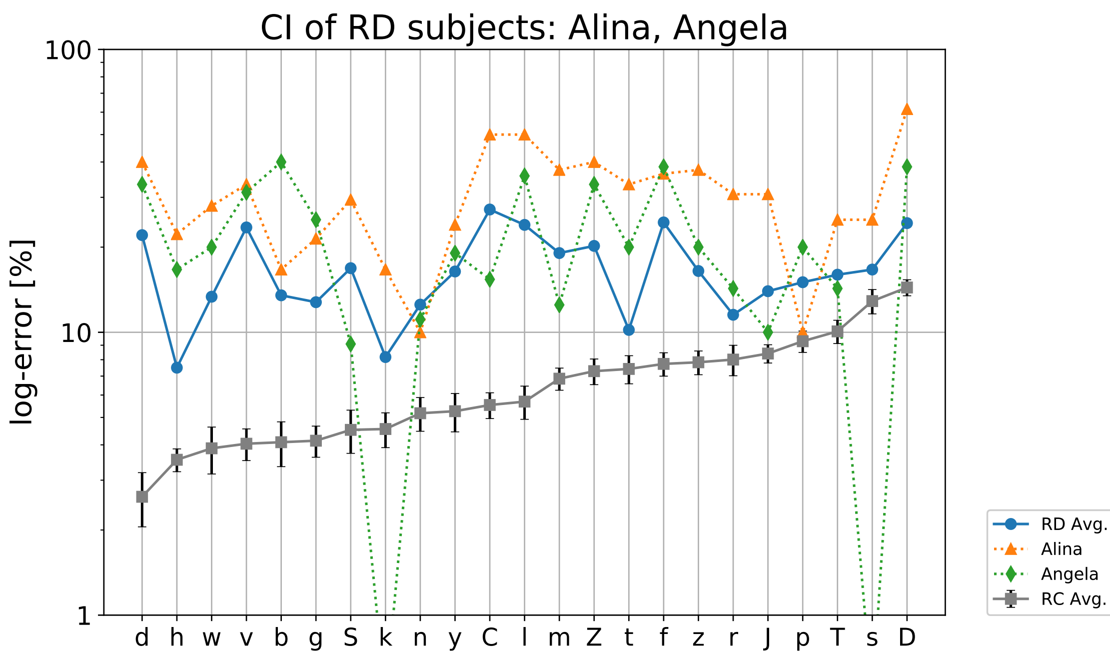

Figure 1. As indicated in the legend on the right, the two solid lines give the sorted average error for the RD (gray squares) and RC (blue circles) subjects, for the case of consonant initial (CI) sounds. RC: The RC error-sorted phones are shown along the abscissa (horizontal axis), with the lowest error on the left (/d/) (1.6 [%]) and the largest error on the right (/D/, i.e., /ð/) of 12 [%]. The standard deviation of less than ±1 [%] is shown for the RC-average subject, superimposed on the grey line. The ordinate (vertical axis) gives the phone error in [%]. Note the log scale. RD: Two typical RD subjects, RD-Alina (◇) and RD-Angela (△) are also shown. Note also how the RD-average error (solid-blue circles) is approximately constant, between 10-20%. This error is not typical of the RC-average subjects, who show an increasing error (and small variance), from around 3% to 15% error. (color online)

Results for random effects are presented in Table 5. We found a main effect of subject ($\chi^2(1)$=983.918, p<0.001), indicating that individual subjects (ignoring RD status) varied in their error rates. We also found an interaction between subject and phone type ($\chi^2(1)$=33.472, p<0.001), indicating that subjects differed in the likelihood of C versus V errors. Similarly, we found an interaction between subject and phone position ($\chi^2(1)$=33.545, p<0.001), indicating that subjects differed in the likelihood of making errors for sounds initial versus final position. There was also an interaction between phone and position ($\chi^2(1)$=4.218, p=0.040), suggesting that the effect of syllable position varied depending on the specific phone. Lastly, we found a three-way interaction between subject, phone type, and phone position ($\chi^2(1)$=15.275, p<0.001), indicating that subjects varied in their error rates as a function of both phone type and position within the syllable. Other interactions in the random effects analysis

[4]Note that this test depends on the residuals being Gaussian, which may or may not be the case. Thus, the raw data is displayed in several figures to support these assumptions.

[5]These mean error rates are computed at the individual trial level. Because some subjects had more trials than others, the subject-level error rates are slightly different (mean RD: 18.8%, mean RC: 6.7%).

were not significant. In summary, the statistical analyses confirm the observations of Figs. 1-5 that the RD listeners perform significantly worse for the SCO task than the RC listeners.

In the next section, we explore these results in greater detail, examining both the group differences (RD vs. RC performance) and individual differences between subjects.

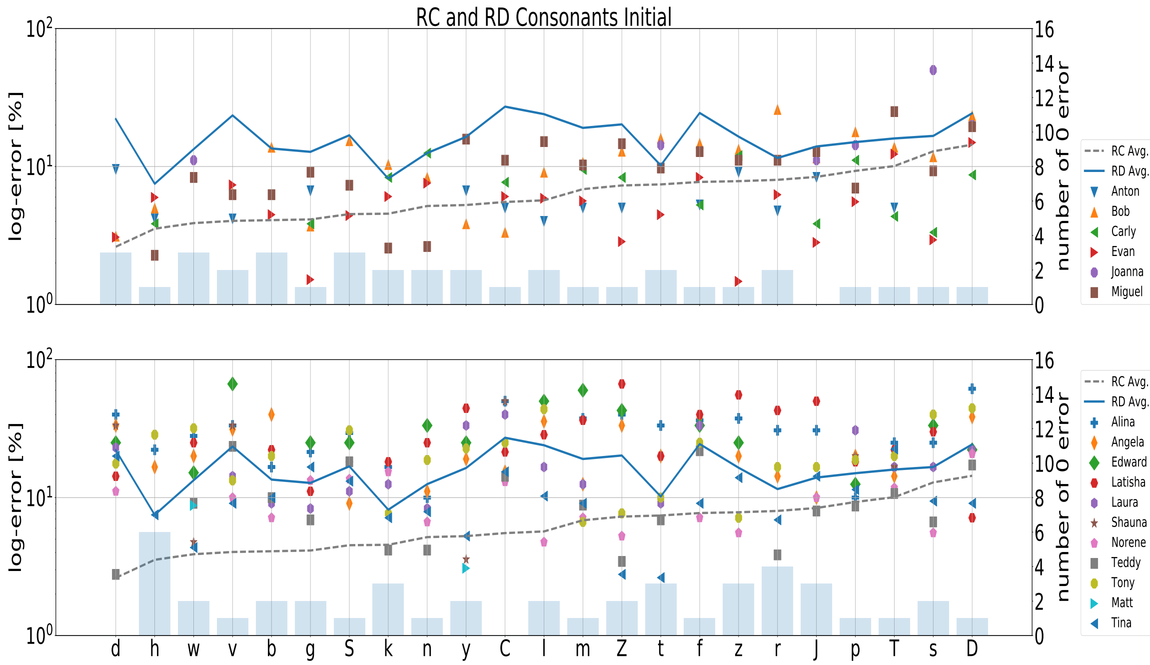

Figure 2. Sorted error for RD- and RC-subjects for CI sounds, following the same format as Fig. 1. The upper panel is the sorted error plot for the RC group. Data points show each RC subject's error. The histogram indicates the number of RC individuals having 0 CI error. Recall that the RC group has approximately half the number of children (6) as the RD group (11). Note how the RC points almost entirely lie below the RD-average (blue solid line), consistent with a significant separation of the two groups. The lower panel is the sorted error plot for the RD Group. The dashed grey line represents the RC-average error for CI sounds, while the solid colored line is the RD average error (same as the upper panel). Data points show each RD subject's error, and the histograms give the number of individuals with 0 error for the CI phones. Note how most of the RD points lie above the RC-average (grey dashed line). (color online)

Sorted errors SCO

To illustrate how the SCO graphs are organized, consider the case of CI stimuli, shown in Fig. 1. The abscissa of the graph represents the phones, sorted by their RC-average error. Special attention must be paid to the ordinate (vertical) axis which shown the phone probability of error, on a log % scale. The bottom of the scale is 1 [%] error, the light thin shaded line half way up shows 10 [%] error, while the top is 100 [%] error.

