AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

© 2023 The Author(s)
ISSN: 0096-3445

2024, Vol. 153, No. 4, 957–981
https://doi.org/10.1037/xge0001522

Can Speech Perception Deficits Cause Phonological Impairments? Evidence From Short-Term Memory for Ambiguous Speech

Harriet J. Smith[1], Rebecca A. Gilbert[1, 2], and Matthew H. Davis[1]
[1] MRC Cognition and Brain Sciences Unit, University of Cambridge
[2] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

Poor performance on phonological tasks is characteristic of neurodevelopmental language disorders (dyslexia and/or developmental language disorder). Perceptual deficit accounts attribute phonological dysfunction to lower-level deficits in speech-sound processing. However, a causal pathway from speech perception to phonological performance has not been established. We assessed this relationship in typical adults by experimentally disrupting speech-sound discrimination in a phonological short-term memory (pSTM) task. We used an automated audio-morphing method (Rogers & Davis, 2017) to create ambiguous intermediate syllables between 16 letter name–letter name ("B"–"P") and letter name–word ("B"–"we") pairs. High- and low-ambiguity syllables were used in a pSTM task in which participants ($N = 36$) recalled six- and eight-letter name sequences. Low-ambiguity sequences were better recalled than high-ambiguity sequences, for letter name–letter name but not letter name–word morphed syllables. A further experiment replicated this ambiguity cost ($N = 26$), but failed to show retroactive or prospective effects for mixed high- and low-ambiguity sequences, in contrast to pSTM findings for speech-in-noise (SiN; Guang et al., 2020; Rabbitt, 1968). These experiments show that ambiguous speech sounds impair pSTM, via a different mechanism to SiN recall. We further show that the effect of ambiguous speech on recall is context-specific, limited, and does not transfer to recall of nonconfusable items. This indicates that speech perception deficits are not a plausible cause of pSTM difficulties in language disorders.

Public Significance Statement
Individuals with dyslexia or developmental language disorder often have problems with short-term memory for speech which some theories attribute to difficulties in hearing speech sounds. By making speech sounds that are intermediate between different syllables we can mimic this challenge for typical adults remembering sequences of letter names. This study found that short-term memory difficulties are unlikely to be caused by a speech perception problem.

Keywords: speech perception, dyslexia, developmental language disorder, verbal short-term memory, language development

Dyslexia and developmental language disorder (DLD) are the most prevalent childhood language disorders, with estimated rates of 7.1% (Yang et al., 2022) and 7.58% (Norbury et al., 2016), respectively, at primary school age. These disorders are also highly comorbid, with approximately 58% of 8-year-olds with dyslexia also meeting criteria for DLD (Snowling, Lervåg, et al., 2019; Snowling, Nash, et al., 2019). Poor performance on phonological tasks, which involve identifying, manipulating, or maintaining speech sounds, is

This article was published Online First December 14, 2023.
Harriet J. Smith https://orcid.org/0000-0003-4314-6571
Some of the data and ideas included in this article were previously presented at meetings of the Experimental Psychology Society (Online Meeting, January 2022; Swansea, July 2023) and at the International Workshop on Reading and Developmental Dyslexia (June 2023). This work was funded by the U.K. Medical Research Council (Grant MC_UU_00030/6 to Matthew H. Davis). The authors thank the jsPsych team and Prolific participants for their time and effort. Full data, code, and audio materials will be available via Open Science Framework (https://osf.io/wc3gn/) at the time of publication. Data and materials (publicly available at time of publication) are available at: https://osf.io/wc3gn/.

Harriet J. Smith served as lead for conceptualization, data curation, formal analysis, investigation, methodology, project administration, validation, visualization, and writing–original draft and contributed equally to funding acquisition and software. Rebecca A. Gilbert served in a supporting role for methodology, resources, software, and writing–review and editing. Matthew H. Davis served as lead for funding acquisition and supervision, contributed equally to conceptualization, and served in a supporting role for data curation, formal analysis, investigation, methodology, project administration, resources, software, validation, visualization, and writing–original draft. Harriet J. Smith and Matthew H. Davis contributed equally to writing–review and editing.

Open Access funding provided by University of Cambridge: This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0; https://creativecommons.org/licenses/by/4.0). This license permits copying and redistributing the work in any medium or format, as well as adapting the material for any purpose, even commercially.

Correspondence concerning this article should be addressed to Harriet J. Smith, MRC Cognition and Brain Sciences Unit, University of Cambridge, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom. Email: harriet.smith@mrc-cbu.cam.ac.uk

common in both disorders. These tasks are traditionally described with respect to a triad of phonological skills (Wagner & Torgesen, 1987): phonological awareness, rapid automatized naming (RAN), and phonological short-term memory (pSTM). An extensive body of research has shown that children with dyslexia and DLD underperform on phonological awareness measures such as rhyme and alliteration detection (e.g., Bradley & Bryant, 1983; Claessen & Leitão, 2012; Fawcett & Nicolson, 1995). Additionally, both groups are slow at RAN (Bowers & Wolf, 1993; De Groot et al., 2015; Denckla & Rudel, 1976; Katz et al., 1992) and have poor pSTM as assessed by span tasks (Ackerman & Dykman, 1993; Cowan et al., 2017) and nonword repetition (Dollaghan & Campbell, 1998; Ehrhorn et al., 2021; Gathercole & Baddeley, 1990; Melby-Lervåg & Lervåg, 2012; Szenkovits & Ramus, 2005). These difficulties persist into adulthood in dyslexia (Elbro et al., 1994; Pennington et al., 1990) and have led to the dominant view that phonological dysfunction should be considered a core deficit in dyslexia (Snowling, 1998; Stanovich, 1986; Vellutino et al., 2004).

The overlap in phenotype across dyslexia and DLD has resulted in substantial discussion around how the relationship between these disorders is best characterized. The severity model (Kamhi & Catts, 1986; Tallal et al., 1997) describes a phonological processing deficit as the underlying cause of both disorders, with differences stemming from the severity of this deficit. A less severe deficit manifests in reading problems with little effect on oral language (dyslexia), whereas a more severe deficit results in problems with both written and oral language (DLD). Alternatively, Bishop and Snowling (2004) propose a two-factor model, with variation along a phonological dimension and a nonphonological language dimension. Differing degrees of dysfunction along these two dimensions characterizes dyslexia (if phonological skills are impaired relative to nonphonological skills), poor comprehension (vice versa), DLD (if both phonological and nonphonological abilities are impaired), or normal language. The two-dimensional model is supported by longitudinal work by Snowling, Lervåg, et al. (2019) and Snowling, Nash, et al. (2019), who found that dyslexic children have early (preschool) deficits in phonology specifically, with increasing nonphonological language difficulties over time, whilst children with DLD have large deficits in both phonological and nonphonological aspects of language from an early age. Ramus et al. (2013) also describe phonological and nonphonological dimensions, but further distinguish between phonological skill and phonological representations, resulting in a three-factor model. Notably, these models all highlight phonological function as a key aspect of both disorders. There is therefore widespread consensus that phonological deficits are a consistent feature of childhood language difficulties, and that impairments in this domain transcend disorder categories. However, this view is not universal; Catts et al. (2005) propose the "fully distinct hypothesis," which includes phonological and nonphonological dimensions, but characterizes DLD as a disorder of nonphonological language abilities. They describe, at most, mild phonological impairments in children with DLD, in contrast to severe impairments in children with dyslexia, or co-occurring DLD and dyslexia (previously reported phonological difficulties in DLD are attributed to high disorder co-occurrence).

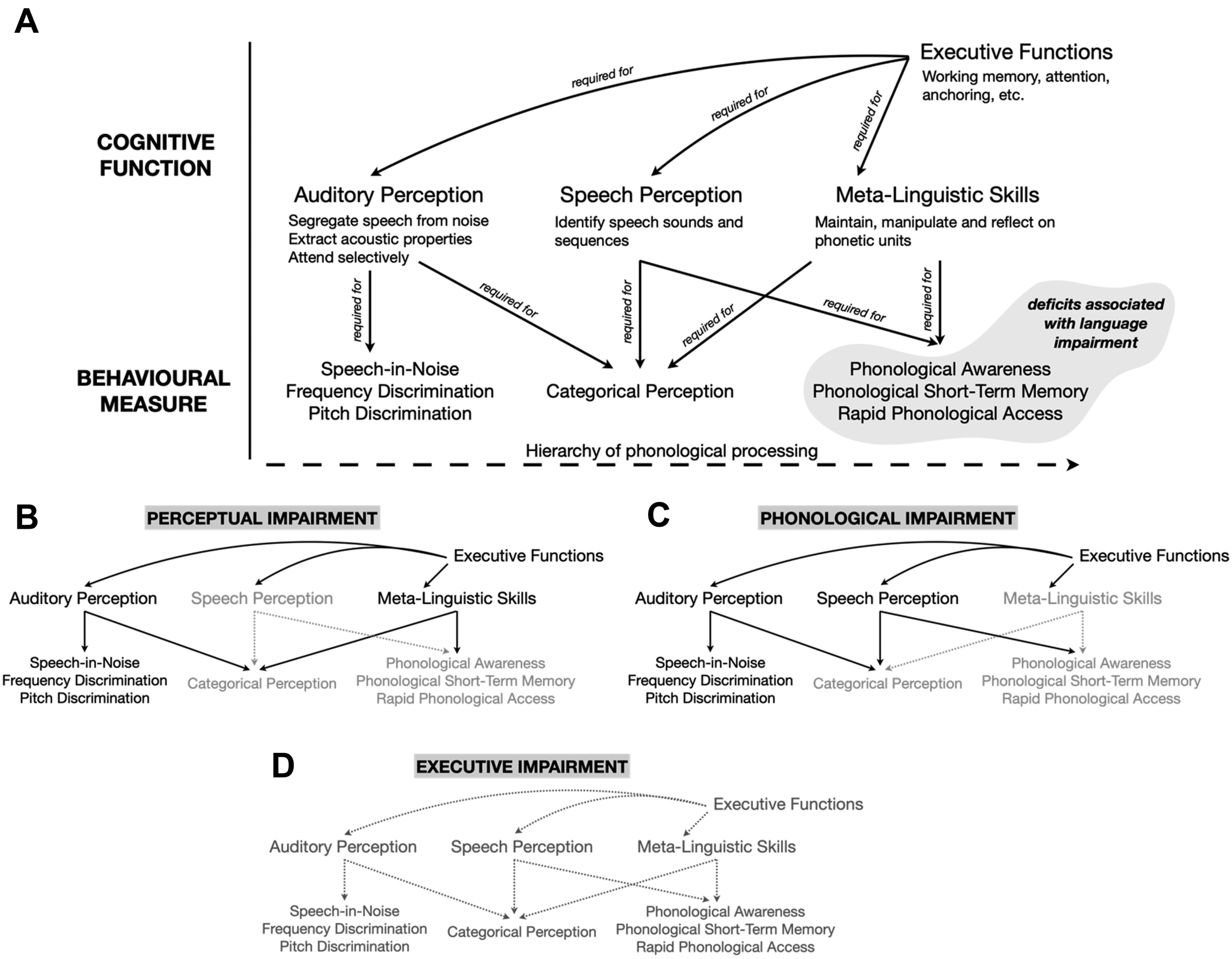
A remaining question is whether these phonological deficits result from dysfunction at the level of speech perception. One prevalent umbrella of theories, here termed "perceptual deficit" accounts, suggest that children who perform poorly on phonological tasks actually have auditory deficits at lower stages of the speech processing hierarchy, which make their discrimination of speech sounds less accurate. This inefficient auditory function throughout language acquisition disrupts acoustic-to-phonological mapping—the learning process by which variable acoustic signals are mapped onto relatively stable internal representations of speech sounds. The result of this dysfunction is that the speech-sound representations children with dyslexia and DLD form are in some way inadequate; either less distinctive, degraded, or otherwise incomplete (Boets, 2014; Elbro et al., 1994; Elbro & Jensen, 2005; Swan & Goswami, 1997). Although the exact nature of the perceptual deficit varies between accounts, the key prediction of these theories is that children who perform poorly on higher-level assessments of phonological skill (i.e., phonological awareness, RAN, and pSTM tasks) also have impairments in auditory perception. Critically, rather than perceptual impairments simply being associated with phonological problems, perceptual dysfunction is described as the underlying cause of these difficulties.

There is substantial evidence to suggest that speech-processing difficulties are associated with childhood language disorders. In categorical perception studies, children with dyslexia are often less consistent at matching unambiguous stimuli to appropriate phonemic categories, and show shallower labeling functions (Godfrey et al., 1981; Joanisse et al., 2000; Manis et al., 1997; Messaoud-Galusi et al., 2007; Noordenbos & Serniclaes, 2015). Furthermore, discrimination of speech sounds at 5½ years predicts phoneme awareness and reading abilities at 6½ years (Snowling, Lervåg, et al., 2019; Snowling, Nash, et al., 2019). Children with DLD also show impairments in identifying speech-in-noise (SiN) (Robertson et al., 2009), and performance accounts for unique variance in oral language ability beyond IQ and low-level auditory perception (Ziegler et al., 2005). Furthermore, inputting distorted speech-sound inputs to connectionist models of sentence processing can simulate DLD-type syntactic deficits (Joanisse & Seidenberg, 2003), which indicates that there is a plausible causal sequence from degraded perceptual input to broader language problems, in artificial neural networks.

However, even if reliable deficits in speech perception can be established in children with phonological impairments, this is not sufficient evidence for a causal link from speech perception to phonological skill. As shown in Figure 1, at least three models can explain this association. The first (panel B) is a causal model in which perceptual dysfunction during language acquisition results in inadequate phonological representations, in accordance with perceptual deficit accounts. In the second model (panel C) the relationship is reversed, and perceptual deficits in typical speech perception tasks such as categorical perception are a consequence of higher-level phonological problems. According to this framework, children perform poorly on perceptual tasks because the lack of specificity in their phonological representations makes discrimination of speech sounds more difficult. Finally, auditory and phonological problems may both be caused by additional cognitive factors (panel D), but not be causally related to each other (a correlational model).

Potential candidates for this additional factor are a procedural learning deficit (Nicolson & Fawcett, 2007, 2011; Ullman & Pierpont, 2005), impaired speech-motor function (Catts, 1989; Heilman et al., 1996), or anchoring difficulties (Ahissar, 2007). These alternative models argue that poor performance on speech perception and phonological tasks should be attributed to executive function issues, and hence perceptual deficits are not causally related to phonological impairments.

Figure 1
Alternative Models Showing Three Possible Relationships Between Speech Perception and Phonological Task Performance



Note. Panel A outlines how cognitive functions within and outside the speech-sound processing hierarchy contribute to performance on phonological measures. Panel B shows the causal pathway associated with phonological deficits according to perceptual impairment accounts. Panel C highlights an alternative account, in which metalinguistic dysfunction results in poor performance on phonological and speech perception tasks. Panel D shows an executive impairment account, where domain-general difficulties account for poor phonological task performance.