To quantify the differences in performance for the RC and RD subjects, the abscissa is sorted according to the the RC averages for each phone, as show by the shaded-grey line with squares, with standard deviation (SD) marks). Since the RC subjects have uniformly small error, they may be averaged to create an "average-normal" pseudo-subject. As shown, /d/ has the lowest RC average error, between 2 and 3 [%], /h/ has the second smallest RC-average error (3.5 [%]), while to the far right /D/ (i.e., /ð/) has the largest error of 12 [%].

The solid line with filled circles is the average of the RD subjects, which mainly fall between 10 and 25 [%] error, independent of the target consonant (abscissa). This is in contrast to the average-RC curve which is monotonically increasing (from 2.5-12 [%]). Also shown for reference are the errors of two typical RD subjects, RD-Alina (△) and RD-Angela (◇). RD-Angela hovers around the average while RD-Alina is mostly above the average, between 15 and 35 [%].

Notable are the huge differences between the individual RD scores. Angela makes zero error for /k, s/ (independent of the vowel) whereas RD-Alina makes around 20 [%] error for these two sounds. For /n/ they both make slightly more than 10 [%] error. For /p/ RD-Alina makes 10 [%] error while Angela makes twice that. The idiosyncratic variability is more obvious and detailed in the NSCM directed graph charts discussed below.

Consonants in syllable-initial (CI) position. Next we examine the sorted error plots for individual RC and RD listeners in CI, CF, VI, VF orders. These four charts (Figs. 2–5) provide the most important observations of the manuscript.

A limitation of plotting error on a log scale is that one cannot present zero errors (the log of zero is $-\infty$). Thus, we added histograms along the bottom to show the number of zero-error sounds (see axis labels on the right) for the RC and RD subjects.

These zero-error phones influence the RC-average error. Generally, there are one to three children (out of 6) from the RC group with zero-errors for each phone.

The upper panel of Fig. 2 is similar to Fig. 1 with the exception of the raw subject data points are shown for the RC subjects, as the scatter of points around the dashed-gray average-RC curve, which ranges from 1.6 to 12 [%]. The dashed curve is lower than the means of the subject points due to the large number of zero-errors. Note the low errors for RC-Evan, who is outstanding at the SCO task. Also of note are RC-Miguel and RC-Bob, who are consistently close to the RD-average, for many (or most) of the initial constants.

The lower panel differs from the upper panel in that the scattered data points correspond to the RD subjects, thus they are centered on the solid-blue average-RC curve. The average RD line is somewhat lower than the means of the scattered points due to the large numbers of zero errors for the RD.

Overall, the individual RD errors (scatter of points) in the lower panel is above the RC-mean (dashed line), indicating that the best (lowest error) individual performance of the RD subjects for individual phones is almost always worse than the average performance of RC subjects. Looking in more detail at the RD listeners, more than half (7/12) of the children in the RD group (e.g., RD-Alina, RD-Angela, RD-Edward, RD-Latisha, RD-Laura, RD-Shauna, RD-Tony) struggle with discriminating all consonants in the initial position, starting with the lowest (best RD performance) of 7% error. A few RD subjects did as well on some consonants as the average of the RC group (e.g., RD-Norene, RD-Teddy, RD-Tina), but these are notable exceptions, as may be seen in the figure (a few of the RD individual consonant data points are below the dashed grey line). It is possible that a careful look would reveal that one or all of these subjects are closer to the RC group than the others in the RD group. Given the complexity of reading disability, one might expect that the SCO consonant test does not apply to every reading disability. Much more data would be required to test this hypothesis. Regardless, the vast majority of RD children were not able to pass the SCO consonant discrimination test (i.e., most of the individual consonant points are above the dashed line).

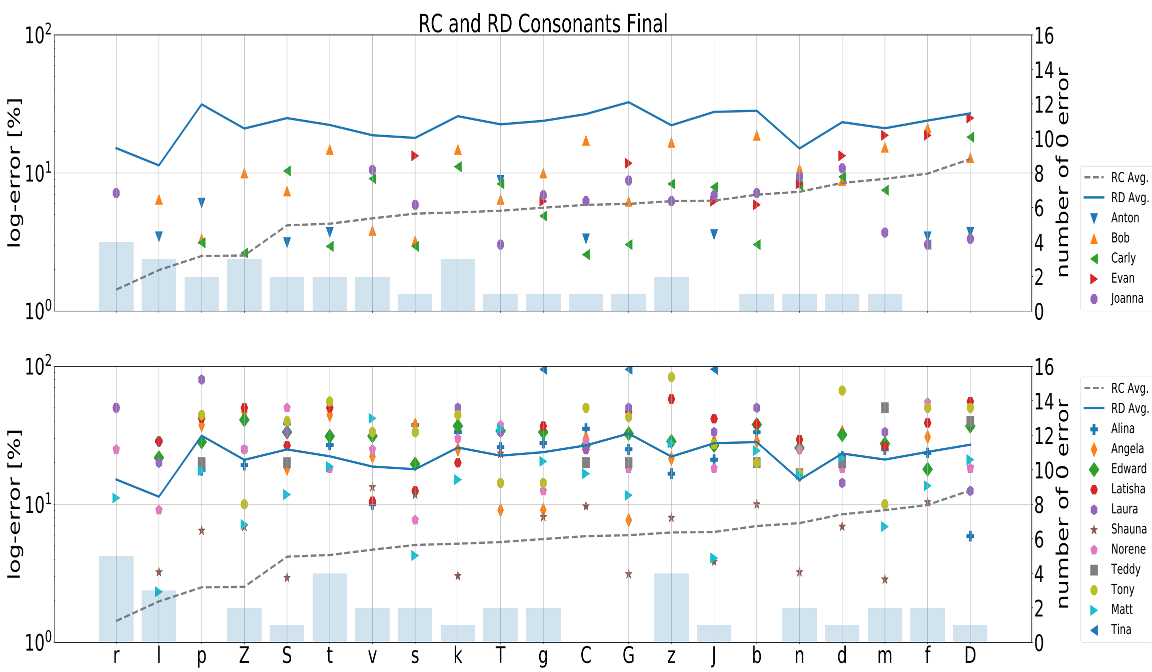

Figure 3. Sorted error for RD- and RC-subjects, for consonant final (CF) sounds, in the same format as Fig. 2. The lowest RC-average error phone on the left is /r/), and the highest error phone on the right is /ð/. As for CI sounds, note how the average of the RD-errors are nearly constant (the solid line is horizontal, hovering around 20-30 [Note the very low errors for RD-Shauna and RD-Matt with several CF scores below the RC-average. (color online)

Consonants in syllable-final (CF) position. Figure 3 show results for consonants in final position, in the same format as Figure 2. The top panel shows that the RD average error is well above the upper bound of the scatter of individual RC data. Thus, CI and CF sounds are similar in terms of their effects on the two subject groups.

In the lower panel showing the individual RD errors, RD-Alina had more than 10% error for almost all consonants in final position. RD-Latisha, RD-Norene, RD-Edward, and RD-Angela are also on the slightly higher error side.

Among the RC children, RC-Anton had no trouble discriminating 10 out of 24 consonants, with the worst (largest) probability of error for all sounds being slightly more than 10%. RC-Bob had the largest average error, approaching 20 [%] for /f/. RC-Joanna, RC-Evan and RD-Teddy performed with less than 10% error, with RC-Joanna having no error for 7 out of 24 consonants. RC-Miguel and RD-Shauna had a slightly higher error, ranging from 5 to 30%.

Individual differences

Also, RC-Carly, RD-Alina, and RD-Angela had high error rates for some phones and very low error rates for others, meaning they showed more abrupt increases in perceptual difficulties across consonants than other children. Errors for RD-Latisha, RD-Norene, and RD-Edward are stable for different phones, meaning every phone caused similar difficulties. RC subjects, along with RD-Shauna and RD-Teddy had relatively uniform smaller probabilities of error for all consonants. For CF sounds, RD-Shauna and RD-Teddy are the cross-over participants. Overall, there are differences between the individual RD subjects, showing idiosyncratic errors.

There are also individual differences for subjects between their CI and CF errors: Data collected from RC-Carly for CF sounds revealed a minimum probability of error of around 3%, while for CI sounds, there was zero-error for 7 out of 24 consonants.