Thus far, the nature of the proposed causal pathway from impairments in speech perception to poor performance on higher-level phonological tasks has not been adequately established. As previously described, co-occurrence of these difficulties is not sufficient to infer causality. In order to verify the perceptual deficit framework, at least two assertions must be demonstrated; firstly, temporal precedence (Moreno & Martínez, 2008)—perceptual problems must occur earlier in development than phonological issues. This question requires longitudinal investigations, which are not the approach taken in the present work. One such study by Snowling, Lervåg, et al. (2019) and Snowling, Nash, et al. (2019) showed that poor categorical speech perception at 5½ years predicted poorer reading at 6½ years, in line with a causal association. However, they also found that speech perception did not mediate the relationship between phoneme awareness and reading. This suggests that poor performance on categorical perception tasks in reading-impaired children may be best attributed to extraneous factors (e.g., decision-making or attentional issues), rather than a model in which perceptual difficulties affect reading via poor phonological awareness. These findings are therefore more consistent with the model described in panel D of Figure 1.

A second approach that we adopt here is to assess whether a targeted disruption at the level of speech perception has a knock-on effect on phonological task performance. We specifically focus on pSTM as our phonological measure, as pSTM is very consistently poor in both dyslexia and DLD. Indeed, measures of pSTM (sentence and nonword repetition) are the most consistent clinical markers for DLD (Bishop et al., 1996; Conti-Ramsden et al., 2001). In laboratory experiments we will disrupt the discrimination of speech sounds by varying the degree of perceptual ambiguity in speech.

Perceptual ambiguity is here manipulated through the use of audio-morphed speech, which is created by averaging acoustic features of speech sounds from opposite ends of natural phonetic continua (e.g., speech tokens that are acoustically intermediate between the syllables "bee" /biː/ and "pea" /piː/). Resolving the identity of speech sounds is made more challenging if they are ambiguous, as the phonemes are less distinctive. Increased perceptual ambiguity results in slower and less accurate perceptual categorization, and shallower labeling functions; these are the same dysfunction previously described in language-impaired children. Processing ambiguous stimuli thus introduces a perceptual challenge for typically functioning adults that mimics the hypothesized dysfunction in language impairments and we can use this manipulation as a laboratory model to assess the validity of perceptual deficit models. The ambiguous speech sounds will form the stimuli in a span task, a classic measure of pSTM performance. By manipulating the ambiguity of the to-be-recalled speech, we will observe if ease of perceptual discrimination has a corresponding effect on pSTM recall. If speech perception problems cause poor performance on phonological tasks, we would expect that more perceptually challenging speech would result in poorer pSTM performance.

There is already substantial evidence to indicate that making perceptual discrimination more difficult can cause a reduction in pSTM. When participants recall lists of sequential items, lists containing words that rhyme and words with co-occurring phonemes result in poorer performance (Baddeley, 1968; Laughery & Pinkus, 1966), termed the "phonological similarity effect." Recall is also poorer for SiN compared with clear speech, even when noise does not impair identification accuracy (Rabbitt, 1968; Surprenant, 1999). This indicates that even if speech sound identity can be resolved accurately, noise can impact the maintenance of verbal information in STM. Recall for synthetic speech is also poorer than for natural speech (Luce et al., 1983; Smither, 1993). Overall, these findings suggest that pSTM retention is more challenging for more perceptually ambiguous speech. However, alternative explanations remain, as understanding degraded speech requires additional cognitive processes to be recruited, including performance monitoring via attention (Peelle, 2018).

Previously, these effects have been explained by increased effort in encoding and/or rehearsal of ambiguous speech. Resolving the phonological identity of ambiguous speech is associated with processing costs, as indicated by slower identification of audio-morphed speech compared to clear speech (Rogers & Davis, 2017). Heald and Nusbaum (2014) attributed the impact of synthetic speech on STM to a one-to-many mapping of synthetic acoustic cues to phonemes at the encoding phase, and showed that learning mappings between synthetic speech segments and particular phonemes reduced the memory impairment. However, Rabbitt (1968) showed that when the second half of a span list was presented in noise, memory was impaired for items in the first half of the list, even when these items were presented without noise. These results indicate that difficult-to-perceive speech also disrupts the rehearsal-based maintenance of previously presented items.

However, the methods used to induce ambiguity in previous studies do not challenge the perceptual system in a way that closely resembles the proposed perceptual dysfunction in phonologically impaired individuals. Furthermore, SiN and synthetic speech challenge the perceptual system in a nonspecific manner, which plausibly impacts perception of all speech sounds equally. In contrast, perceptual difficulties due to poorly specified speech-sound representations are likely to be specific to phonological neighbors (i.e., highly similar sounds). For example, a classic letter span task may include phonological neighbors which are easily confusable, such as the letter for "B" /biː/ and "D" /diː/. However, the set of items used for a digit span task does not include equivalently confusable items, and therefore a perceptual deficit might not cause misidentification of digit sounds. Yet, individuals with phonological impairment typically show reduced pSTM in letter and digit span tasks alike (see Ackerman & Dykman, 1993; Smith-Spark & Fisk, 2007 for direct comparisons).

Impairments on digit span tasks may therefore be more easily reconciled with an executive dysfunction account than a core perceptual deficit. One explanation is that language-impaired individuals may be less able to adequately tune to task-relevant information (a form of executive dysfunction similar to the anchoring deficit; Ahissar, 2007). This may lead to interference from irrelevant items in the lexicon (e.g., the phonological representation of "tree" could plausibly interfere with "three" in a digit span context). We do not consider perceptual and executive dysfunction to be (necessarily) mutually exclusive; poorer perception likely increases "listening effort" (Peelle, 2018; Pichora-Fuller et al., 2016), requiring greater recruitment of cognitive resources to achieve speech comprehension. Therefore an individual with poorer perceptual abilities may be especially susceptible to higher-level phonological dysfunction, if they also cannot engage cognitive processes effectively to compensate for increased listening effort. However, in this framework the causal factor leading to poor pSTM includes executive dysfunction (as in panel D of Figure 1), rather than a core perceptual deficit. Relatively weak speech-sound discrimination abilities may act a risk factor for phonological problems, but in isolation be insufficient to cause pSTM deficits. This would constitute a multiple-deficit explanation (Pennington, 2006), whereby difficulties at several levels of the phonological processing hierarchy contribute to the likelihood of poor performance on speech perception and phonological tasks.

To compare the perceptual-only and perceptual-plus-executive explanations we tested two forms of perceptual ambiguity in our letter span task: ambiguity between two letter names (e.g., "B" and "P," hereafter letter–letter) and ambiguity between a letter name and a monosyllabic spoken word (e.g., "B" and "we," "letter–word"). Importantly, the perceptual challenge induced by both forms of ambiguity was equivalent (see Stimulus Pretesting section below). However, in the case of letter–letter ambiguity, competition between response options is high, as both letter names would be appropriate responses in the letter span task. Conversely, for the letter–word items, the context of the letter span task likely biases participants against giving a response that is a spoken word, rather than a letter name.

The use of top-down contextual information to resolve perceptual ambiguity is well-documented at multiple levels of speech processing. Ambiguous or acoustically degraded phonemes can be disambiguated by lexical context (e.g., Connine & Clifton, 1987; Ernestus et al., 2002; Gwilliams et al., 2018; Rogers & Davis, 2017), semantic context (e.g., Obleser & Kotz, 2011), and talker familiarity (e.g., Eisner & McQueen, 2005), among other factors (for a review, see Mattys et al., 2012). Beyond the linguistic level, perception is affected by overarching task context. Variation in attentional demands (e.g., single task or dual task: Casini et al., 2009; Saltzman et al., 2021), stimulus presentation (e.g., interstimulus interval: Coady et al., 2005) and the variability of the stimulus set (e.g., open-set or closed-set: Clopper et al., 2006; Yu & Schlauch, 2019) and can all impact interpretation of ambiguous speech. Furthermore, interactive models such as TRACE (McClelland & Elman, 1986) provide a computational framework in which context constrains perceptual interpretation of ambiguous speech via bidirectional connections between different processing levels (see McClelland et al., 2006, for a review). It is therefore highly plausible that the top-down influence of task set may make letter–word ambiguity easier to resolve than letter–letter ambiguity.

In summary, these two forms of ambiguity are equally perceptually challenging, but differ in their executive demands in the context of a letter span task. If a perceptual deficit alone is sufficient to impair pSTM, we would predict that recall would be similarly affected by both forms of perceptual ambiguity. However, if the impact of poor perception on pSTM is mediated by executive processes, we would expect letter–letter ambiguity to have a greater impact on recall than letter–word ambiguity. Testing these two forms of perceptual ambiguity therefore allows us to distinguish between a "pure" perceptual deficit explanation, and an account in which perceptual and executive difficulties combine to impair pSTM (a "double-deficit" framework).

Overview of the Present Studies

In the present study we therefore assessed pSTM performance in typical adults for speech sounds with varying perceptual ambiguity. We used audio-morphing of natural speech tokens to create highly naturalistic speech sounds with fine-grained, tightly controlled distinctions between stimuli. We propose that audio-morphed speech represents the most direct simulation of the hypothesized perceptual dysfunction in dyslexia/DLD, if speech-sound representations are degraded or poorly specified. By examining the mechanisms by which ambiguous speech affects pSTM, we can establish whether there is sufficient evidence for a causal pathway from dysfunctional speech perception to poor performance on a pSTM task. In other words, we will simulate a perceptual deficit, and test the consequences on phonological performance.

Our core hypothesis is that perceptually ambiguous speech will be more difficult to maintain in pSTM, showing that a disruption at the level of speech perception can impair performance on a classic phonological task. To assess this, we administered an auditory letter span task to adults with typical language abilities, and assessed recall of more ambiguous letter names compared with less ambiguous letter names. We included two forms of ambiguity: (a) ambiguity between a letter name and another letter name, and (b) ambiguity between a letter name and a word. This manipulation allowed us to evaluate whether a perceptual challenge can impair recall independently from executive demands. In our second experiment, we assessed if perceptual ambiguity can affect the rehearsal-based maintenance of unambiguous information in pSTM. If observed, this would further show that perceptual ambiguity—like SiN—can lead to more widespread disruption of nonperceptual processes that contribute to pSTM performance.

Method

Stimulus Generation

A letter span task was designed to test pSTM. In order to specifically test for cognitive effects of ambiguity, the stimulus set needed to consist of items in which ambiguity varied, but other linguistic dimensions were controlled for. Spoken letter names were chosen as stimuli, as these items are uniformly monosyllabic (unlike digits), lack the semantic associations of words and are equally and highly familiar to all native English speakers.

We selected 10 letter names which could be transformed with only a minimal phonetic change (a single change in voicing, manner or place-of-articulation; see Ladefoged, 1975) to produce both a second letter name, and a monosyllabic spoken word that was not a letter name. For example, the letter name "B" (produced as "/biː/") was paired with the letter name "D" ("/diː/") and the monosyllabic word "we" ("/wiː/"). In both cases, a change of a single articulatory feature is sufficient to change the identity of the letter name or spoken word (for "B"–"D" a change in place of articulation, from labial to coronal, and for "B"–"we" a change in manner of articulation from stop to approximant). We identified 10 letter name to letter name minimal pairs and 10 letter name to word minimal pairs, hereafter letter–letter and letter–word pairs as listed in Table 1.

Each letter or word in these 20 pairs was spoken by a female native speaker of Southern British English, speaking in time to a 1 Hz metronome so as to ensure that all speech tokens had their p-center aligned to the beat and hence could be used to generate rhythmic letter name sequences (see Morton & Chambers, 1976). Stimuli were recorded direct to computer at a sampling rate of 44.1 kHz. Audio preprocessing (noise reduction and temporal alignment) was carried out in Adobe Audition.

Table 1
Letter–Letter and Letter–Word Pairs With Minimal Phonetic Difference

Letter 1 | phonetic transcription		Letter 2 | phonetic transcription		Phonetic difference Letter 1–Letter 2	Word | phonetic transcription		Phonetic difference Letter 1–word
D	/diː/	T	/tiː/	Voicing	Knee	/niː/	Manner
B	/biː/	D	/diː/	Place	We	/wiː/	Manner
C	/siː/	Z	/ziː/	Voicing	She	/ˈʃiː/	Place
F	/ɛf/	S	/ɛs/	Place	Ev	/ɛv/	Voicing
M	/ɛm/	N	/ɛn/	Place	Ebb	/ɛb/	Manner
N	/ɛn/	M	/ɛm/	Place	Ed	/ɛd/	Manner
P	/piː/	B	/biː/	Voicing	Key	/kiː/	Place
T	/tiː/	P	/piː/	Place	Key	/kiː/	Place
V	/viː/	Z	/ziː/	Place	Fee	/fiː/	Voicing
Z	/ziː/	V	/viː/	Place	Thee	/ðiː/	Place

The STRAIGHT channel vocoder (Kawahara et al., 1999; Kawahara & Morise, 2011) was used for time-aligned averaging (morphing) of periodic, aperiodic, and F0 representations between each pair of speech sounds. As described in Rogers and Davis (2017), we used custom MATLAB scripts (The MathWorks, Inc., Natick, Massachusetts), which used dynamic time-warping (code supplied by Dan Ellis; https://www.ee.columbia.edu/~dpwe/resources/matlab/dtw/) to place temporal anchor points at 50 ms intervals in one syllable sound and mark maximally similar positions in the second sound. We could then use time-aligned averaging functions within the STRAIGHT morphing code to automatically average equivalent marked sections of the two sounds and create natural sounding phonetic continua between the two syllables.

This audio-morphing code also incorporated a form of automated p-centre adjustment to maintain the rhythmic properties of speech sequences, building on methods initially described by Scott (1993). p-Centre adjustment builds on the observation due to Morton and Chambers (1976) that the perceived rhythmicity of syllable sequences depends not on the timing of the acoustic onset or offset of speech sounds, but rather on the timing of "perceptual centers" (p-centers). That is, the subjective moment of occurrence of a spoken syllable is most closely related to the acoustic onset of the vowel and can be defined based on the time of the maximum rise in acoustic energy within a particular frequency band (based on an algorithm described by Cummins & Port, 1998). We incorporated this within the audio-morphing procedure to ensure that despite p-centers varying between different letter name sounds, and thus being altered by STRAIGHT's time-aligned averaging, we could still create equivalent, rhythmic speech sound sequences irrespective of the degree of perceptual ambiguity created.

For each of the 20 phonetic continua (10 letter–letter, 10 letter–word) we generated 10 acoustically intermediate speech tokens, which can be described in terms of the relative proportions of the two spoken syllables that are synthesized. Intermediate tokens were generated at 10% acoustic steps from 5% to 95% of the second sound. For example, a sound from the "B"–"D" continuum created by averaging 5% of the "B" sound with 95% of the "D" sound would be highly perceptually similar to "D." Conversely, the 95% "B"/5% "D" sound would be heard as "B." Previous work (Rogers & Davis, 2017) suggests that sounds from the middle of these continua are more perceptually ambiguous, and require more time to identify, than sounds from toward the ends of the continua. On average, perceptual boundaries for these phonetic continua are typically at the 50% morph points, but can vary as a function of acoustic, phonetic, and lexical variables. To assess the impact of these factors we therefore conducted a perceptual pretest using other individuals from the same participant population as the main study.

Stimulus Pretesting

Participants

Data were collected from 28 participants; this sample size was derived from a power analysis for a *z*-test of proportions with .8 power and binomial hypotheses of null = .5 and alt = .75, conducted in MorePower 6.0 (Campbell & Thompson, 2012). Participants were recruited from Prolific (https://www.prolific.co). All were prescreened to ensure that they self-reported as native English-speaking adults aged 18–45 years, who had grown up in the United Kingdom and had no reported history of dyslexia, speech-and-language therapy or hearing problems. In answer to the question, "How would you describe your gender?" (free response box), 16 participants responded "male" or "man" and 12 participants responded "female" or "woman" (no other responses were given). No further demographic information was collected.

Procedure

In order to identify sounds with behaviorally determined degree of perceptual ambiguity, we pretested the intermediate tokens from our two phonetic continua in a two-alternative forced choice identification task. This identification task, like the subsequent pSTM experiments was developed in JavaScript, utilizing the JsPsych plugin library (De Leeuw, 2015), and hosted on an in-house JATOS server (Lange et al., 2015) based at the CBU in Cambridge. Participants heard each of the 200 novel tokens (10 tokens from each of 20 phonetic continua) in isolation, in a randomized order. Five hundred milliseconds after sound offset, written representations of the two sounds that had been combined to create that token were presented as response options. For example, a participant might hear a token from the "B"–"D" continuum and see the capital letters "B" and "D" on screen. Participants made key presses to indicate which option best matched the sound they had heard. A third option was also available, if participants felt that the speech token that they heard did not match either of the labeled end-point. Each participant produced two judgements for each token, resulting in 400 trials. As repeated exposure to speech sounds in categorical perception tasks may induce perceptual learning and alter individuals' category boundaries, separate groups of participants were used for the identification and short-term memory tasks.

Results

Responses faster than 300 ms (0.01% of the data) and trials on which neither of the two response alternatives was heard (0.2% of trials), or no response was given (0.02% of trials), were excluded. For every participant, the proportion of "sound one" responses for the 95%-sound one morphs was at least .9 ($M = .98$, $SD = .028$), and the proportion of "sound two" responses for the 95%-sound two morphs was at least .9 ($M = .97$, $SD = .026$), indicating a low lapse rate.

The proportion of responses matching the two end points of the continua were averaged over participants for each continua independently, and logistic regression models were fit. Responses followed the expected sigmoidal profile, indicating that participants responded to tokens at the end-points of the continua by consistently selecting the most similar option, with relatively steep changes in responding toward the middle of the continua.

From these functions the point of subjective equality (point at which the predicted probability of a "sound one" response was .5) and slope values were calculated. Fit was poor for the "Z"–"thee" continuum (McFadden's $R^2 = .31$). Fit for the other continua varied from moderate to good (McFadden's R^2 range = .52 to .8).

Continua that varied from the expected psychometric function—a relatively steep slope toward the middle of the continua, with a point of subjective equality around 50%—were excluded. This was quantified as: β1 (log slope) values that were > 2 *SD*s below the average slope, or less than 90% expected responses for the most extreme continuum values (e.g., less than 90% "sound one" responses for the

95%-sound one token). These criteria resulted in the exclusion of two letter–word continua ("Z"–"thee" and "F"–"ev"). Due to the span task design, the letter–letter pairs that included the same letters ("Z"–"C" and "F"–"S") were also excluded.

A low-ambiguity and a high-ambiguity token were selected from each of the 16 remaining continua. Our intention was to select two speech tokens both of which were consistently labeled as the same sound (to ensure that errors in the span task could not be attributed to misidentification), but varied in their degree of ambiguity. For each continuum, we identified the relative proportions of sound one and sound two that gave a 99% predicted probability of a "sound one" response (based on the best-fitting logistic model), and used these proportions to generate our low-ambiguity token. The high-ambiguity token was defined as having the relative proportions of sound one and sound two that led to 90% predicted probability of a "sound one" response. Two example logistic functions for "B"–"D" and "B"–"We" phonetic continua, with high and low-ambiguity tokens marked, are shown in Figure 2 (Panels A and B). Despite this small change in response probabilities this degree of perceptual ambiguity led to a noticeable slowing to categorization response time (panel C).

This resulted in a set of four novel tokens for each of the eight letter names; high- and low-ambiguity morphs created from each of the letter–letter and letter–word phonetic continua.

Experiment 1: Method

Participants

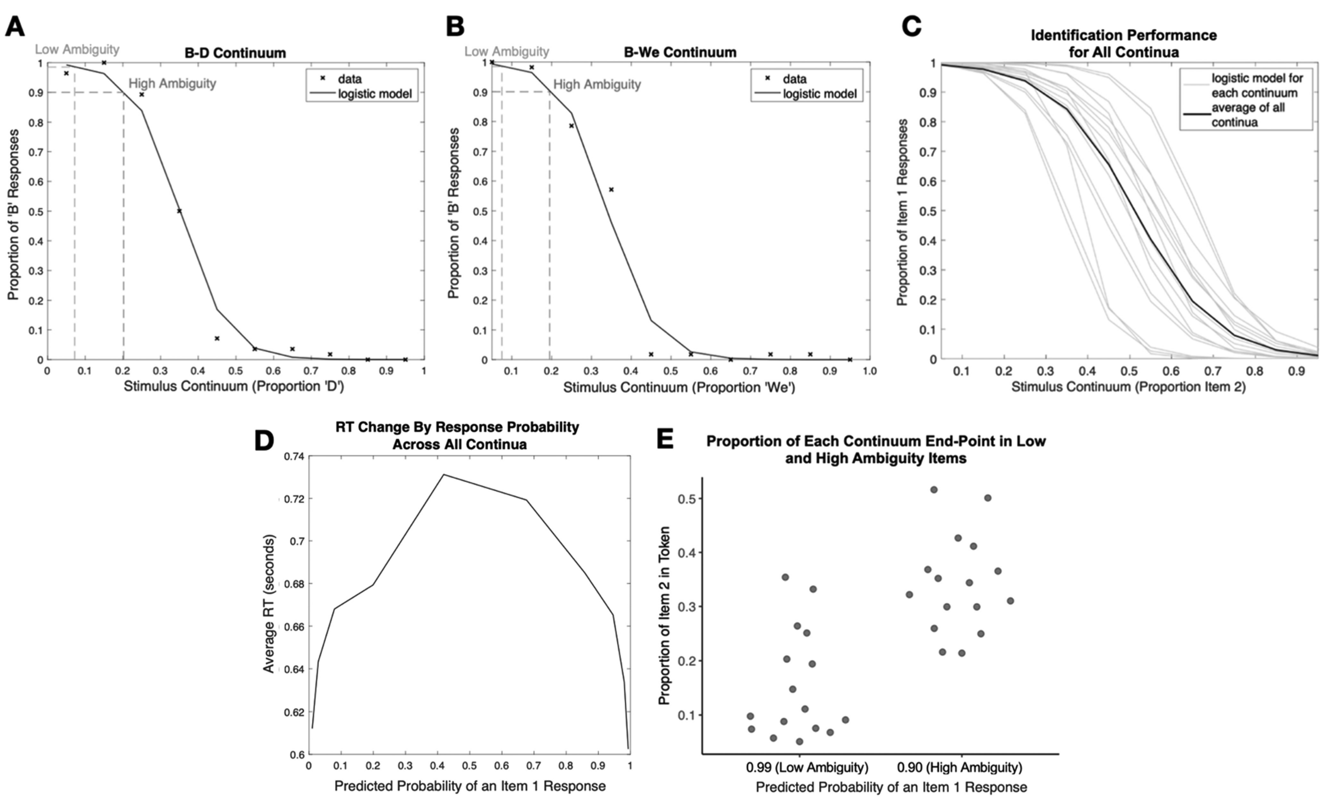
Thirty-six participants were recruited from Prolific. This sample size was determined for a 2 × 2 (Morph Type × Degree of Ambiguity) repeated measures design with .8 power and Cohen's $f = .25$, based on prior SiN literature (Renz et al., 2018). Power calculation was conducted in G*Power 3.1 (Faul et al., 2007). All participants self-reported that they were native English-speaking adults aged between 18 and 45 years, who had grown up in the United Kingdom and had no reported history of dyslexia, speech-and-language therapy or hearing problems. In response to the question, "How would you describe your gender?" (free response box), 15 participants answered "male" or "man," 20 participants answered "female" or "woman," and one participant answered "transgender man." No further demographic information was collected. None of the participants in this experiment had taken part in the stimulus pretesting.

Procedure

The letter span task was also developed in JavaScript, utilizing the JsPsych library, and hosted on the JATOS server. Stimulus lists were constructed in accordance with three contrasts, each with two levels: length (six/eight letters) ambiguity (high/low), and continuum type

Figure 2

Performance on Identification Task (Stimulus Pre-Testing) and Stimulus Selection



Note. Panels A and B show two example continua, with identification responses for the "B"–"D" and "B"–"We" pairs and fitted psychometric functions. Responses are averaged over all participants. For each pair, the point on the response curve at which the predicted probability of a "B" response was .9 was chosen as the high-ambiguity stimulus for Experiments 1 and 2, and the point at which the predicted probability of a "B" response was .99 was chosen as the low-ambiguity stimulus. Panel C shows the identification responses for all 16 final continua. Panel D shows average RT for predicted identification responses. Panel E shows the proportion of each continuum end-point in the final 16 high-ambiguity and low-ambiguity stimuli. RT = response time.

(letter–letter/letter–word). Eighty unique letter combinations were created by pseudorandomly ordering the eight letter name sounds using MATLAB randomization functions (for 40 eight-item lists) or pseudorandomly ordering six letter name sounds from the set (for 40 six-item lists). Each of the eight letter names appeared an equal number of times in the six-item lists, and each letter name appeared an equal number of times at each list position, for both eight- and six-item lists. Lists were checked to exclude U.K. and U.S. English acronyms, using common acronym dictionaries from Laszlo and Federmeier (2007) and Izura and Playfoot (2012), as well as an additional self-devised set to include more modern acronym usage.

High-ambiguity and low-ambiguity versions of each of the 80 lists were created, consisting either of entirely high-ambiguity letter name sounds, or entirely low-ambiguity letter name sounds. This ensured that the high- and low-ambiguity conditions contained lists with the same combination of letter names, in the same order. Fifty percentage of the high-ambiguity list stimuli were drawn from letter–letter continua, and 50% from letter–word continua, and the same for the low-ambiguity lists. This resulted in 20 lists for each combination of length, ambiguity, and continuum type.

All participants ($N = 36$) heard each of the 160 unique lists once, presented at a rate of 1 Hz (one item per second). The order of list presentation was randomized using a JsPsych randomization function and was unique to every subject. One second after the offset of the final list item, a response box was shown and participants were instructed to type the list of letter names they had just heard, in the same order as they were presented.

Analysis Strategy

Two measures of recall performance were calculated for each of the eight conditions, for each participant: (a) the proportion of entire sequences recalled accurately (i.e., responses that consisted of the correct letters reported in the correct order), which we will refer to as "recall accuracy" with high values indicating good performance; (b) the proportion of letters recalled correctly for each sequence, quantified using the Damerau–Levenshtein (D–L) distance.

The D–L distance (Damerau, 1964; Levenshtein, 1966) is a string metric which quantifies the edit distance between two sequences. It is described as the minimum number of transformations required to change one sequence into another. Transformations may consist of insertions, deletions, or substitutions of one item, or a transposition of two adjacent items. The D–L distance provides a more sensitive measure of the similarity of a participant's response to the target sequence than accuracy alone and is particularly useful for the supraspan eight-letter sequences. D–L distance was calculated for each response sequence, using the editDistance function in MATLAB (Text Analytics Toolbox). For clarity, a higher D–L distance indicates a greater difference between the target and response sequences, and therefore poorer recall. To aid interpretability, this metric was transformed into the proportion of letters recalled correctly per sequence (by subtracting the D–L distance from the length of the target sequence and dividing by the length of the target sequence, for each sequence independently).

To assess the hypothesis that recall would be poorer for high-ambiguity sequences compared with low-ambiguity sequences, we used a series of linear mixed effects models (LMMs). For our two recall metrics (proportion of sequences recalled and proportion of letters recalled) independently, we applied LMMs with ambiguity (high vs. low ambiguity), list length (six vs. eight), and continuum type (letter–letter vs. letter–word) as fixed effects, and participant as a random effect. LMMs were used to compare conditions because this method can account for by-participant variance in within-subject designs. These analyses allowed us to evaluate whether recall was affected by ambiguity, and whether this hypothesized "ambiguity cost" varied depending on the length of the sequence (within-span or supraspan) and/or the task relevance of the ambiguity. When significant interaction effects were identified, least-squares means (LS means) analysis with Tukey adjustment was used to compare recall performance for individual conditions.

All analyses were conducted in R (R Version 4.1.3; R Core Team, 2022), using the lme4 package (Bates et al., 2014) to calculate LMMs, and the lsmeans package (Lenth, 2016) for LS means analyses.

Experiment 1: Results

Proportion of Sequences Recalled Correctly

An LMM with list length (six vs. eight), ambiguity (high vs. low ambiguity), and continuum type (letter–letter vs. letter–word) as fixed effects, and participant as a random effect, was used to compare the proportion of sequences recalled accurately in each condition (Figure 3, panels A, B, and C). This revealed a three-way interaction of list length, ambiguity, and continuum type, $F(1, 259) = 5.55$, $p = .019$, $\eta_p^2 = .02$. Subsequent LS means analysis with Tukey adjustment showed that the proportion of sequences recalled correctly was consistently higher for the six-letter sequences than the eight-letter sequences ($p < .001$ for all comparisons of a six-letter with an eight-letter condition), as expected.

For the six-letter sequences, recall of the low-ambiguity lists was better than for the high-ambiguity lists, but only when the sounds in both sequences were taken from letter–letter continua, $t(259) = 5.26$, $p < .001$, $d = 0.61$. This ambiguity effect was absent for the sequences made up of sounds from letter–word continua, $t(259) = 0.25$, $p = .1$.