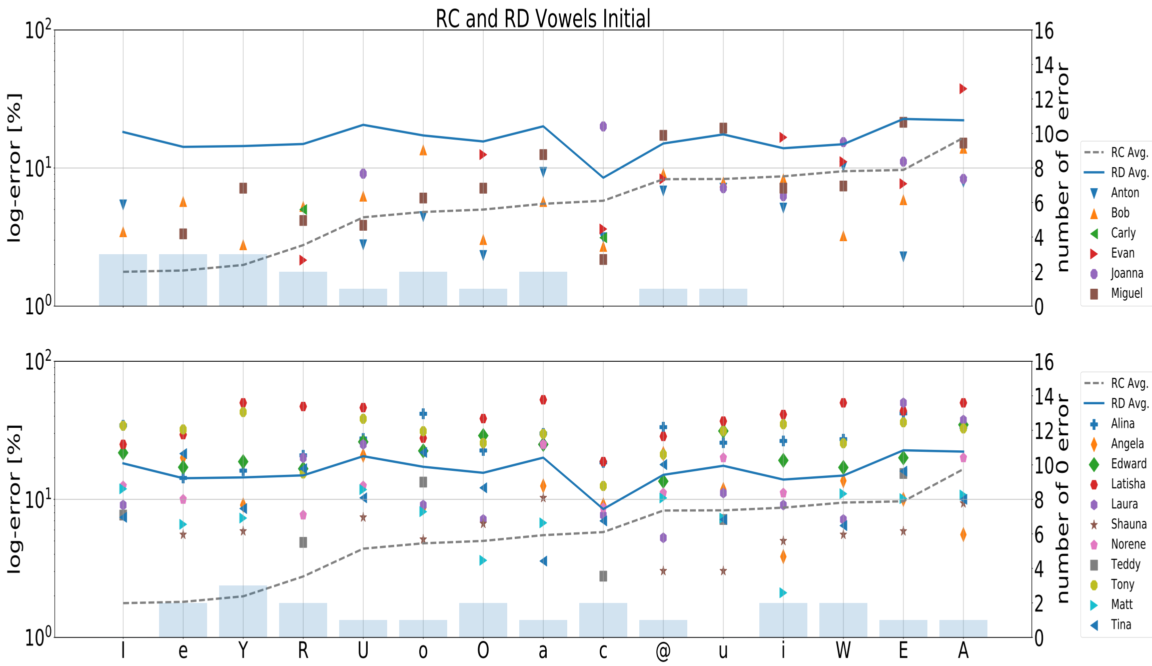

Figure 4. Sorted error for RD- and RC-subjects, for VI sounds, in the same format as Fig. 2. The lowest RC-average error phone on the left is /ɪ/, and the highest error phone on the right is /ɑ/. As observed in Figs. 2 and 3, the RD-average VI error is nearly constant, between 10 and 20 [%] error. (color online)

Vowels in syllable-initial (VI) position. Figure 4 shows data for vowels in syllable initial position in the same format as Figure 2. Previous high performers like RC-Anton and RD-Teddy still outperformed everyone, having between 2% and 20% error. RC-Bob showed improved discrimination of vowels in initial position, compared with discrimination of consonants in initial position. RD-Latisha struggled the most with discriminating vowels in syllable-initial position with an error rate of 20% or greater. Along with RD-Alina and RD-Latisha, RD-Tony and RD-Edward were among the worst performers, with probabilities of error ranging from 10% to slightly higher than 40%; while RD-Angela, RD-Laura, and RD-Norene along with RC-Joanna, RC-Evan, RC-Carly and RC-Miguel were in the middle with 2–30% error.

Vowels in syllable-final (VF) position. Figure 5 shows results for vowels in final position, in the same format as the previous figures. As in the other conditions, the VF sounds show a clear separation between the RD and RC groups. RC-Anton had the lowest error of all subjects, with zero-error for 8 of the 15 vowels; the highest error rate was 7%. RC-Joanna was the second best, with maximum error for all vowels less than 10%. The data across all conditions showed that RC-Joanna's phone discrimination was better when the phone was in syllable-final position, whether the phone was a consonant or a vowel. The rest of the lower error subjects were RC-Evan, RC-Carly and RC-Bob. RC-Miguel had the highest error among all children, starting at 30%. The rest of the high error group were RD-Tony, RD-Teddy, RD-Latisha, RD-Edward, RD-Alina, RD-Angela and RD-Norene.

NSCM results

While the SCO confusions may be visualized as *sorted error plots*, the NSCM confusion errors are best shown as directed graphs (Figs. 6–10).

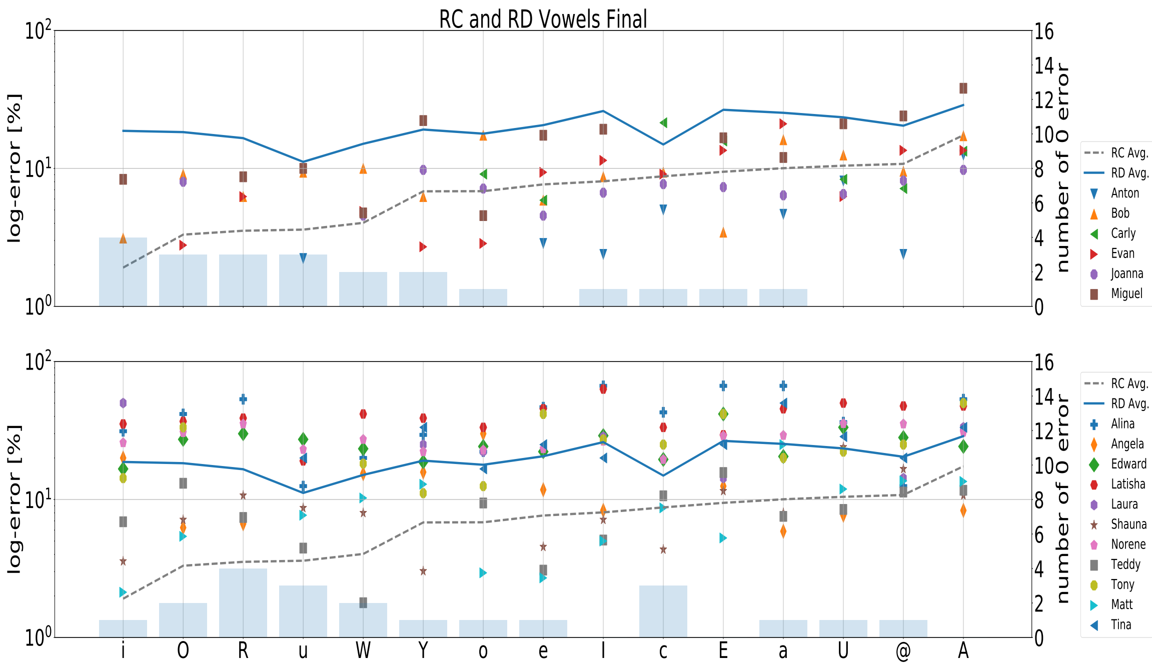

Figure 5. Sorted error for RD- and RC-subjects, for VF sounds, in same format as Fig. 2. The lowest RC-average error phone on the left is /i/), and the highest error phone on the right is /ɑ/. As before, the RD-average VF error is nearly constant, ranging between 10 and 30 [%] error. (color online)

Given the large difference in errors between the two subject groups for the SCO task (Figs. 2–5), one would expect large differences in the NSCM task for the same subjects. What we found is that the RC, as a group, make very few errors (as expected) while the RD individual subjects make very large idiosyncratic (different) errors, based on phone type and position.

Figures 6–9 reveal that there are large differences between RD and RC. For example, RC-Anton and RD-Noreneprovide a nice comparison as to the specific phone errors across the groups. Similar directed graphs (not shown) for the remaining subjects reveal that these idiosyncratic differences are valid for all the RD subjects while low-error RC subjects are similar, but differ in those confusions when they do make errors. Due to the much smaller error for the RC subjects, they are sufficiently similar that we may treat them as a single average-normal group.

The results of the SCO task reveal substantial differences between the RD and RC subjects and show a large number of individual differences between subjects. The data from the NSCM task provide a useful way to capture these individual differences. Here, we present a few examples of the NSCM data to illustrate this point. As summarized above, in the NSCM task, subjects heard a single MaxEnt syllable (CV or VC) and repeated it back. Two expert transcribers wrote down which phone they heard the subject produce. This results in a *confusion matrix*, which codes the confusions. While the effect is small, errors in the NSCM task were slightly smaller than those in the SCO task. It is not clear why, and this was unexpected.