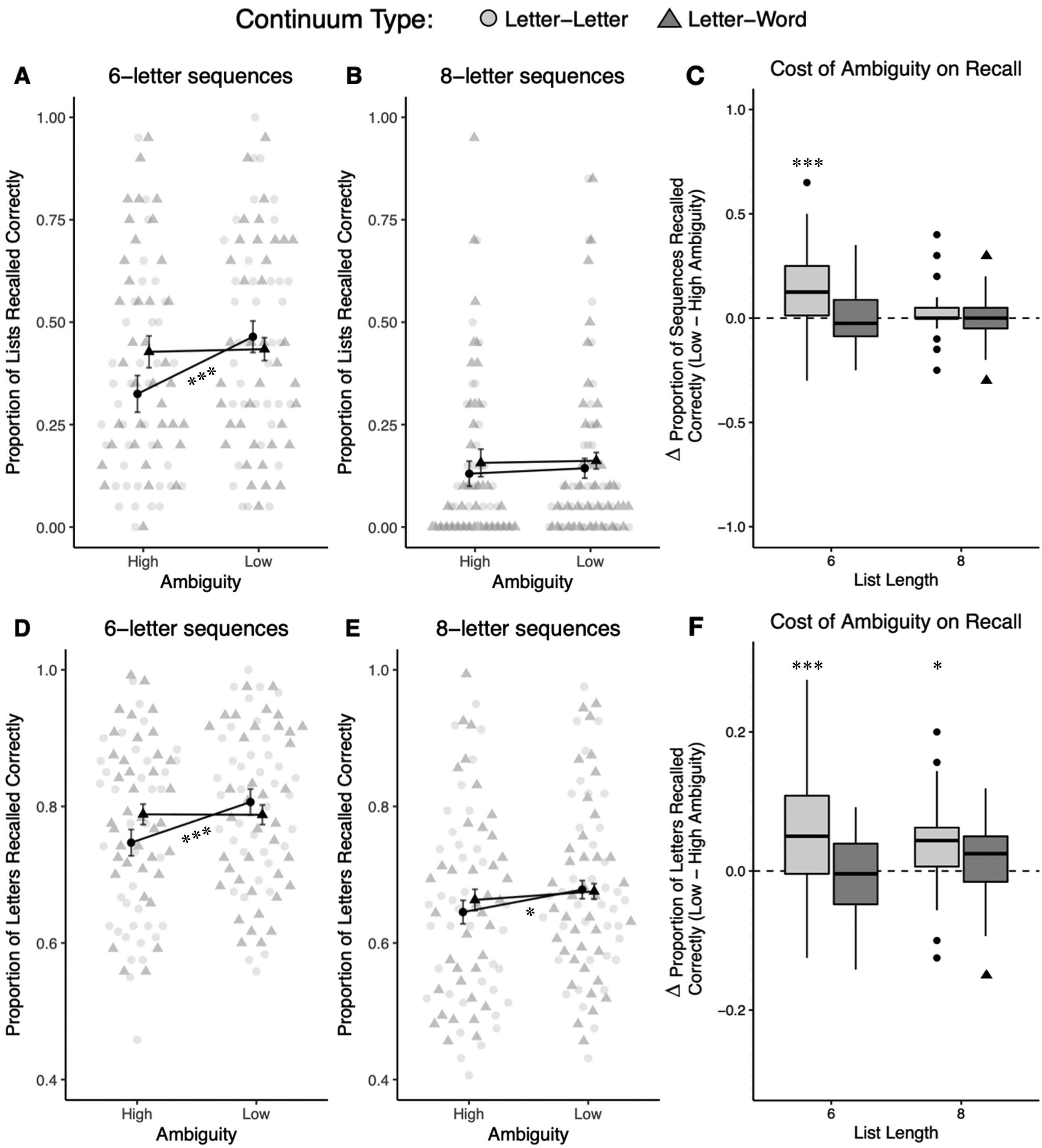
Equivalent comparisons between eight-letter conditions showed no statistically significant differences in recall accuracy. This is likely because these data were heavily skewed toward very low rates of correct sequence recall (mean proportion of sequences recalled correctly, across all conditions = .15).

Proportion of Letters Recalled Correctly

A second LMM with list length (six vs. eight), ambiguity (high vs. low ambiguity), and continuum type (letter–letter vs. letter–word) as fixed effects, and participant as a random effect, was used to compare the proportion of letters recalled correctly in each sequence, between conditions. This data are shown in Figure 3 (panels D, E, and F). This analysis identified a main effect of list length, $F(1, 259) = 838.02$, $p < .001$, $\eta_p^2 = .76$, such that recall was better for six-letter sequences than eight-letter sequences. This analysis also highlighted a two-way interaction of ambiguity and continuum type, $F(1, 259) = 7.65$, $p = .006$, $\eta_p^2 = .03$. Subsequent LS means analysis with Tukey adjustment showed that this interaction was due to better recall of low-ambiguity letter–letter sequences than high-ambiguity letter–letter sequences, $t(259) = -5.69$, $p < .001$, $d = -0.38$. As with the recall accuracy metric, there was no

Figure 3

Recall Performance for the High-Ambiguity and Low-Ambiguity Sequences, in Experiment 1



Note. In Panels A and B, the semitransparent points show the proportion of sequences recalled perfectly in each condition, out of a maximum of 20, for each participant. In Panels D and E, the semitransparent points show the average proportion of letters recalled correctly from sequences in each condition, for each participant. The opaque points represent group averages. The error bars show 95% confidence intervals, calculated according to the Cousineau–Morey method for within-subjects designs (Morey, 2008). Panels C and F show the differences in recall performance between the high-ambiguity conditions and low-ambiguity conditions, termed the ambiguity cost. For interpretation of panels C and F, the line within the box depicts the median, the lower and upper hinges correspond to the first and third quartiles, the upper/lower whiskers extend to the highest/lowest values within 1.5× of the interquartile range, and outliers are plotted individually.
* $p < .05$. ** $p < .01$. *** $p < .001$.

difference in the proportion of letters recalled correctly between high- and low-ambiguity letter–word sequences, $t(259) = -0.62$, $p = .89$, $d = -0.04$. In contrast to the recall accuracy measure, the ambiguity cost (poorer recall of high-ambiguity sequences) was present for both six-letter and eight-letter sequences, when examining the proportion of letters recalled correctly—L–S means with Tukey adjustment: $t(259) = -4.82$, $p < .001$, $d = -0.45$, for six-letter sequences; $t(259) = -2.34$, $p = .03$, $d = -0.30$, for eight-letter sequences.

Frequency of Misidentification

The extent to which participants performed poorly in the high-ambiguity condition due to misidentification of the target letters was also examined. To do this, the participant's actual responses were compared with "substitute" sequences. The substitute sequences were made up of the letters from the opposite ends of the same continua as the letters in the target sequence, in the same order as the target sequence. For example, if the first letter in the target sequence was a "B–D" morph with a high probability of being identified as "B," the first letter in the substitute sequence would be a "D." This comparison was performed for the letter–letter conditions only. If the ambiguity cost was driven by participants misidentifying the more ambiguous tokens as the letter from the opposite end of the continuum (for example, identifying a "B–D" morph as "D" rather than "B"), we would expect a higher degree of similarity between the responses in the high-ambiguity conditions and these substitute sequences, compared to the similarity between responses in the low-ambiguity conditions and these substitute sequences. To examine this, the D–L distance between each response and the corresponding substitute sequence was calculated, and this distance was converted to the proportion of matching letters in the two sequences, as described previously. A LMM with sequence length (six or eight letters) and ambiguity (high or low) as fixed effects, and participant as a random effect, found no effect of ambiguity on similarity to the substitute sequences ($p = .72$). This indicates that responses in the high-ambiguity conditions did not contain a higher proportion of same-continuum substitutions than responses in the low-ambiguity lists.

Experiment 1: Discussion

Our results indicate that even a limited degree of perceptual ambiguity (letter name sounds which are correctly identified 90% of the time, compared with 99%) is sufficient to produce a marked impairment of pSTM recall. For six-letter sequences the proportion of sequences recalled correctly was markedly reduced by perceptual ambiguity. For our more sensitive measure of recall performance (proportion of letters recalled correctly), the ambiguity-induced impairment was also observed for eight-letter sequences. This suggests that a deficit in identifying speech sounds could result in a reduction of pSTM.

However, the detrimental impact on short-term memory was limited to ambiguous speech sounds which were made by averaging between two letter names (letter–letter morphs). Ambiguity did not affect recall of the letter–word morphs, despite equivalent identification performance for the letter–letter morphs and the letter–word morphs in the categorization pretest. This suggests that in order for ambiguity to affect pSTM, the ambiguity must increase the likelihood that a target item will be misidentified as a task-relevant alternative. In the case of the letter–letter morphs, both of the natural speech sounds that were combined to create the morph would be appropriate responses in a letter span context. For example, for a letter sound created by averaging between the natural letter name sounds "B" and "D," both "B" and "D" would be a possible stimulus in a letter span task. Conversely, for a letter–word morph, the word sound would not be a task-appropriate response for example, in the "B"–"We" case, "We" would not be a legitimate response in a letter span task. This may bias listeners toward a letter name interpretation, improving the ease with which ambiguous speech can be categorized. Based on these results, we conclude that perceptual ambiguity impairs recall only when it increases confusability between possible responses.

Interestingly, we did not find that the ambiguity cost was driven by misidentification of the audio-morphed sounds as the letter from the alternative end of the continuum. Responses to high-ambiguity sequences did not contain more alternative-end letters than responses to low-ambiguity sequences. This suggests that the detrimental impact of perceptual ambiguity may not be because the ambiguous letters themselves are misidentified, but because the increased demand of identifying ambiguous speech disrupts pSTM encoding and ongoing rehearsal processes more generally.

This finding has implications for perceptual impairment accounts of phonological deficits which we will return to in the general discussion. We also conducted a second experiment to replicate the effect of letter–letter ambiguity on pSTM performance, and explore whether perceptual ambiguity of speech sounds had prospective and retroactive effects on pSTM for unambiguous sounds, equivalent to effects previously shown for SiN by Rabbitt (1968).

Experiment 2: Justification

The purpose of our second experiment was to determine if processing incoming ambiguous speech disrupts ongoing STM maintenance of previously heard speech sounds (irrespective of their ambiguity). This design builds on previous studies which show that when background noise is used to mask later speech sounds in STM sequences, recall of clearly heard speech earlier in a sequence is retroactively impaired (Guang et al., 2020; Rabbitt, 1968). This finding for SiN suggests that perceptual challenges at the level of speech identification can additionally disrupt ongoing recall beyond the immediate period during which perceptual difficulty is experienced, perhaps by inhibiting STM rehearsal processes or by occupying cognitive capacity that would otherwise support accurate STM recall.

Our second experiment therefore followed a similar design to Rabbitt's (1968) study, with the aim of investigating whether the perceptual challenge posed by ambiguous speech parallels these previous observations for SiN with regards to retroactive effects. If observed, these retroactive effects might suggest that perceptual difficulties for individuals with DLDs could similarly disrupt the process of STM maintenance more broadly with effects that continue beyond the time period in which perceptually ambiguous input is heard. That is, perceptual impairment may affect performance on STM tasks even when not all the items that are heard and maintained are ambiguous. Alternatively, if perceptual challenges only impair processing or STM recall of particularly confusable or ambiguous items then less widespread or dramatic effects of perceptual difficulty on STM tasks would be expected since effects would not be expected for recall of less confusable items.

Experiment 2: Method

Participants

Twenty-six participants were recruited from Prolific, none of whom had taken part in the stimulus pretesting or Experiment 1. This sample size was determined in G*Power 3.1 (Faul et al., 2007) based on a repeated measures design with two within-subjects factors: one two-level factor (ambiguity) and one five-level factor (list type). Power was estimated at .95 and Cohens $f = .22$ (the effect size for the smallest significant effect in the previous experiment).

As previously, participants were prescreened, and all self-reported that they were native English-speaking adults aged between 18 and 45 years, who had grown up in the United Kingdom and had no reported history of dyslexia, speech-and-language therapy or hearing problems. In response to the question, "How would you describe your gender?" (free response box), 11 participants answered "male" or "man" and 15 participants answered "female" or "woman" (no other responses were given). No further demographic information was collected.

Design

The span task used in Experiment 2 was identical to the Letter Span task used in Experiment 1, with the following exceptions: (a) only letter–letter morphs were included, as reliable ambiguity costs were not observed for letter–word sounds in the previous experiment; (b) stimulus lists comprised five different combinations of high- and low-ambiguity letter names.

In the first sequence type, the first half of each sequence consisted of high-ambiguity letter names, and the second half of low-ambiguity letter names (termed the "high–low" condition). Thus, a six-item sequence in this condition would consist of three high-ambiguity letter name sounds followed by three low-ambiguity sounds. The second sequence type had the reverse arrangement of high- and low-ambiguity letter names (the "low–high" condition). The third and fourth sequence types consisted of only high-ambiguity sounds ("high–high") or low-ambiguity sounds ("low–low") respectively, and therefore replicated the letter–letter sequences in Experiment 1. The final sequence type was constructed of half high-ambiguity and half low-ambiguity letter names, arranged in a pseudorandom order (the "mixed" condition).

A novel group of participants ($N = 26$) attempted to recall 16 six- and eight-item lists of each sequence type, for a total of 32 trials per sequence type, and 160 trials in the experiment overall.

Analysis Strategy

Replication of Experiment 1

Firstly, we assessed whether the negative effect of letter–letter ambiguity on recall (as identified in Experiment 1) replicated in this independent participant group. We calculated the proportion of sequences recalled correctly, for each condition, and the proportion of letters recalled correctly, for each sequence, via the same methods as Experiment 1 (method: analysis strategy). We then applied LMMs with ambiguity (high or low) and sequence length (six or eight letters) as fixed factors and participant as a random factor, for each recall measure. This allowed us to evaluate whether recall was poorer for the low-ambiguity sequences compared with the high-ambiguity sequences, as we predicted, and whether the impact of ambiguity depended on short-term memory load (list length). Any interactions were investigated with LS means analysis, with Tukey adjustment.

Retroactive and Prospective Effects

The primary aim of Experiment 2 was to investigate if ambiguity in either the first or second half of the list would affect recall of the other half of the sequence. In line with the analyses conducted by Rabbitt (1968) and the replication of that study by Guang et al. (2020), we used half-list accuracy as one dependent measure, and additionally calculated the proportion of letters recalled correctly in each half-list.

To examine potential retroactive effects, we first calculated the proportion of first half-lists recalled correctly. Responses were scored as correct if the first three letters (in the six-item lists) or the first four letters (in the eight-item lists) of the participants' overall response matched the first three or four letters of the target sequence, regardless of the total length of the response. We then compared the proportion of first half-lists recalled correctly for the low–low, low–high, high–low and high–high conditions. This was achieved by applying an LMM with first-half ambiguity (low or high), second-half ambiguity (low or high) and length (six or eight letters) as fixed effects, and participant as a random effect. This analysis allowed us to assess whether second-half ambiguity impacted first-half recall, when first-half ambiguity was equated (i.e., would recall of the low-ambiguity first half of a low–low list be superior to recall of the first half of a low–high list).

We also calculated the proportion of letters recalled correctly for the first half-list, using a custom MATLAB function. This function first identified the best item-by-item alignment between the participant's response and the target sequence (the alignment which resulted in the lowest possible D–L distance for the entire sequence). For example, if the target sequence was "vndcmb" and the recalled response was "vdcmb," the best alignment between the sequences would be "v-dcmb" (a deletion of one character, for a D–L distance of 1). Each target sequence was separated into a first half-list and a second half-list, with the first three letters (in the six-item lists) or the first four letters (in the eight-item lists) comprising the first half-list. For the response sequences, the characters in the realigned response sequence that corresponded to the first half of the target sequence were taken as the first half-list. For the example above, the target sequence would be divided into "vnd" and "cmb," and the response would be divided into "v-d" and "cmb." The difference between each target half-list and the corresponding response half-list was then calculated. As previously described, a penalty of 1 was assigned for each insertion, deletion, or substitution of a character, and for each adjacent-character transposition. For adjacent-character transpositions which occurred in the middle of a sequence, transposing the last letter of the first half-list with the first letter of the second half-list, a .5 penalty was assigned to both half-lists.

The total number of penalties per half-list was taken as the D–L distance. In our example, the D–L distance for the first half-list would be 1, and the distance for the second half-list would be 0. As previously described, the D–L distance was transformed into the proportion of letters recalled correctly for each half-list, by subtracting the D–L distance from the length of the target half-list, and dividing by the length of the target half-list.

To further examine retroactive effects, we assessed the impact of first-half and second-half ambiguity on the proportion of letters recalled correctly from the first half of the list. We calculated the average proportion of letters recalled correctly from the first half-list in the high–high, high–low, low–high, and low–low conditions. These were compared with an LMM with first-half ambiguity (low or high), second-half ambiguity (low or high). and sequence length (six or eight letters) as fixed effects, and participant as a random effect.

Following a similar analysis strategy, we calculated the proportion of second half-lists recalled correctly, to identify potential prospective effects (whereby high ambiguity at the beginning of a list might impair recall of low-ambiguity items later in the list). Again, the proportion of second half-lists recalled accurately in the high–high, high–low, low–high, and low–low conditions were entered into an LMM with first-half ambiguity (low or high), second-half ambiguity (low or high), and length (six or eight letters) as fixed effects, and participant as a random effect. The proportion of letters recalled correctly in the second half of each list was also calculated, using the same method as for the first half-lists. A further LMM was applied, with identical structure as described in the previous paragraph.

Comparison of All Conditions

We also compared performance across all five sequence types, using both the proportion of sequences recalled correctly per condition and the proportion of letters recalled correctly as metrics. This allowed us to further examine whether the position of highly ambiguous items within a sequence affected recall, by comparing sequences where high-ambiguity items were randomly distributed ("mixed" condition) with lists where high-ambiguity items appeared at either the beginning or end ("low–high" or "high–low"). All analyses consisted of LMM, with sequence length (six or eight letters) and sequence type (high–high vs. high–low vs. low–high vs. low–low vs. mixed) as fixed effects and participant as a random effect.

Bayes Factor Analysis of Null Results

Following the hypothesis testing described above, Bayes Factor analyses were used to investigate nonsignificant effects (Dienes, 2014, 2021). These analyses compare evidence in favor of two competing hypothesis; in this case, we assessed whether our nonsignificant results from linear modelling were best interpreted as true null effects of perceptual ambiguity, or whether the data were merely insufficient for the ambiguity cost to reach significance in some situations (e.g., for letter–word ambiguity in Experiment 1). The alternative hypothesis was calculated based on estimates of the true ambiguity cost for letter–letter lists. To ensure that this hypothesis was informed by all available data, data from both experiments were combined using a meta-analytic Bayesian method. Here the prior distribution was constructed using the first dataset (Experiment 1), the likelihood distribution was based on the second dataset (Experiment 2), and the mean and standard error of the resulting posterior distribution were taken as the new estimates. Bayes Factors were then calculated using the bayesplay package (V0.9.2; Colling, 2023) to compare the alternative (ambiguity cost) hypothesis to the null hypothesis.

Transparency and Openness

All manipulations and measures used in the experiments, as well as sample size determinations and exclusions, are described in the Method section. This is in line with the Journal Article Reporting Standards (Kazak, 2018). Data, analysis code and audio materials are available via the Open Science Framework (https://osf.io/wc3gn/) at the time of publication. All analyses were conducted in R (R Version 4.1.3; R Core Team, 2022), using the lme4 package (Bates et al., 2014) to calculate LMMs, the lsmeans package (Lenth, 2016) for LS-means analyses, and the bayesplay package (v0.9.2; Colling, 2023) for Bayes factor calculation. The design and analyses were not preregistered.

Experiment 2: Results

Replication of Experiment 1

Analysis of the number of sequences recalled correctly in the high–high condition compared with the low–low condition successfully replicated the ambiguity cost identified in Experiment 1. Our LMM showed a main effect of ambiguity, such that fewer high-ambiguity sequences were recalled perfectly, compared with low-ambiguity sequences, $F(1, 75) = 27.29$, $p < .001$, $\eta_p^2 = .27$.

In contrast to Experiment 1, we did not observe an interaction effect between sequence length and ambiguity for proportion of sequences recalled correctly; L–S means comparisons with Tukey adjustment showed that the ambiguity cost was present for both six-letter, $t(75) = -4.48$, $p < .001$, $d = -0.71$, and eight-letter, $t(75) = -2.92$, $p = .02$, $d = -0.46$, sequences.

A similarly compelling replication of previous results was observed for analysis of the proportion of letters recalled correctly in each condition. Again, a main effect of ambiguity was observed (LMM), $F(1, 75) = 19.31$, $p < .001$, $\eta_p^2 = .2$, reflecting better recall of low-ambiguity versions of the sequences (low–low trials) than high-ambiguity versions (high–high trials). As previously, the ambiguity cost was present for both six-letter, $t(75) = -3.55$, $p = .004$, $d = -0.43$, and eight-letter, $t(75) = -3.66$, $p = .003$, $d = -0.45$, sequences (Figure 4).

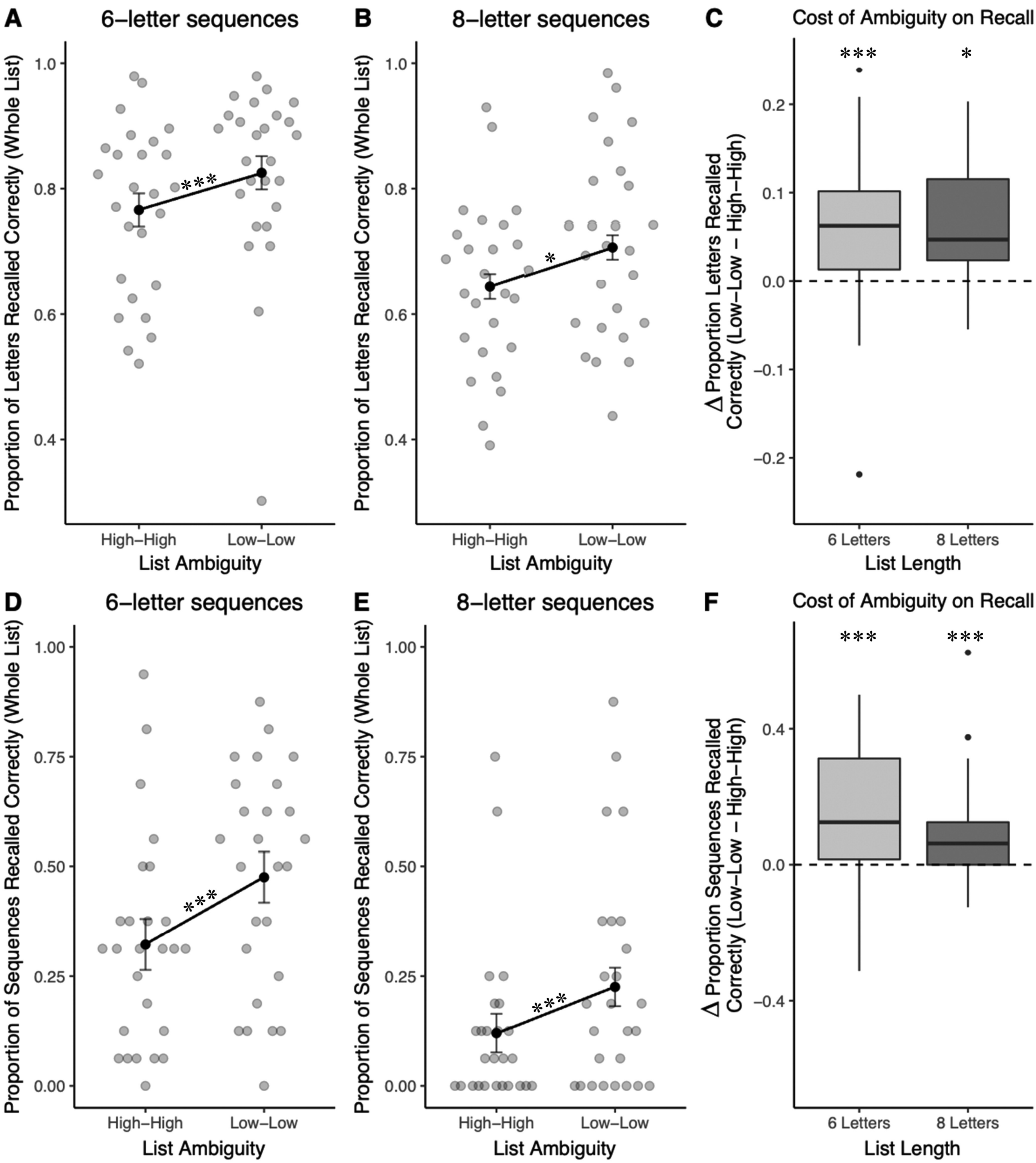
To summarize, Experiment 2 successfully replicated the ambiguity cost identified in Experiment 1, in a novel group of participants.

Retroactive and Prospective Effects

To assess potential retroactive effects, an LMM was conducted with the proportion of first half-lists recalled correctly as the dependent measure. First-half ambiguity (low or high), second-half ambiguity (low or high) and length (six or eight letters) were included as fixed effects, and participant as a random effect. Results showed the expected main effects of length, $F(1, 201) = 33.71$, $p < .001$, $\eta_p^2 = .14$, such that shorter lists were recalled more accurately, and of first-half ambiguity $F(1, 201) = 8.82$, $p = .003$, $\eta_p^2 = .04$, such that a higher proportion of first half-lists were recalled perfectly if they consisted of low-ambiguity letters, compared with high-ambiguity letters. However, there was no effect of second-half ambiguity ($p = .61$), and no interactions emerged. In other words, recall of low-ambiguity letters at the beginning of a sequence was not affected by the presence of high-ambiguity letters later in the sequence, in contrast to findings from the SiN literature.

Figure 4

Recall Performance for the High-Ambiguity and Low-Ambiguity Sequences in Experiment 2



Note. In Panels A and B, the semitransparent points show the average proportion of letters recalled correctly in sequences from each condition, for each participant. In Panels D and E, the semitransparent points show the proportion of sequences recalled perfectly in each condition, out of a maximum of 16, for each participant. The opaque points represent group averages. The error bars show 95% confidence intervals, calculated according to the Cousineau–Morey method for within-subjects designs. Panels C and F show the differences in recall performance between the high-ambiguity conditions and low-ambiguity conditions (ambiguity cost). For complete details of boxplot interpretation, see Figure 3.

* $p < .05$. ** $p < .01$. *** $p < .001$.

A further LMM was applied to examine the impact of first-half and second-half ambiguity on the proportion of letters recalled correctly from the first half-list (Figure 5). Again, results showed the expected main effects of length, $F(1, 201) = 16.87$, $p < .001$, $\eta_p^2 = .08$, and of first-half ambiguity, $F(1, 201) = 4.17$, $p = .03$, $\eta_p^2 = .02$, such that the first half of the list was recalled more accurately if it consisted of low-ambiguity letters, compared with high-ambiguity letters. However, there was no effect of second-half ambiguity ($p = .5$), and no interactions were found. Thus, for our more sensitive measure of proportion of letters recalled correctly, the ambiguity cost for high-ambiguity letters was maintained when only recall for the first half of a sequence was assessed. However, recall of low-ambiguity letters at the beginning of a sequence was not affected by the presence of high-ambiguity letters later in the sequence.

The proportion of second half-lists recalled accurately in the high–high, high–low, low–high, and low–low conditions were also entered into an LMM with first-half ambiguity (low or high), second-half ambiguity (low or high) and length (six or eight letters) as fixed effects, and participant as a random effect (Figure 6). As expected, when considering only recall for the second half of the list, the effect of second-half ambiguity was significant, $p = .005$, $F(1, 201) = 7.96$, $\eta_p^2 = .04$, with a higher proportion of second half-lists recalled accurately when they consisted of low-ambiguity letters, compared to high-ambiguity letters. This indicates that the ambiguity cost is also present when we only consider recall of the second half of a list. However, first-half ambiguity did not affect the proportion of second half-lists recalled accurately ($p = .55$), and no interactions were identified.

For the comparison of the proportion of letters recalled from the second half of each list type, the expected main effect of second-half ambiguity, $p < .001$, $F(1, 179) = 29.76$, $\eta_p^2 = .14$, was found, such that recall of letters from the second half of a list was higher when these letters were low ambiguity, than when they were high ambiguity. However, first-half ambiguity did not affect the proportion of letters recalled accurately from second half-lists ($p = .11$), and no interactions emerged. Therefore recall of letters from the second half of a list was not affected by the ambiguity of the first half of that list. Thus, our more sensitive measure confirmed the ambiguity cost for second half-lists, but failed to show any prospective effects (Figure 6).

The lack of retroactive or prospective effects in this experiment suggests that the ambiguity cost is limited to recall of the high-ambiguity stimuli only, and does not transfer to affect recall of low-ambiguity items within the same sequence.

Comparison of All Conditions

We used additional LMMs to examine recall across all five combinations of high and low ambiguity (Figure 7). For the proportion of sequences recalled correctly, main effects of sequence length, $F(1, 225) = 213.64$, $p < .001$, $\eta_p^2 = .49$, and sequence type, $F(4, 225) = 9.64$, $p < .001$, $\eta_p^2 = 0.15$, were identified, with no interaction effect. However, post hoc tests (L–S means with Tukey adjustment) showed no significant differences in accuracy between any of the same-length sequences containing both high- and low-ambiguity items (high–low vs. low–high vs. mixed). This suggests that the main effect of sequence type is primarily driven by the difference in recall accuracy between high–high and low–low sequences.

The pattern of results for the proportion of letters recalled correctly was very similar; we identified main effects of sequence length, $F(1, 225) = 596.79$, $p < .001$, $\eta_p^2 = .73$, and sequence type, $F(4, 225) = 6.58$, $p < .001$, $\eta_p^2 = .10$, but no interaction. As with the accuracy data, there were no differences in D–L distance between any of the conditions with a mixture of high- and low-ambiguity letter names irrespective of whether ambiguous letter names were heard at the start, end, or mixed throughout the list. This indicates that the position of ambiguous items within a sequence does not affect recall.