The data from this task are analyzed as directed graphs, which provide a graphical method for summarizing a confusion matrix. In these graphs, nodes represent individual phones that the listener hears, and connections between nodes depict the listener's confusions. Connections from the node back to itself are correct responses. Connections to other nodes are errors, with the percentage written above each connection indicating how often that particular confusion was made.

First, we examine the confusion matrices collected for the two best RC subjects, RC-Anton and RC-Evan. Then, we investigate CI and CF sounds for RD-Norene who is close to the RD average in performance. As we shall see in these comparisons, the RC subjects make few errors, while RD-Norene makes many more errors. Moreover, the directed graphs for each subject's data capture the confusions made for each phone. Given the confusion matrix, interventions could be developed based on the specific phones for which an RD subject needs the most help. These methods could also be used to track improvements in the subject's phone recognition over time.

RC subjects. Figure 6 shows the confusion matrix data for RC-Anton, who has the lowest error of the RC group. Overall, the graph shows almost no error, as indicated by the nodes without connections to other nodes in the graph. A pruned graph highlighting a subset of this full error graph is shown in Figure 8. Confusion data for RC-Evan, who has the next lowest error, are shown in Figure 7; low-level errors for RC-Evan are shown in the right panel of the pruned graph (Fig. 8).

The pruned graphs reveal a number of interesting details about the low-level errors for the RC subjects. For example, for RC-Anton (left) on the top-left of the pruned graph, there is a single arrow, looping back to the node labeled /d/, indicating that this sound is recognized correctly 99% of the time (i.e., 1% error) with a threshold around chance. A more interesting case

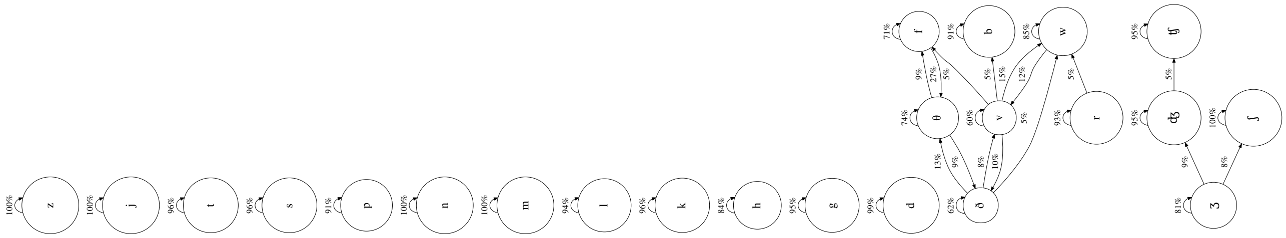

Figure 6. Directed graph for RC-Anton for CI sounds in the NSCM task. RC-Anton made less than 10[%] error on all the sounds other than /ð/ and /θ/. High-error sounds are shown as smaller circles. These three high error sounds are detailed in Fig. 8 (left).

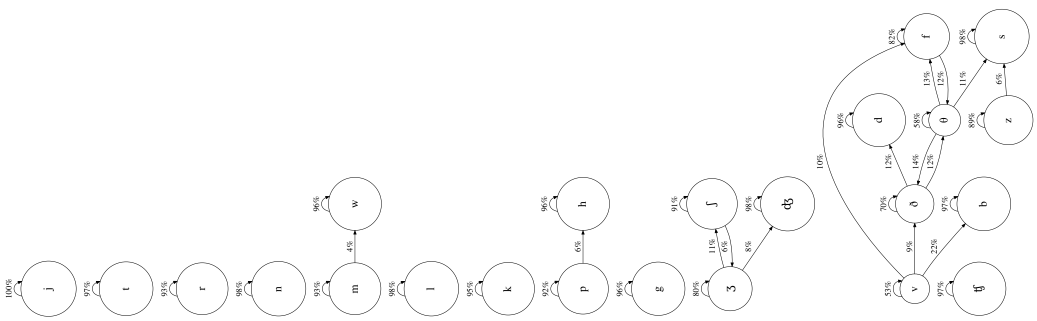

Figure 7. Directed graph for RC-Evan for CI sounds in the NSCM task. RC-Evan made less than 10[%] errors on all the sounds other than /θ,ʒ,v,ð/. These three high error sounds are detailed in Fig. 8 (right).

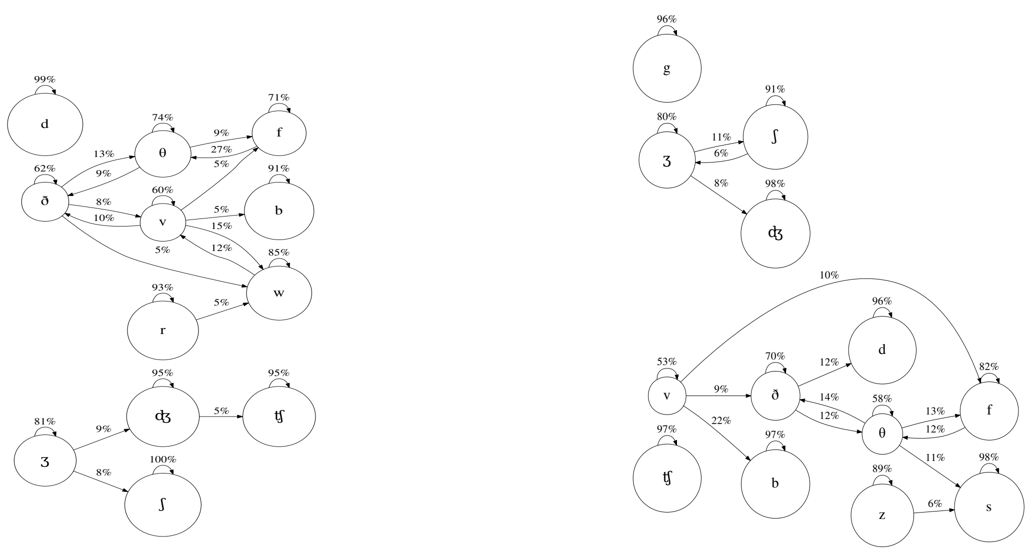

Figure 8. Pruned directed graphs for RC-Anton (left) and RC-Evan (right) for CI sounds. In the pruned graphs, the low error tokens have been removed, and the directed graphs are displayed vertically for improved readability. Smaller circles indicate larger error, such as for /ð,v,f,θ/ (left) and /ʒ,v,ð,θ/ (right).

is /ð/, which has only 62% correct responses and is confused with /θ/ (13%), /v/ (8%), and /w/ (5%). /θ/ has a higher score of 74% correct, having 9% confusions with /f/ and /ð/. Presentations of /f/ are almost entirely confused with /θ/ (27% of the time). This is RC-Anton's largest error phone, which is common for all RC listeners. Another common error for the RC group, that Anton does *not* make, is confusing /f/ as /v/.

RD subjects. The RD subjects have, in general, very complex directed graphs, with many documented confusions between consonants. It is instructive to compare CI and CF sounds for a single listener. This is illustrated in Figs. 9 and 10, which compare the graphs for CI versus CF sounds for RD-Norene. More data like these need to be carefully analyzed in follow up studies; These directed graphs are simple examples to demonstrate how they supplement, complement and amplify the SCO results. The large number of errors, and the specific type, would be very helpful to a speech therapist. It will be very interesting to establish when these errors are correctable, following therapy.

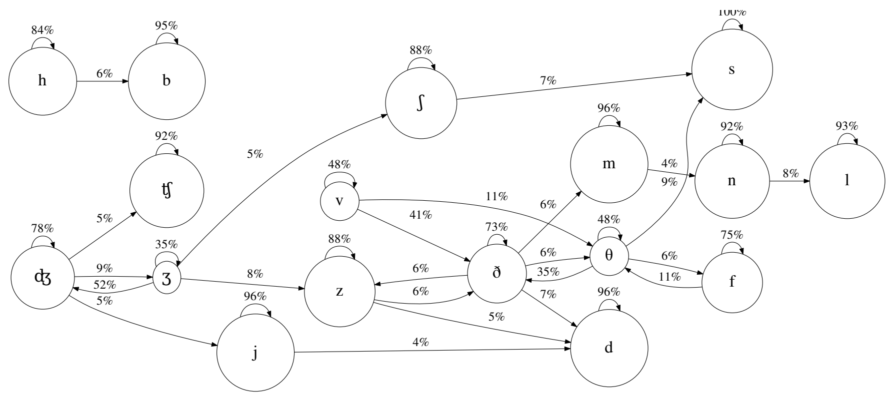

Figure 9. Directed graph for RD-Norene for CI sounds. The number of errors, along with the detail, should be very helpful to a speech therapist, who could concentrate on the most important high error sounds (the smaller circles). It would be very interesting if these errors are correctable, following therapy. It would be even more interesting if correcting these errors would lead to improved reading success.