Bayesian Meta-Analysis of Null Findings

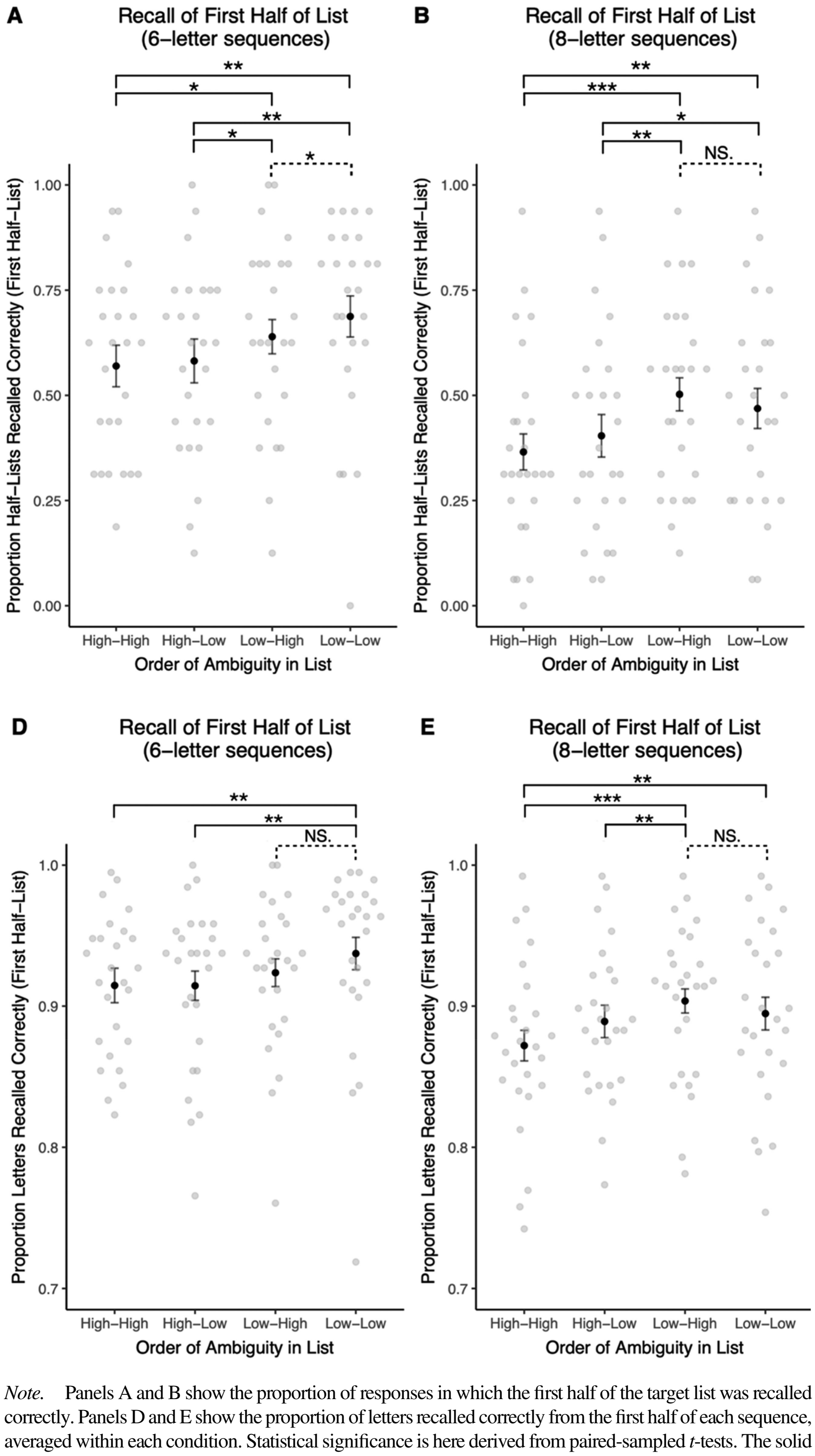
Across Experiments 1 and 2, we identified several situations in which perceptual ambiguity did not significantly affect pSTM performance. Specifically, in Experiment 1 we found an interaction between continuum type and ambiguity, such that the ambiguity cost was significantly reduced for letter–word ambiguity compared with letter–letter ambiguity. Direct comparisons of recall between high-ambiguity and low-ambiguity sequences found no significant differences for letter–word ambiguity. In Experiment 2, we did not observe reliable prospective or retrospective effects of perceptual ambiguity on pSTM performance. We applied Bayesian hypothesis testing to three null effects: (a) effect of letter–word ambiguity on recall in Experiment 1; (b) prospective effects of letter–letter ambiguity on recall of the second half of lists with split-half ambiguity in Experiment 2; and (c) retrospective effects of letter–letter ambiguity on the first half of split-half lists in Experiment 2. In all cases we applied these analyses separately to our two dependent measures (proportion of sequences, and letters reported correctly), with estimates of the true ambiguity cost for letter–letter lists as our alternative hypothesis.

Starting with the letter–word sequences in Experiment 1, the mean ambiguity cost (recall for high-ambiguity sequences subtracted from recall for low-ambiguity sequences) for the proportion of sequences reported correctly was .0058 ($SEM = .014$), based on combined data from six- and eight-item sequences. In contrast, the ambiguity cost for the proportion of letter–letter sequences reported correctly (based on combined data from six- and eight-item sequences in Experiments 1 and 2) was .073 ($SEM = .014$). Based on these values, we computed a Bayes Factor of $BF_{10} = .0051$. This indicates extreme evidence in favor of the null hypothesis (Lee & Wagenmakers, 2014), suggesting that the degree of letter–word ambiguity did not affect recall. A similar outcome was obtained for this comparison when using the ambiguity cost in terms of the proportion of letters reported correctly as the dependent measure. The mean ambiguity cost for letter–word lists was .0056 ($SEM = .0065$) and the ambiguity cost for letter–letter lists was .051 ($SEM = .0067$). Hence, the Bayes factor again provides extreme evidence in favor of a null effect of ambiguity on letter–word lists ($BF_{10} < .0001$). These findings suggest that the effect of letter–word ambiguity on pSTM was not merely smaller than the effect of letter–letter ambiguity, but was most likely entirely absent.

To assess the potential prospective effect of ambiguity in the first half of a sequence on the second half of that sequence, we first considered the proportion of second half-lists recalled correctly. The true ambiguity cost was taken as the difference in second half-list recall between conditions which had the same first-half ambiguity,

Figure 5

Recall Performance for the First Half of Each List, in Every Condition

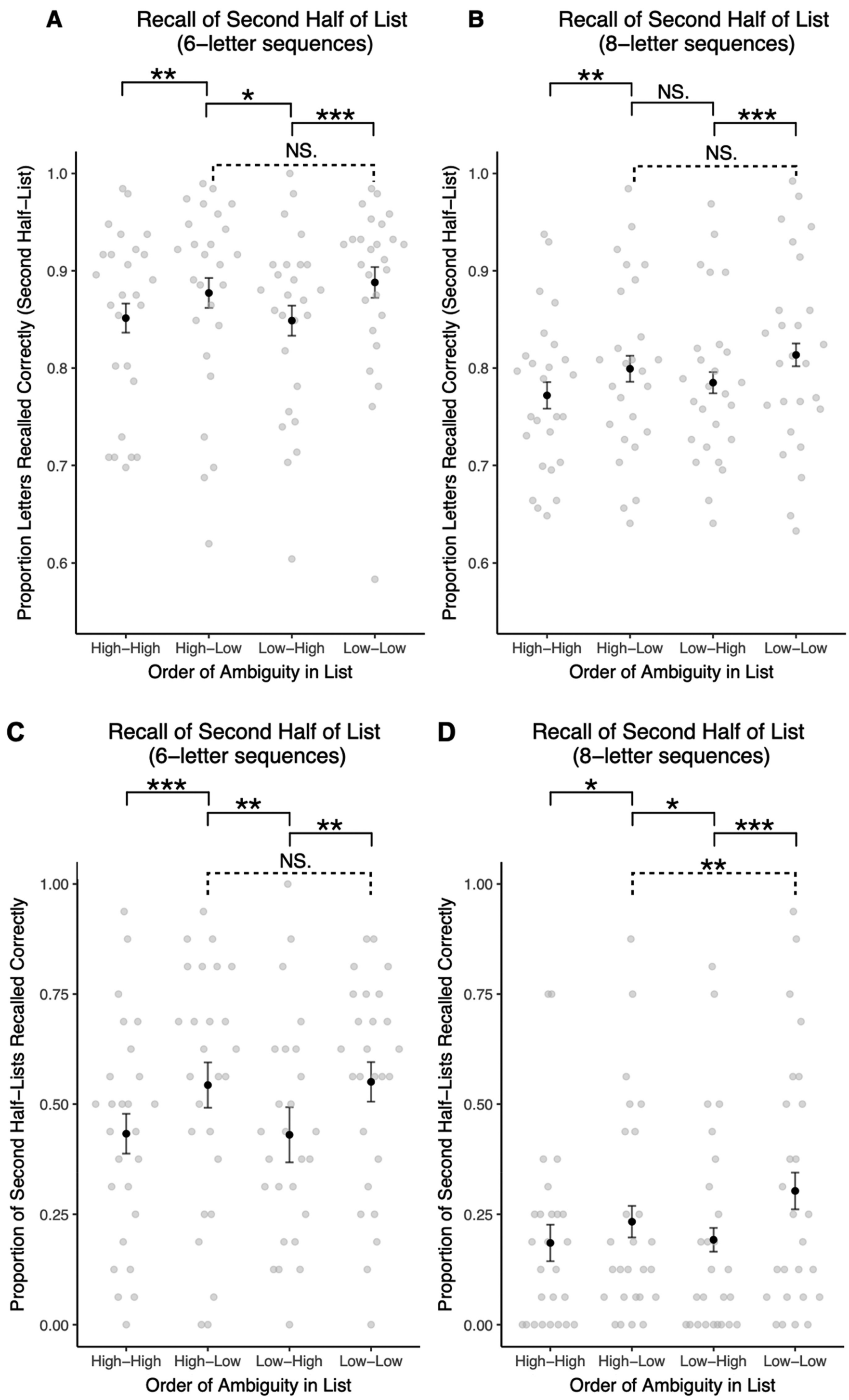


Note. Panels A and B show the proportion of responses in which the first half of the target list was recalled correctly. Panels D and E show the proportion of letters recalled correctly from the first half of each sequence, averaged within each condition. Statistical significance is here derived from paired-sampled *t*-tests. The solid lines linking conditions show the direct effect of ambiguity, comparing first-half recall for sequences where the first half consisted of high-ambiguity letters with conditions where the first half consisted of low-ambiguity letters. For ease of interpretation, only statistically significant direct effects are shown. The dashed line linking the low–high and low–low conditions shows the indirect effect of second-half ambiguity on first-half performance (the retroactive effect). NS = not significant at the $p < .05$ level.

$^{*} p < .05$. $^{**} p < .01$. $^{***} p < .001$.

Figure 6

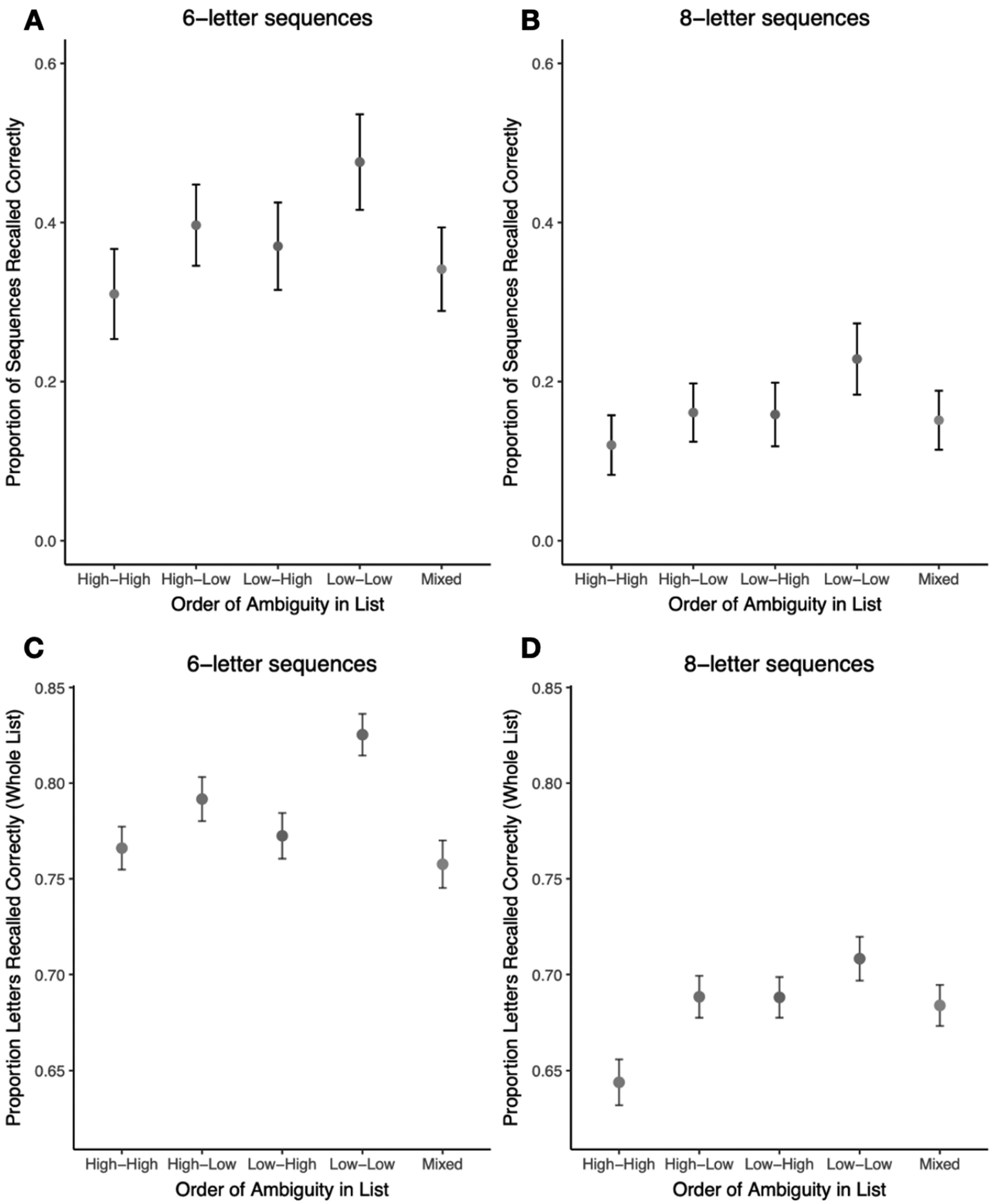
Recall Performance for the Second Half of Each List, in the High–Low and Low–Low Conditions



Note. Panels A and B show the proportion of responses in which the second half of the target list was recalled correctly, for the six- and eight-letter sequences, respectively. Panels C and D show the proportion of letters recalled correctly from the second half of each response, averaged within each condition. Translucent points represent data for individual participants and opaque points represent group averages. The error bars show 95% confidence intervals, calculated according to the Cousineau–Morey method for within-subjects designs, as previously described. NS = not significant at the $p < .05$ level.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Figure 7

Average Proportion of Sequences Recalled Correctly, for Each Condition in Experiment Two (Panels A and B) and Average Proportion of Letters Recalled Correctly, for Each Condition in Experiment Two (Panels C and D)



Note. The error bars show 95% confidence intervals, calculated according to the Cousineau-Morey method, as previously described.

but differing second-half ambiguity (e.g., the cost associated with a "high–high" list compared with a "high–low" list, or with a "low–high" list compared with a "low–low" list). Combining over six- and eight-item lists, the mean ambiguity cost was .094 ($SEM = .013$). The prospective effect was taken as the difference in second half-list recall between conditions which had different first-half ambiguity, but equivalent second-half ambiguity (e.g., the cost associated with a “high–high” list compared with a “low–high” list, or with a “high–low” list compared with a “low–low” list). Combining over six- and eight-item lists, the mean prospective effect was .022 ($SEM = .014$). Assessing the prospective effect with the ambiguity cost as the alternative hypothesis resulted in a Bayes Factor of $BF_{10} = .0044$, indicating “extreme” evidence in favor of the null hypothesis (Lee & Wagenmakers, 2014). For the proportion of letters recalled correctly from the second half-list, the ambiguity cost ($M = .03$, $SEM = .0043$) and prospective effect ($M = .0098$, $SEM = .0047$) were computed in the same way, resulting in “strong” evidence in favor of the null hypothesis ($BF_{10} = .04$).

In a similar fashion, we investigated the potential retrospective effect of ambiguity in the second half of a sequence on the first half of that sequence. Here the ambiguity cost was taken as the difference in first half-list recall between conditions which had the same second-half ambiguity but differing first-half ambiguity. For the proportion of first half-lists recalled correctly, combining six- and eight-item lists, the mean ambiguity cost was .099 ($SEM = .017$). The retrospective effect was taken as the difference in first half-list recall between conditions which had the same first-half ambiguity, but contrasting second-half ambiguity. Combining over six- and eight-item lists, the mean retrospective cost was .016 ($SEM = .014$). Assessing the retrospective effect with the ambiguity cost as the alternative hypothesis resulted in a Bayes factor of $BF_{10} = .003$, again interpreted as “extreme” evidence for the null hypothesis. For the proportion of letters recalled correctly from the first half-list, the ambiguity cost ($M = .019$, $SEM = .0038$) and retrospective effect ($M = .0047$, $SEM = .0034$) were calculated in the same way, resulting in “strong” evidence for the null hypothesis ($BF_{10} = .04$).

Experiment 2: Discussion

Our comparisons of recall for the entirely high-ambiguity lists with the entirely low-ambiguity lists in Experiment 2 clearly replicated the ambiguity cost identified in Experiment 1. As in the previous experiment, phonetic ambiguity impaired short-term recall of letter sequences.

However, in contrast to findings from the SiN literature, we failed to identify either retroactive or prospective effects of phonetic ambiguity. The presence of high-ambiguity letters in a sequence had no significant effect on recall of the surrounding low-ambiguity items, whether they occurred before or after the high-ambiguity sounds. Furthermore, Bayesian hypothesis testing suggests that the appropriate way to interpret our nonsignificant results is to conclude that perceptual ambiguity does not prospectively or retrospectively impact short-term recall. This indicates that the challenge of identifying incoming ambiguous speech does not disrupt ongoing rehearsal of information already held in pSTM, or lead to subsequent difficulty in identifying and maintaining low-ambiguity speech. Hence, the disruption of pSTM due to perceptual ambiguity is limited just to the high ambiguity letters irrespective of when they occur in a list. This finding supports the interpretation that the detrimental effects of perceptual ambiguity on pSTM occur because ambiguous speech is more challenging to identify and not because ambiguous speech also impacts on other higher-level processes that contribute to pSTM, such as rehearsal.

General Discussion

In two experiments with independent participant groups, we have reliably found that perceptual ambiguity impairs short-term recall of speech. Recall of more ambiguous letter name sounds was consistently poorer than recall of less ambiguous letter name sounds, despite the relatively minor difference in ambiguity (90% compared with 99% identification accuracy), and both sounds belonging to the same phonetic category. The ambiguity cost on recall is consistent and of moderate size, with Cohen’s d ranging from $d = 0.71$ to $d = 0.3$ (the smallest of these for a condition involving eight-letter sequence recall with near floor levels of performance). In the studies reported here this effect has been replicated across two different participants groups and found for two difference performance measures (lists or letters recalled correctly). It is even observed (in Experiment 2) for lists in which some but not all letters are perceptually ambiguous.

The cost of perceptual ambiguity on pSTM we have identified here is consistent with the well-documented negative impact of phonological confusability on short-term memory, including the phonological similarity effect (Baddeley, 1968; Laughery & Pinkus, 1966), and poor recall of SiN (Rabbitt, 1968; Surprenant, 1999). However, our findings also advance the existing literature considerably, as they demonstrate that a pure manipulation of speech perception difficulty can result in a pSTM impairment. Our perceptually ambiguous tokens comprise a single acoustic source, are created by averaging between natural speech sounds, and have a precisely controlled degree of perceptual ambiguity. Therefore they do not place high demands on the same domain-general executive resources (particularly selective attention) required to identify SiN. They also sound substantially more natural—for the low levels of perceptual ambiguity that we tested here it may not be immediately obvious which letter sequences include ambiguous speech sounds. Furthermore, this manipulation provides a highly ecological simulation of the kinds of categorical perception deficits that have previously been documented in individuals with phonological impairment. Thus, we have shown that pSTM can be compromised by a subtle reduction of discriminability at the acoustic level. Further analysis showed that this reduction in pSTM is not explained by simple perceptual confusions—that is, participants are not simply reproducing the alternative percept of the ambiguous letters. In summary, our perceptual ambiguity cost demonstrates that even a minor challenge to the speech perception system can reliably diminish pSTM capacity.

Despite these observations, we also observed several situations in which perceptual ambiguity did not significantly affect pSTM performance. Notably, in our first experiment the ambiguity cost was dependent on the type of continuum the ambiguous items had been drawn from (letter–letter or letter–word). Furthermore, Bayesian hypothesis testing provided strong evidence for a null effect of perceptual ambiguity on recall of letter–word morphs (rather than the continuum-by-ambiguity interaction merely reflecting a smaller ambiguity cost for letter–word morphs, for example). Therefore we can conclude that the ambiguity cost was only present for recall of letter–letter morphs, and not for recall of letter–word morphs. Similarly, we found strong evidence against any impact of high-ambiguity items on recall of surrounding low-ambiguity items within the same list, in Experiment 2.

The Effect of Ambiguity on Short-Term Memory Is Specific, Distinct, and Limited

We therefore conclude that, although perceptual ambiguity can impair pSTM, this effect appears to be limited to specific perceptual circumstances. Firstly, the lack of ambiguity cost for letter–word morphs suggests that perceptual ambiguity will only impair recall when the ambiguity is between two task-relevant responses. In our letter span task, both of the natural speech stimuli used to create each letter–letter morph would be valid, task-appropriate responses—for example, both "B" and "D" are possible items in a letter sequence, and participants had difficulty recalling items created by averaging between "B" and "D" sounds. The impairment in recalling high-ambiguity letter–letter stimuli may derive from competition between these two possible interpretations of the speech sound, in the letter span task context. For letter–word morphs, which were created by averaging between a letter name sound and a word sound, only one of the original speech stimuli (the letter) would be a valid response in a letter span task. This makes ambiguity resolution less challenging, as the task context heavily biases the listener toward a letter name interpretation. The lack of ambiguity cost for letter–word morphs suggests that participants can establish that the present task involves recalling lists of letters, and are therefore biased against interpreting stimuli as other words, even when they are perceptually ambiguous. This indicates that top-down influences such as task set can modulate the degree to which perceptual ambiguity impacts pSTM recall.

Additionally, we did not identify either retroactive or prospective effects of ambiguity on recall—memory for low-ambiguity letters was unaffected by the presence of high-ambiguity letters either earlier or later in the same sequence. These results are in contrast to Rabbitt's (1968) finding that recall of clear speech is impaired by subsequent masking noise in subsequent syllables (SiN). One possible explanation for this difference which we can reject is that SiN identification is more difficult than phonological ambiguity resolution, and therefore the disruption the perceptual challenge imposes on pSTM is more substantial. Although Rabbitt (1968) did not report the signal-to-noise ratios used, Guang et al. (2020) used a noise level of −3 dB, which was confirmed via pilot study as the highest noise level resulting in 99% speech identification accuracy. Our high-ambiguity stimuli were created to produce an identification accuracy of 90%, and therefore ambiguity resolution is more perceptually challenging than Guang's SiN stimuli. Following the logic of this explanation, ambiguity resolution would be expected to have a greater impact on pSTM than SiN. This means that a difficulty-based explanation is unconvincing.

A more persuasive explanation is that SiN and phonological ambiguity resolution create different challenges for speech identification. In SiN perception, listeners must segregate a complex auditory scene into two sound sources (speech and noise), and focus on one whilst suppressing the other. Attentional systems must be engaged to segregate the signals and select the relevant speech information, in addition to the challenge of identifying individual speech units. Selective attention (Heald & Nusbaum, 2014), inhibition (Stenbäck et al., 2015; Veneman et al., 2013) and working memory (McCreery et al., 2017; Rönnberg et al., 2010, 2014) all account for independent variance in SiN performance (see Dryden et al., 2017 for review), which highlights the importance of executive skills in dealing with this challenge. In contrast, in the phoneme blending technique the signals of two natural speech sounds are averaged, resulting in a single acoustic object (one spoken syllable) which to some degree resembles two different letter names (e.g., B and D). This poses a purer challenge to the speech perception system than SiN, as the additional source segregation problem is removed. With this difference in mind, our findings may indicate that the retroactive effects observed in the SiN literature may be the result of domain-general executive mechanisms, rather than a speech identification challenge per se. We note that there is a substantial literature showing that extraneous nonspeech noises can impair pSTM performance (e.g., the suffix effect: Rowe & Rowe, 1976; the irrelevant sound effect: Hughes & Jones, 2001; Schlittmeier et al., 2012), highlighting the importance of auditory attention for successful recall. The retrospective effects observed by Rabbitt might similarly reflect this. Future work should consider whether other attentional challenges—for example, distraction, or dual task situations—can also produce pSTM impairments without also creating perceptual difficulty.

In summary, we were only able to demonstrate a causal link from disrupted speech perception to pSTM difficulties when the speech-sound ambiguity occurred between highly confusable alternatives, such as two letter names with minimal phonetic difference. Furthermore, certain task contexts (e.g., letter lists) were able to eliminate effects of task-irrelevant perceptual ambiguity on pSTM, and the negative impact of the perceptual challenge did not transfer to other items being maintained in pSTM. These findings suggest that although speech perception deficits can impair pSTM, this only occurs in specific task contexts in which perceptual demands result in response uncertainty. Furthermore, Experiment 2 indicates that pSTM impairment is due to the difficult of correctly identifying ambiguous items, rather than a perceptual challenge disrupting non-perceptual processes such as rehearsal of information already held in pSTM. As we will explain in the subsequent section, these findings substantially challenge perceptual deficit accounts that have been invoked to explain reduced pSTM for stimuli (such as digits) that are largely not confusable when presented in the context of digit span tasks (see Introduction for overview).

Implications for Perceptual Deficit Accounts of Phonological Impairment

These findings challenge a perceptual deficit explanation for poor pSTM in DLDs. In our experiments, the perceptual system was challenged by creating ambiguity between speech sounds, making the letter name sounds less distinctive. This simulates the type of perceptual dysfunction that has been proposed for individuals who have failed to develop adequately specified speech-sound representations. However, this perceptual disruption did not lead to the type of pSTM difficulties seen in children with DLDs. Instead, the effect of perceptual ambiguity on recall was restricted to highly confusable response alternatives.

Children with DLDs perform poorly on a broad range of pSTM tasks, including word, letter, and digit span (Baddeley & Hitch, 1974; Gathercole & Baddeley, 1990). Although some items in a traditional digit, letter or word span task may be perceptually confusable, most items used in these tasks are highly distinctive. If ambiguity between two very similar sounding speech tokens, such as "B" and "we," is insufficient to impair pSTM (as observed in Experiment 1), it seems implausible that any two digits would be sufficiently confusable to impact pSTM capacity, even for a perceptual system with very poor discrimination (one illustration of this difference in confusability comes from the widespread use of the North Atlantic Treaty Organization phonetic alphabet—Alpha, Bravo, Charlie, etc.—in place of letter names; the only similar change made for the numbers 0 to 9 is the use of "niner" for "nine" to avoid confusion with the number "five"). For this pair of digit confusions to affect recall of digit sequences would require that the impact of the perceptual challenge must transfer to less confusable or ambiguous items, as has previously been observed for SiN. However, the lack of prospective or retroactive effects of perceptual ambiguity in our second experiment indicates that the presence of perceptually challenging speech is not sufficient to affect recall of less-confusable verbal information. On the basis of the present findings, therefore, we do not think that a perceptual deficit account can sufficiently explain impaired digit span in phonologically impaired individuals.

As we did not find evidence in favor of a perceptual deficit account, we should return to alternate explanations for poor performance on both pSTM tasks and speech perception measures. We therefore suggest separating the processes involved in perceiving speech sounds, from the processes involved in maintaining and reflecting on those speech sounds. This is in line with work by Ramus, Szenkovits, et al., who distinguish between the quality of phonological representations, which are proposed to be intact in dyslexia, and access to these representations, which they suggest is impaired (Ramus & Ahissar, 2012; Ramus & Szenkovits, 2008). Poor performance on pSTM tasks may not reflect a failure in matching auditory input to phonetic representations during encoding, but rather a deficit in maintaining and/or accessing this information at the rehearsal and retrieval phases. More generally, difficulties on phonological tasks are likely to reflect deficits in higher-level metalinguistic abilities, rather than basic auditory processes.

Returning to the multiple-deficit framework, we should also consider that a range of domain-general cognitive functions interact with speech processing (Ganong, 1980; Gordon-Salant & Cole, 2016; McClelland et al., 2006). These could act as protective or risk factors to modulate the impact that individual variation in perceptual acuity has on task performance and phonological development. In the current experiments, our participants may have been able to utilize additional cognitive resources to compensate for perceptual ambiguity. Indeed, the null effect of letter–word ambiguity suggests that the impact of perceptual difficulties on pSTM is, at least in part, moderated by the ability to use contextual information effectively. It is possible that children with dyslexia and DLD may have weaker speech perception, or speech perception at the lower end of the normal range (Hazan et al., 2009), and be less able to engage these top-down resources to compensate for this disadvantage.

We can only offer speculation about the precise nature of the executive dysfunction that might be required (independently or in addition to perceptual impairment) to account for pSTM reductions. Our results could be interpreted in-line with the anchoring hypothesis, which suggests that a deficit in the ability to tune to incoming stimuli affects both speech perception and the efficiency of pSTM (Ahissar, 2007; Oganian & Ahissar, 2012). Our finding that the ambiguity cost could be abolished by integrating knowledge of task-set suggests that pSTM problems would persist if individuals were unable to tune their encoding to task-specific information, which mirrors the hypothesized anchoring deficit in dyslexia. However, evidence regarding the importance of anchoring for phonological processing remains mixed (Di Filippo et al., 2008; Ziegler, 2008). Rather than a single independent cause, anchoring issues may be one of a set of domain-general processing deficits that could explain co-occurring speech perception and phonological problems. Indeed, other executive functions (e.g., working memory, procedural learning) are also plausible candidates, in addition to sensorimotor aspects of speech processing. We believe it is premature to offer any definitive statement on these points at the present time since evidence from our own work is lacking. Furthermore, although our results suggest that a perceptual deficit is not sufficient to impair pSTM, we cannot completely exclude perceptual problems as a contributory factor in some individuals.

If phonological difficulties are not caused by perceptual dysfunction, the question remains as to why children with these difficulties often perform poorly on speech perception measures. A likely explanation is that the tasks used to study speech perception place high demands on nonperceptual abilities, including pSTM (Manis et al., 1997). In an ABX discrimination paradigm, the classical categorical perception procedure, participants must decide whether stimulus X matches stimulus A, presented first, or stimulus B, presented second (Liberman et al., 1957). This requires the participant to hold three consecutive sounds in memory for comparison, which may be particularly challenging for a child with a pSTM deficit. Indeed, Coady and colleagues (Coady et al., 2007) manipulated pSTM demands in an identification task and reported comparable performance between children with DLD and age-matched peers when memory load was minimized (through short inter-stimulus intervals and natural speech).

It is also important to contextualize our findings within a developmental perspective. Phonological abilities in typically developing children, and phonological impairment in those with language and reading disorders, emerge as part of a complex and extremely interactive developmental process (e.g., Bentin & Leshem, 1993; Carroll et al., 2003; Dickinson et al., 2003; Hogan, 2010). Auditory perceptual theories suggest that speech perception deficits interfere with the typical acquisition of information about speech structure, and therefore the development of normal phonological representations, during early language learning (Elbro et al., 1994; Goswami, 2011; Joanisse et al., 2000). Thus, speech perception issues as a young child may affect the development of neural mechanisms that support pSTM, without poor perception necessarily persisting into later childhood or adulthood. By this view, speech perception abilities act to "scaffold" the acquisition of phonological awareness, and literacy, but might not continue to influence adult performance in pSTM tasks.