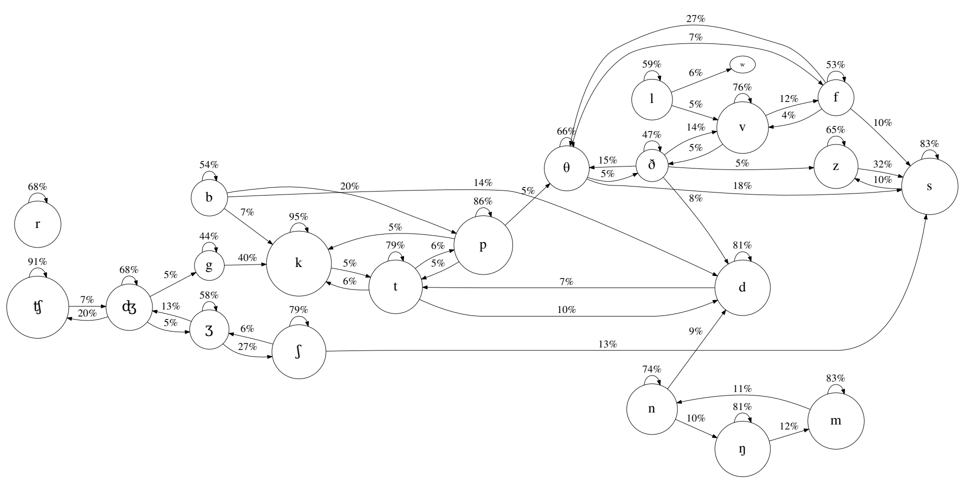

Figure 10. Directed graph for RD-Norene for CF sounds, showing a large number of CF errors compared to most of the RD subjects. This shows the overall poor discrimination of CF sounds. The CF scores are much worse than RD-Norene's CI scores shown in Figure 9.

Discussion

Without an understanding of the underlying causes of RD, there can be little hope of a successful diagnosis and treatment. To validate a theory it must be supported by experimental data. Past theories about reading dysfunction, beginning in the 1950s, were based on the widely held observation that *phonological awareness* was the key to understanding reading, due to the observed direct correlation between phonological awareness and RD (Torgesen, 2004). However this did not explain the exact nature of RD. Moreover, models of reading emphasize the importance of decoding skills in leading to read, while tacitly assuming encoding is fully functional. Based on our observation that our RD children have a large phone errors, thus compared to RC children, encoding is not fully functional. Thus the need for fully functional encoding seems to have been under appreciated.

Based on our experimental results, we propose that the intervention based on phonological awareness and decoding skills *assumes* that the RD child has fully intact phone encoding. Given the large difference in error between RC and RD, this assumption is clearly wrong. As discussed in the Introduction, many studies have attempted to experimentally link RD and speech perception. These experiments were largely unsuccessful, and it now seems clear why they had negative results. First, most experiments aimed at linking speech perception and RD have emphasized phoneme categorization by using meaningful stimuli (i.e., phonemes that incorporate information from the context channel). Context results in reduced error, making the test much less sensitive (Allen, 1996, 2005).

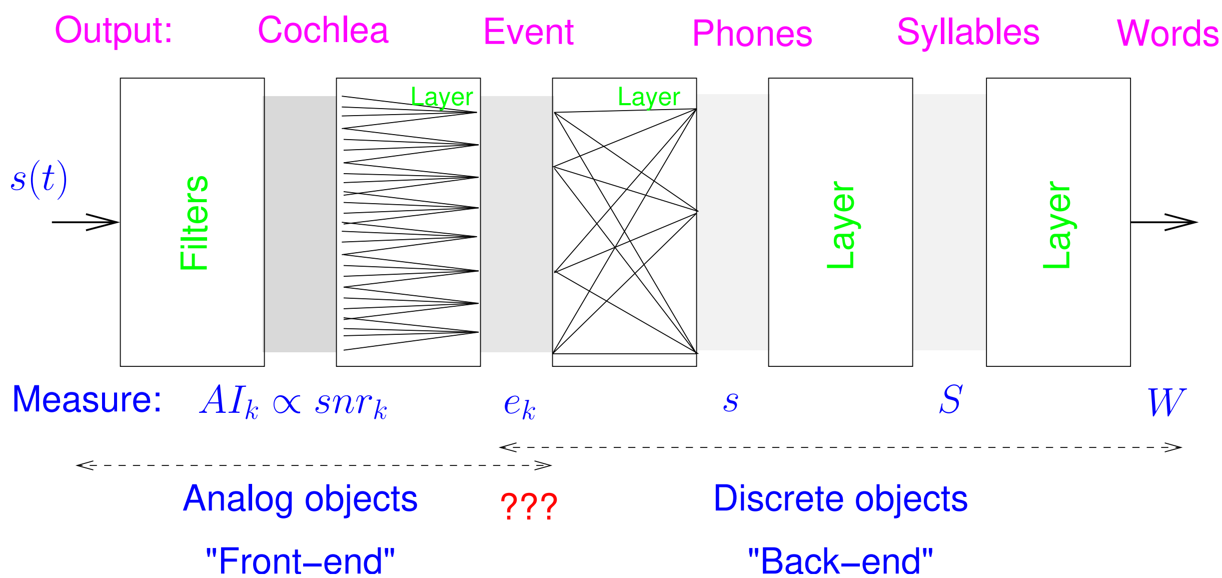

Figure 11. Model block diagram summary of human speech recognition, based on Fletcher's Articulation Index (AI). At the top of each block is a label that identifies the physical process. The equations below the boxes indicate the probability measure defined at that level. Speech $s(t)$ enters on the left and is processed by the cochlea (first block), breaking the signal into a filtered continuum of band-passed responses. The output of the cochlea is characterized by the specific AI_k, a normalized signal to noise ratio (SNR), expressed in dB units. The second box represents the work of early auditory areas of the brain, which are responsible for event identification in the speech signal, such as onset transients (Li and Allen, 2011). The third block puts these basic events together, defining phones. The remaining blocks account for context processing (Allen, 2005). The formulas for the measures are discussed in Allen (1996, 2005). (color online)

Based on our present understanding of speech perception, as outlined in Fig. 11 (Allen, 2005), low-level phone perception plays a critical role understanding speech, thus phone perception must be in place when one learns to talk. Based on our experimental results, it appears that learning to read requires a deeper cognitive foundation in phone perception than required for conversational speech. Namely decoding requires a better developed cortical phone map. For example, minor disfluencies can go unnoticed in spoken language communication, but impair, or even block decoding. This may be because reading requires a much more more detailed error free representation of the phone set. We are suggesting that the neural circuits that encode phones need to be more deeply embedded to support decoding, than required for conversational speech (Li and Allen, 2011; Singh and Allen, 2012; Toscano and Allen, 2014). The experimental evidence presented here supports this theory.

When teaching a child to read, it is important to verify that they have a full command of the underlying phone set. The SCO and NSCM paradigms seem to be one way to assure this. In the past if was assumed that if the child could talk, that the phone set was fully developed. It appears that this assumption is not support by our experimental results.

We next explore the possible sources of this deficit and review how the two tasks used in this study provide insights into RD.

Learning to read

The processes of learning to understand spoken language and learning to read can be viewed in terms of three *information channels*: auditory, contextual, and visual (Miller et al., 1951; Bronkhorst et al., 1993; Allen, 2005; Kamhi and Catts, 2012). In order to understand speech, infants must first recognize phones or syllables (i.e., speech sounds via the auditory channel) and then associate speech with words (i.e., meaning via the contextual channel). These processes occur early for spoken language (between 1-18 months; Eimas et al., 1971; Jusczyk, 2000), long before learning to read.

Around 3-4 years, TD children can begin recognizing letters (the visual channel), and by around 5-6 years, they can begin learning to decode letters. Note that, critically, reading must be taught—it is *not* spontaneously acquired—unlike spoken language. Thus, the auditory and contextual channels may be viewed as fundamental to language comprehension, and ultimately to reading ability, in that they provide the foundation for decoding. This view is similar to that offered by connectionist approaches and other models of reading (Gough and Tunmer, 1986; Harm et al., 2003; Hoover and Gough, 1990; Plaut et al., 1996; Seidenberg and McClelland, 1989), which demonstrate that successful reading depends on decoding. These basic models can also explain the effects of word frequency and spelling-sound consistency on reading (Plaut et al., 1996), and they help to explain the effectiveness of RD interventions that focus on learning to decode (Harm et al., 2003). Other work has argued that children learn to read by blending onsets and rimes to recognize single-syllable words in print (e.g., the onset of "c" and the rime "at", blended to read the word "cat"; Goswami and Bryant, 1990).

Taken as a whole, current models of reading demonstrate that the key factor in children's success depends on their ability to decode letters. However, this view assumes that the mapping from sounds onto words (i.e., encoding) is already well-established so that children can accurately recognize basic speech sounds (phones). To study when and how speech perception (i.e., the contribution of the auditory channel) impacts reading ability, one must first eliminate contributions from the visual and contextual channels, particularly given that higher-level linguistic information can influence early speech sound encoding in meaningful contexts (Getz and Toscano, in press). Therefore, in the current study we avoid the use of text and meaning, in order to specifically test the auditory perception abilities of RD subjects. To achieve this, we use tasks that measure basic phonetic encoding abilities, following approaches used in studying perception of natural speech sounds by Fletcher (1995); Miller et al. (1951); Miller and Nicely (1955); Li and Allen (2011), as discussed in Allen (2005).

To see how disruptions in phonetic encoding may play a role in learning to read, consider the processes involved in speech perception. These were described in Harvey Fletcher's early work on the Articulation Index (AI) model of speech recognition, outlined in Fig. 11. To account for the identification of speech sounds, Fletcher identified a cascade of processing tasks, starting with cochlear filtering, followed by the identification of speech features (events), phones, syllables and eventually, meaningful words (Allen, 1996). Previous work does not suggest that a speech perception model is relevant to reading, but here we argue that, based on our experimental results, Fletcher's model provides key insights into the reading process. In particular, it suggests that accurate phone processing (encoding) must be functional before a child can begin learning to read.

RD interventions

Given the models described above, what determines whether children successfully learn to read or whether they have a RD? One challenge in answering these questions is that the criteria used to define and diagnose reading problems are necessarily complex. For example, dyslexia, a type of RD, specifically refers to difficulties with recognizing words in print (Kamhi and Catts, 2012).

However, dyslexia may be characterized on the basis of not only word recognition ability, but also spelling and visual decoding skills. Furthermore, dyslexia is correlated with performance on other aspects of reading, such as vocabulary, grammar, comprehension, and reasoning. Thus, dyslexia involves a broad range of abilities, many of which are interrelated. Despite the complexities involved in defining RD, we would expect that interventions focusing on letter decoding would be most successful, given the models of reading described above. Indeed, there is evidence that teaching decoding skills (aka phonics) specifically improves reading outcomes (NICHD/NRP, 2000a,b).

Lit review seems late.

Other interventions have focused on *phonological awareness*, another key skill for children to master in order to become successful readers (Cunningham, 1990). However, this view assumes a skill of explicitly attending to, judging, and manipulating the structure of speech sounds at the syllable level. It typically aims to assess how children recognize speech sounds in a meaningful contexts (i.e., words). There have been at least three decades of studies on phonological awareness (Goswami and Bryant, 1990; Kamhi and Catts, 2012), and research on the topic is ongoing (DeGroot et al., 2015). Interventions based on phonological awareness have been argued to improve reading ability. Bradley and Bryant (1983) presented a study on word training using CVCs with 65 children who had significant reading difficulties, performing at least one standard deviation below the mean in reading level. A subset of the RD group received phonemic awareness training via visual presentations, which caused them to improve greatly in reading compared with the group that did not receive such training. Because the cohort was taken from a group of subjects from a previous test of reading ability on the same words, the experiment was longitudinally controlled. Hence, the experiment provided one of the first demonstrations of a causal relation between phonological awareness and improvement in reading ability. Bradley and Bryant (1978) found a similar relationship between phonological awareness and improvement in reading abilities.

However, as noted above, it is difficult to isolate speech perception problems using tasks that only involve meaningful words, due to context effects (Miller et al., 1951). Thus, a useful strategy is to break the problem into smaller parts, along its natural boundaries, and tackle them one after another, as summarized in Fig. 11 (Allen, 2005, Fig. 1.3, p. 18). This is the approach we take in the current study.

RD and speech perception

Both decoding and phonological awareness interventions assume that the child can accurately encode speech sounds, and both of these skills depend on processes that lie at the very last stage of the speech recognition process. This is illustrated by Fletcher's model (Fig. 11), where phones (i.e., individual speech sounds) are mapped onto syllables, which are woven into meaningful contexts such as words. Thus, the most basic (i.e., earliest) step in perceiving speech is phone recognition (Li and Allen, 2011). As a result, perception of phone differences may well determine the success of reading and language comprehension. Given this, we aim to specifically investigate low-level phone perception as the first step in the analysis of reading problems. This approach is supported by other reading research as well. For example, Share and Stanovich (1995) note that contextual information plays a less important role than phonological sensitivity in reading comprehension, suggesting a reason to focus on phone perception.

Is this a repeat?

Several earlier studies have investigated the role of speech perception in RD. Many of these studies have used the *categorical perception* paradigm, where speech sounds are varied along specific acoustic-phonetic continua and listeners' identification and discrimination responses are compared (Liberman et al., 1957).[6] Brandt and Rosen (1980) used this approach to measure perception of speech sounds in 12 RD children and four TD children who served as reading control (RC) subjects. Listeners were tested with synthetic CV syllables varying along a voice onset time (VOT) continuum distinguishing voicing (/b,p/) and formant transition continua distinguishing place of articulation (/b,d,g/), each presented in the context of the vowel /ɑ/. They concluded that there was no significant difference in categorical perception between the RD and RC children.

Other studies have examined RD children's perception of speech sounds for meaningful words. Manis et al. (1997) investigated this with 25 dyslexic (DYS) children (4th-10th grade), whose performance was compared with 25 chronological age (CA) matched children (5th-8th grade) and 24 reading level (RL) matched children (2nd-3rd grade). The spoken words *bath* (/bæθ/) and *path* (/pæθ/) were used as testing materials to measure /b/-/p/ categorization ability of the groups of subjects. In general, the slopes (b) of the categorization curves for the DYS group were shallower than those of the CA group but were not different from the RL group. Similarly, Joanisse et al. (2000) looked at phoneme categorization with 61 DYS 3rd graders (7-10 years old), 52 CA matched 3rd graders, and 37 RL matched 1st and 2nd graders (6-8 years old), testing phoneme categorization with the word pairs *dug-tug* and *spy-sky*. They found no overall difference in categorization between the DYS and control groups, but a subset of DYS subjects showed shallower categorization functions for both speech sound contrasts.

Possible sources of RD speech perception errors

The results of this study suggest a similarity between RD children and hearing impaired listeners. RD children typically have normal hearing and may have normal speech production. Thus, on the surface, there is no obvious reason to question whether they might have a speech perception problem. But given the experimental results presented here, we now see that there are large differences between RC and RD children in their speech perception abilities, and strong similarities between hearing impaired and RD children, with the important difference being that RD children have normal hearing.

It also seems clear from phonological awareness data, that children with RD develop problems early, before first grade, which we interpret as the early failure to learn to discriminate a subset of phones. It is unclear what the source of these speech perception deficits is, but one plausible explanation is a lack of exposure to speech sounds during early development. Given the idiosyncratic nature of the errors, cochlear hearing loss does not seem to be a likely cause.

It is known, however, that episodic middle ear problems can cause up to a 50 dB hearing loss, for months at a time (Williams and Jacobs, 2009). Perhaps this is a major source of the problem, but the impact is not obvious because most speech sounds can be adequately discriminated, thus speech perception appears to be normal. Yet, when test in their ability to discriminate individual phones, RD children make many more errors than RC children.

If temporary middle ear loss is the main cause, then it seems likely, given their normal cochlear function and normal cognition, that speech perception training, highly focused on the specific errors that the listener makes, might dramatically improve their phone discrimination, and thus their reading. Only with early detection, along with the proper diagnostic feedback, and drilling on the high error sounds, can we determine if this will resolve the problems. There is significant evidence, that given enough resources, reading problems can be resolved. But with no resources and no diagnostics, these subtle problems are untreatable.

Results from the SCO and NSCM tasks

The two tasks used in the current study (SCO and NSCM) nicely complement each other. The SCO task can quickly determine which phones cause errors, but it does not indicate which sounds they are confused with. In contrast, the NSCM task gives a detailed map of the exact nature of the errors, in the form of a confusion matrix (directed graph). It should be possible for a speech therapist to use this information to improve phone recognition for RD children. It would be even more encouraging if, after improving their speech recognition scores, reading ability improves. This suggests experiments for future research.

[6]These studies have often used synthetic speech in order to create sounds that vary along specific acoustic dimensions and control for variability between speech sounds (Liberman et al., 1967). We now better understand the nature of this variability and can control for it (Li and Allen, 2011), and more recent work has used natural speech sound continua instead (Toscano and McMurray, 2012; Toscano and Lansing, in press). In the current study, natural speech is chosen over synthetic speech as it provides a rich set of acoustic cues used by the auditory system that are distinct from those found in synthetic speech sounds (Li et al., 2010; Li and Allen, 2011).

The results of the SCO task also clearly demonstrate how much worse the RD children performed compared with the RC children in phone recognition. Previous work has found mixed evidence for speech perception deficits in children with dyslexia (Hazan and Adlard, 1998; Ziegler et al., 2009; Messaoud-Galusi et al., 2011). However, from our results, we conclude that notable phonetic deficits exist for RD children compared with children who have normal reading ability. As shown by the sorted error plots, in general, the RC group performed much better than the RD group on the SCO task, with more zero-error phones and lower probabilities of error than the RD children.

The phonetic encoding process was engaged, perhaps even taxed, during the SCO task. The children were not told whether the change of phone was in the syllable-initial position or in the syllable-final position. They were presented with utterances with the only instructions being that they would hear three syllables and then should find the one that had a change in the phone. The children with RD struggled with mismatched phone representations during the SCO task, requiring greater effort to store in memory all the three spoken sounds, instead of processing the phones easily and correctly, without effort, as was done by the RCs. Some of the RC children (e.g., RC-Miguel, RC-Bob, RC-Evan) had the same error magnitude, but as a group, the RCs significantly outperformed the RD children. More experiments on the SCO task with various numbers of potentially confusing phones (e.g. a triad, quartet, or quintet of syllables) would be needed to fully characterize reasons why the 3-trial SCO task had larger error than the single-trial NSCM task, but a memory overload on a task that has poor cortical representation, seems a likely candidate.

Outliers in the phonetic encoding tasks

Although the RC group performed significantly better than the RD group in the SCO task, not all individual RD subjects performed worse than individual RC subjects. Based on close inspection of Figs. 2–5, RD-Norene and RD-Teddy performed at the RC-average error for consonants in initial position, RD-Shauna and RD-Teddy performed at the RC-average error for consonants in final position and vowels in initial position, and RD-Shauna performed at the RC-average error for vowels in final position. These outliers in the RD group all demonstrated normal phone recognition, similar to the RC children. Thus, either there are other factors than a weakness in phonetic perception, or the patterns reflect the highly idiosyncratic nature of the reading disorders for these individual children.

Conclusions

Given the fundamental role of phone encoding and decoding in human communication (e.g., Fig. 11), it naturally follows that phone encoding must play a role in learning and its disorders. **(1)** RD children have a significant speech perception problem in identifying open set syllables, despite normal pure-tone hearing and normal language processing abilities. **(2)** Individual RD subjects were highly different in their confusions (idiosyncratic), yet were consistent in the task (low variance, high error) **(3)** For the NSCM task, the errors were somewhat reduced, indicating that the single-interval task was easier (lower error) than the SCO three-interval task. **(4)** These results are at odds with several previous studies which shows weak indication of phoneme identification impairment in RD children. These differences are likely due to (i) the differences in methods of the present study (SCO and NSCM), (ii) with large numbers of trials per child (e.g., 1,500), (iii) the use of natural speech (iv) the large numbers of mixed-gender talkers (e.g., 18). Edits-2-here.

Additional conclusions: The results of the current study demonstrate that the model of speech perception in Figure 11 can help us understand reading difficulties: accurate phone encoding at early stages of processing is a pre-requisite for other steps in the process. If children can complete this fundamental step, then they can learn to read. This seems similar to what we have learned abut the deaf and the need to implant the cochlear implant in the first two years of life, after which it is much less likely to function. It is assumed that when implanted early, the auditory signal is decoded by the auditory brain with sufficient accuracy that the CI child can *main-stream* into the classroom.

This is one of the take-home message of the current study, and it is confirmed by the results presented above. RD subjects made considerably more phone recognition errors than RC subjects, and individual differences were observed in the type of phone (consonant vs. vowel) and syllable position (initial vs. final) that caused errors. Overall, the data suggest that early phone perception is a critical factor in normal reading development.

In conclusion, in response to our initial questions:

1. Without access to visual (i.e., letter) and contextual (i.e., word) information, do children with RD show a phonetic encoding deficit? **Yes.**

2. If so, what type of task is most effective in diagnosing RD? **SCO.** What task is most informative about RD regarding perception of specific speech sounds? **NSCM.**

3. Do RD and TD children have common or unique phonetic perception patterns (i.e., do they have similar or dissimilar patterns of speech sound confusions)? Alternatively are they idiosyncratic? If so, to what extent? **Errors are idiosyncratic to a large extent.**

Author Contributions

Historical work flow: Author JA designed the SCO and NSCM experiment and wrote the code for the SCO task (c. 2005). CJ initiates a relationship with *The Reading Group, Urbana IL* and collects the data, with help from several graduate students (2006). JL analyzes the data using ANOVA in her doctoral dissertation (c. 2016). JT recommends, and then provides an extended logit statistical analyses (2017); YW and JA finalized the thesis (2018), and JL graduates (July, 2018); YW writes the present manuscript base on JL thesis, including an extended analysis, introducing Figs 2-5. JA and JT proofread the manuscript, and YW submits (2019).

Acknowledgments

We would like to thank the many students who were involved in collecting and coding data for this study. Sandeep Phatak (who also wrote the NSCM task data collection program), Bryce Lobdel, and Mary Kubalanza were involved in various aspects of the data organization. Angela Menon, Christoph Scheidenger, and Jonathan Buei helped translate the computer log files into Matlab© format. The project was then completed as partial requirements of author Jie Lu's dissertation. Jie did the final analysis of the data. Without her contribution, this work would not have seen the light of day. This research was partially funded by Phonak Hearing Systems, Stäfa, Schweiz. Contact Person Stefan Launer, Director Research and Technology.

References

Allen, J. B. (1996). Harvey Fletcher's role in the creation of communication acoustics. *Journal of the Acoustical Society of America*, 99(4):1825–1839.

Allen, J. B. (2005). *Articulation and Intelligibility*. Morgan and Claypool, 3401 Buckskin Trail, LaPorte, CO 80535. ISBN: 1598290088.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4):390–412.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

Bradley, L. and Bryant, P. E. (1978). Difficulties in auditory organisation as a possible cause of reading backwardness. *Nature*, 271(5647):746–747.

Bradley, L. and Bryant, P. E. (1983). Categorizing sounds and learning to read - a causal connection. *Nature*, 301(5899):419–421.

Brandt, J. and Rosen, J. (1980). Auditory phonemic perception in dyslexia: Categorical identification and discrimination of stop consonants. *Brain and Language*, 9(2):324–337. cited By 60.

Bronkhorst, A., Bosman, A., and Smoorenburg, G. (1993). A model for context effects in speech recognition. *Journal of the Acoustical Society of America*, 93(1):499–509.

Cunningham, A. E. (1990). Explicit versus implicit instruction in phonemic awareness. *Journal of experimental child psychology*, 50(3):429–444.

DeGroot, B. J., den Bos, K. P. V., der Meulen, B. F. V., and Minnaert, A. E. (2015). Rapid naming and phonemic awareness in children with reading disabilities and/or specific language impairment: Differentiating processes? *Journal of Speech, Language, and Hearing Research*, 58:1538–1548.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., and Vigorito, J. (1971). Speech perception in infants. *Science*, 171(3968):303–306.

Ellis, W. and Cramer, S. C. (1994). Learning disabilities: A national responsibility. report of the summit on learning disabilities. Report on learning disabilities ERIC Number: ED378783, ERIC, National Center for Learning Disabilities, 381 Park Ave., South, Suite 1420, New York, NY 10016. https://eric.ed.gov/?id=ED378783.

Fletcher, H. (1995). Speech and hearing in communication. In Allen, J. B., editor, *The ASA edition of Speech and Hearing in Communication*, pages A1–A34,1–487. Acoustical Society of America, New York.

Fousek, P., Svojanovsky, P., Grezl, F., and Hermansky, H. (2004). New nonsense syllables database: Analyses and preliminary asr experiments. In *Proceedings of the International Conference on Spoken Language Processing*, pages 2749–2752.

Getz, L. M. and Toscano, J. C. (in press). Electrophysiological evidence for top-down lexical influences on early speech perception. *Psychological Science*.

Goswami, U. and Bryant, P. (1990). *Phonological Skills and Learning to Read*. Hove : Lawrence Erlbaum.

Gough, P. B. and Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7(1):6–10.

Hanford, E. (2018). Hard words: Why aren't kids being taught to read. American Public Media: APM Reports. https://www.apmreports.org/story/2018/09/10/hard-words-why-american-kids-arent-being-taught-to-read.

Harm, M. W., McCandliss, B. D., and Seidenberg, M. S. (2003). Modeling the successes and failures of interventions for disabled readers. *Scientific Studies of Reading*, 7(2):155–182.

Hazan, V. and Adlard, A. (1998). Speech perception in children with specific reading difficulties (dyslexia). *Quarterly Journal of Experimental Psychology: Section A*, 51(1):153 – 177.

Hazan, V., Messaoud-Galusi, S., Rosen, S., Nouwens, S., and Shakespeare, B. (2009). Speech perception abilities of adults with dyslexia: Is there any evidence for a true deficit? *Journal of Speech, Language, and Hearing Research*, 52:1510 – 1529.

Hazan, V., Messaoud-Galusi, S., Rosen, S., Schlauch, R., and Wright, B. (2013). The effect of talker and intonation variability on speech perception in noise in children with dyslexia. *Journal of Speech, Language, and Hearing Research*, 56(1):44 – 62.

Hoover, W. A. and Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, 2(2):127–160.

Jaeger, T. F. (2008). Categorical data analysis: Away from anovas (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4):434–446.

Joanisse, M. F., Manis, F. R., Keating, P., and Seidenberg, M. S. (2000). Language deficits in dyslexic children: Speech perception, phonology and morphology. *Journal of Experimental Child Psychology*, 77:30–60.

Johnson, C. J., Kubalanza, M. C., Allen, J., Scheidiger, C., Biller, M., and Buie, J. (2015). Speech perception and reading disabilities in individual children. In *Seminar Presented at the American Speech-Language-Hearing Association*. Denver, Colorado.

Jusczyk, P. W. (2000). *The Discovery of Spoken Language*. MIT press.

Kamhi, A. G. and Catts, H. W. (2012). *Language and Reading Disabilities*. Pearson.

Li, F. (2009). *Perceptual Cues of Consonant Sounds and Impact of Sensorineural Hearing Loss on Speech Perception*. PhD thesis, University of Illinois at Urbana-Champaign.

Li, F. and Allen, J. B. (2011). Manipulation of consonants in natural speech. *IEEE Transactions on Audio, Speech and Language Processing*, 19(3):496–504.

Li, F., Menon, A., and Allen, J. B. (2010). A psychoacoustic method to find the perceptual cues of stop consonants in natural speech. *Journal of the Acoustical Society of America*, 127(4):2599–2610.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74(6):431–461.

Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54(5):358.

Manis, F. R., McBride-Chang, C., Seidenberg, M. S., Keating, P., Doi, L. M., Munson, B., and Petersen, A. (1997). Are speech perception deficits associated with developmental dyslexia. *Journal of Experimental Child Psychology*, 66:211–235.

Messaoud-Galusi, S., Hazan, V., and Rosen, S. (2011). Investigating speech perception in children with dyslexia: Is there evidence of a consistent deficit in individuals?. *Journal of Speech, Language, and Hearing Research*, 54(6):1682 – 1701.

Miller, G., Heise, G., and Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*, 41(5):329–335.

Miller, G. A. and Nicely, P. E. (1955). An analysis of perceptual confusions among some english consonants. *Journal of the Acoustical Society of America*, 27(2):338–352.

NICHD/NRP (2000a). Report of the national reading panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Technical Report Tech. Rep. No. 00-4769, NIH, Washington DC. http://www.nichd.nih.gov/publications/nrp/report.cfm.

NICHD/NRP (2000b). Report of the national reading panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Report of the subgroups. Technical Report Tech. Rep. No. 00-4754, NIH, Washington DC. http://www.nichd.nih.gov/publications/nrp/findings.cfm.

Phatak, S. and Allen, J. B. (2007). Consonant and vowel confusions in speech-weighted noise. *Journal of the Acoustical Society of America*, 121(4):2312–26.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., and Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological Review*, 103(1):56.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Seidenberg, M. S. and McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96(4):523.

Share, D. L. and Stanovich, K. E. (1995). Cognitive processes in early reading development: Accommodating individual differences into a model of acquisition. *Issues in Education*, 1(1):1–57.

Singh, R. and Allen, J. B. (2012). The influence of stop consonants' perceptual features on the Articulation Index model. *Journal of the Acoustical Society of America*, 131(4):3051–3068.

Tallal, P. (2000). The science of literacy: From the laboratory to the classroom. *Proc. Nat. Acad. Sci.*, 97(6):2402–2406.

Torgesen, J. K. (2004). Learning disabilities: An historical and conceptual overview. In *Chapter 1: Learning about learning disabilities*, pages 3–40. Elsevier.

Toscano, J. and Allen, J. B. (2014). Across and within consonant errors for isolated syllables in noise. *Journal of Speech, Language, and Hearing Research*, 57:2293–2307.

Toscano, J. C. and Lansing, C. R. (in press). Age-related changes in temporal and spectral cue weights in speech. *Language and Speech*.

Toscano, J. C. and McMurray, B. (2012). Cue-integration and context effects in speech: Evidence against speaking-rate normalization. *Attention, Perception, and Psychophysics*, 74(6):1284–1301.

White-Schwoch, T., Woodruff Carr, K., Thompson, E. C., Anderson, S., Nicol, T., Bradlow, A. R., Zecker, S. G., and Kraus, N. (2015). Auditory processing in noise: A preschool biomarker for literacy. *PLOS Biology*, 13(7):1–17.

Williams, C. J. and Jacobs, A. M. (2009). The impact of otitis media on cognitive and educational outcomes. *Medical Journal of Australia*, 191(S9):S69–S72.

Ziegler, J. C., Pech-Georgel, C., George, F., and Lorenzi, C. (2009). Speech-perception-in-noise deficits in dyslexia. *Developmental Science*, 12:732–745.

Table A1

Conversion from Darpabet to International Phonetic Alphabet for LDC unvoiced consonants, voiced consonants, and vowels.

Unvoiced Consonants		**Voiced Consonants**		**Vowels**		
Dbet	IPA	Dbet	IPA	Dbet	IPA	L/T
C	ʧ	D	ð	@	æ	
S	ʃ	b	b	A	ʌ	T
T	θ	d	d	E	ɛ	L
f	f	g	g	I	ɪ	L
h	h	J	ʤ	O	ɔɪ	
k	k	l	l	R	ɝ	
p	p	m	m	U	ʊ	L
s	s	n	n	W	aʊ	
t	t	r	r	Y	aɪ	
H	ʍ	Z	ʒ	c	ɔ	T
		G	ŋ	a	ɑ	
		v	v	e	e / eɪ	T
		w	w	i	i	T
		y	j	o	o / oʊ	T
		z	z	u	u	T
					ə	L

Appendix