However, further research is required in order to conclude that the fully developed adult language system responds to a challenge to speech perception (through ambiguity) in a different way than a developing language system would. Children as young as 5 years can use both lexical and nonlinguistic context to resolve perceptual ambiguity (Hufnagle et al., 2013; McQueen et al., 2012), which indicates that typically developing children deal with ambiguity using the same combination of bottom-up and top-down mechanisms as adults. This suggests that the compensatory executive mechanisms observed in adults in our study would also support children's comprehension when speech perception is ambiguous (whether due to internal or external sources of ambiguity). If perceptual dysfunction is particularly detrimental for pSTM during development, we might expect children who have experienced temporary hearing loss (e.g., due to otitits media [OME]) to show corresponding pSTM problems. Majerus et al. (2005) identified speech perception deficits in such children (aged 7–9 years, with recurrent OME before 3 years), but their pSTM capacity was equivalent to controls. This suggests that temporary changes in perceptual acuity (parallel to our perceptual ambiguity manipulation, although longer lasting) do not impact pSTM development, even when the perceptual dysfunction occurs during a key period for language acquisition (before 3 years).

We might also consider whether speech perception and pSTM skills are differentially associated in children and adults—if perceptual abilities are a stronger predictor of pSTM in childhood, this could indicate that pSTM is more reliant on perception during development. We are not aware of research directly comparing this relationship in children and adults, but some studies have reported positive correlations between perceptual function (SiN tolerance) and pSTM in children without language impairments (Magimairaj et al., 2018; Roman et al., 2017). In contrast, a meta-analysis by Füllgrabe and Rosen (2016) concluded that individual differences in span are not consistently related to speech perception for unimpaired adults with typical hearing. This suggests that perceptual and pSTM abilities may be more closely related in development than in adulthood, but this does not provide evidence for a causal relationship. Furthermore, SiN is an imperfect measure of speech-sound acuity, as discussed earlier in this paper. In summary, we cannot exclude the possibility that speech perception influences pSTM differently during development, but neither does the existing evidence provide a strong argument in favor of this view.

A natural progression would be to apply the ambiguity cost paradigm to a group of (typically developing) children, to test whether effects of perceptual ambiguity on pSTM are equivalent in children and adults. For example, can children effectively integrate task context to resolve this form of ambiguity, or are they (unlike adults) disrupted by letter–word ambiguity? Another fruitful line of research might be to simulate perceptual deficits in computational models of short-term memory (e.g., Burgess & Hitch, 2006; Hurlstone et al., 2014). Both Joanisse and Seidenberg (2003) and Jones and Westermann (2022) found that connectionist models trained on perceptually degraded speech learned language abnormally (with resulting deficits in word comprehension and the quality of phonological representations, respectively). In a similar fashion, contrasting (a) a pSTM model which has been trained on perceptually unclear stimuli with (b) a model initially trained on normal speech, which then has a perceptual deficit added after it is fully functioning, would help determine whether perceptual difficulties differentially affect a developing pSTM system and a mature system.

The nature of the phonological problems observed in dyslexia and DLD is also worth addressing. Here we have taken a transdiagnostic approach, focusing on phonological difficulties as a well-documented feature of both disorders. However, phonological issues are only one possible dimension of DLD. DLD is a complex, multifactorial diagnosis, and individuals with DLD can exhibit poor language skills in a wide range of linguistic domains (Bishop et al., 2017). Our findings indicate that poor speech perception is unlikely to produce the type of phonological STM problems observed in some individuals with DLD. However, our interpretation of these results is limited to the phonological component of this disorder—other deficits (such as in grammar or lexical processing) may have a different relationship to speech perception.

In the case of dyslexia, the notion of a phonological core deficit is more broadly accepted (Snowling, 1998; Stanovich, 1986; Vellutino et al., 2004), though not uncontroversial (see, e.g., Bosse et al., 2007; Vidyasagar & Pammer, 2010). However, the present study has only examined the relationship between speech perception and pSTM, and not between perception and the other components of Wagner and Torgesen's "phonological triad" (Wagner & Torgesen, 1987)—namely, phonological awareness and rapid automatized naming (RAN). These three components have been shown to differentially contribute to reading ability (Gathercole et al., 1991; Wagner et al., 1997; Wolf & Bowers, 1999), and reading progress is better explained by a model which treats them as distinct skills, rather than a single phonological proficiency (Nelson et al., 2012). These findings indicate that the relationship between speech perception and phonological awareness/RAN should be investigated independently from pSTM. We believe that introducing perceptual ambiguity to a phonological awareness task (e.g., phoneme segmentation) would be a productive next step.

The present approach of conducting experimental studies with typical adults has obvious limitations for studying developmental language and reading disorders. However, we believe that this approach complements ongoing work with disordered groups. Simulating a perceptual deficit in adults with typical language skills allowed us to directly assess the impact of changing one system (perception) on another system (short-term memory). The ability to manipulate the levels of the proposed causal factor is critical to testing causal explanations, which is not possible when observing developmental trajectories, and not ethical if impairments in language or reading outcomes are a likely consequence of an experimental manipulation. We propose, therefore, that laboratory research on the interactions between different aspects of speech and language processing, and longitudinal developmental studies, can both make an important contribution. In the long term this combination of approaches can resolve the decades-long debate surrounding the nature of the "core deficit" or "deficits" responsible for phonological impairments in dyslexia and DLD.

Constraints on Generality

We would expect these results to generalize to other studies investigating recall of audio-morphed speech in a letter span task. We believe the ambiguity cost (the negative impact of perceptual ambiguity on short-term recall) would replicate in situations comparing high- and low-ambiguity stimuli, where ambiguity is created through naturalistic audio-morphing between sounds from either end of phonetic continua. We recommend that an identification pretest be conducted, to calibrate low-ambiguity and high-ambiguity stimuli to 90% and 99% identification accuracy, respectively. The ambiguity cost effect has here been observed for six- and eight-letter sequences, and may vary depending on sequence length (particularly if comparing subspan with supraspan sequences). However, other work from the current authors has identified the ambiguity cost for three-, five- and seven-letter sequences (Smith, 2023). We cannot predict if the ambiguity cost would occur if perceptual ambiguity were caused by more natural variations in speech sounds (e.g., unfamiliar accent). We would expect these results to generalize to a young adult population without preexisting language or hearing difficulties. We have no reason to believe that the results depend on other characteristics of the participants, materials, or context (Simons et al., 2017).

References

Ackerman, P. T., & Dykman, R. A. (1993). Phonological processes, confrontational naming, and immediate memory in dyslexia. *Journal of Learning Disabilities*, *26*(9), 597–609. https://doi.org/10.1177/002221949302600910

Ahissar, M. (2007). Dyslexia and the anchoring-deficit hypothesis. *Trends in cognitive sciences*, *11*(11), 458–465. https://doi.org/10.1016/j.tics.2007.08.015

Baddeley, A. D. (1968). How does acoustic similarity influence short-term memory? *The Quarterly Journal of Experimental Psychology*, *20*(3), 249–264. https://doi.org/10.1080/14640746808400159

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–89). Academic Press.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4* (R package Version 1.1-7) [Computer software]. https://cran.r-project.org/web/packages/lme4/index.html

Bentin, S., & Leshem, H. (1993). On the interaction between phonological awareness and reading acquisition: It's a two-way street. *Annals of Dyslexia*, *43*(1), 125–148. https://doi.org/10.1007/BF02928178

Bishop, D. V., North, T., & Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. *Journal of Child Psychology and Psychiatry*, *37*(4), 391–403. https://doi.org/10.1111/j.1469-7610.1996.tb01420.x

Bishop, D. V., & Snowling, M. J. (2004). Developmental dyslexia and specific language impairment: Same or different? *Psychological Bulletin*, *130*(6), 858–886. https://doi.org/10.1037/0033-2909.130.6.858

Bishop, D. V., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & Catalise-2 Consortium. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, *58*(10), 1068–1080. https://doi.org/10.1111/jcpp.12721

Boets, B. (2014). Dyslexia: Reconciling controversies within an integrative developmental perspective. *Trends in Cognitive Sciences*, *18*(10), 501–503. https://doi.org/10.1016/j.tics.2014.06.003

Bosse, M. L., Tainturier, M. J., & Valdois, S. (2007). Developmental dyslexia: The visual attention span deficit hypothesis. *Cognition*, *104*(2), 198–230. https://doi.org/10.1016/j.cognition.2006.05.009

Bowers, P. G., & Wolf, M. (1993). Theoretical links among naming speed, precise timing mechanisms and orthographic skill in dyslexia. *Reading and Writing*, *5*(1), 69–85. https://doi.org/10.1007/BF01026919

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read—A causal connection. *Nature*, *301*(5899), 419–421. https://doi.org/10.1038/301419a0

Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language*, *55*(4), 627–652. https://doi.org/10.1016/j.jml.2006.08.005

Campbell, J. I., & Thompson, V. A. (2012). Morepower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. *Behavior Research Methods*, *44*(4), 1255–1265. https://doi.org/10.3758/s13428-012-0186-0

Carroll, J. M., Snowling, M. J., Stevenson, J., & Hulme, C. (2003). The development of phonological awareness in preschool children. *Developmental Psychology*, *39*(5), 913–923. https://doi.org/10.1037/0012-1649.39.5.913

Casini, L., Burle, B., & Nguyen, N. (2009). Speech perception engages a general timer: Evidence from a divided attention word identification task. *Cognition*, *112*(2), 318–322. https://doi.org/10.1016/j.cognition.2009.04.005

Catts, H. W. (1989). Speech production deficits in developmental dyslexia. *Journal of Speech and Hearing Disorders*, *54*(3), 422–428. https://doi.org/10.1044/jshd.5403.422

Catts, H. W., Adlof, S. M., Hogan, T. P., & Weismer, S. E. (2005). Are specific language impairment and dyslexia distinct disorders? *Journal of Speech, Language, and Hearing Research*, *48*(6), 1378–1396. https://doi.org/10.1044/1092-4388(2005/096)

Claessen, M., & Leitão, S. (2012). Phonological representations in children with SLI. *Child Language Teaching and Therapy*, *28*(2), 211–223. https://doi.org/10.1177/0265659012436851

Clopper, C. G., Pisoni, D. B., & Tierney, A. T. (2006). Effects of open-set and closed-set task demands on spoken word recognition. *Journal of the American Academy of Audiology*, *17*(05), 331–349. https://doi.org/10.3766/jaaa.17.5.4

Coady, J. A., Evans, J. L., Mainela-Arnold, E., & Kluender, K. R. (2007). Children with specific language impairments perceive speech most categorically when tokens are natural and meaningful. *Journal of Speech, Language, and Hearing Research*, *50*(1), 41–57. https://doi.org/10.1044/1092-4388(2007/004)

Coady, J. A., Kluender, K. R., & Evans, J. L. (2005). Categorical perception of speech by children with specific language impairments. *Journal of Speech, Language, and Hearing Research*, *48*(4), 944–959. https://doi.org/10.1044/1092-4388(2005/065)

Colling, L. J. (2023). *bayesplay: The Bayes factor playground* (R Package Version 0.9.2) [Computer software]. https://cran.r-project.org/web/packages/bayesplay/index.html

Connine, C. M., & Clifton, C., Jr. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(2), 291–299. https://doi.org/10.1037//0096-1523.13.2.291

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry*, *42*(6), 741–748. https://doi.org/10.1111/1469-7610.00770

Cowan, N., Hogan, T. P., Alt, M., Green, S., Cabbage, K. L., Brinkley, S., & Gray, S. (2017). Short-term memory in childhood dyslexia: Deficient serial order in multiple modalities. *Dyslexia*, *23*(3), 209–233. https://doi.org/10.1002/dys.1557

Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, *26*(2), 145–171. https://doi.org/10.1006/jpho.1998.0070

Damerau, F. J. (1964). A technique for computer detection and correction of spelling errors. *Communications of the ACM*, *7*(3), 171–176. https://doi.org/10.1145/363958.363994

De Groot, B. J., Van den Bos, K. P., Van der Meulen, B. F., & 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*(5), 1538–1548. https://doi.org/10.1044/2015_JSLHR-L-14-0019

De Leeuw, J. R. (2015). Jspsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Denckla, M. B., & Rudel, R. G. (1976). Rapid 'automatized' naming (R.A.N.): Dyslexia differentiated from other learning disabilities. *Neuropsychologia*, *14*(4), 471–479. https://doi.org/10.1016/0028-3932(76)90075-0

Dickinson, D. K., McCabe, A., Anastasopoulos, L., Peisner-Feinberg, E. S., & Poe, M. D. (2003). The comprehensive language approach to early literacy: The interrelationships among vocabulary, phonological sensitivity, and print knowledge among preschool-aged children. *Journal of Educational Psychology*, *95*(3), 465–481. https://doi.org/10.1037/0022-0663.95.3.465

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*, Article 781. https://doi.org/10.3389/fpsyg.2014.00781

Dienes, Z. (2021). How to use and report Bayesian hypothesis tests. *Psychology of Consciousness: Theory, Research, and Practice*, *8*(1), 9–26. https://doi.org/10.1037/cns0000258

Di Filippo, G., Zoccolotti, P., & Ziegler, J. C. (2008). Rapid naming deficits in dyslexia: A stumbling block for the perceptual anchor theory of dyslexia. *Developmental Science*, *11*(6), F40–F47. https://doi.org/10.1111/j.1467-7687.2008.00752.x

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, *41*(5), 1136–1146. https://doi.org/10.1044/jslhr.4105.1136

Dryden, A., Allen, H. A., Henshaw, H., & Heinrich, A. (2017). The association between cognitive performance and speech-in-noise perception for adult listeners: A systematic literature review and meta-analysis. *Trends in Hearing*, *21*, Article 233121651774467. https://doi.org/10.1177/2331216517744675

Ehrhorn, A. M., Adlof, S. M., Fogerty, D., & Laing, S. (2021). Probing phonological processing differences in nonword repetition for children with separate or co-occurring dyslexia and developmental language disorder. *Scientific Studies of Reading*, *25*(6), 486–503. https://doi.org/10.1080/10888438.2020.1849223

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, *67*(2), 224–238. https://doi.org/10.3758/BF03206487

Elbro, C., & Jensen, M. N. (2005). Quality of phonological representations, verbal learning, and phoneme awareness in dyslexic and normal readers. *Scandinavian Journal of Psychology*, *46*(4), 375–384. https://doi.org/10.1111/j.1467-9450.2005.00468.x

Elbro, C., Nielsen, I., & Petersen, D. K. (1994). Dyslexia in adults: Evidence for deficits in non-word reading and in the phonological representation of lexical items. *Annals of Dyslexia*, *44*(1), 203–226. https://doi.org/10.1007/BF02648162

Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, *81*(1–3), 162–173. https://doi.org/10.1006/brln.2001.2514

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fawcett, A. J., & Nicolson, R. I. (1995). Persistence of phonological awareness deficits in older children with dyslexia. *Reading and Writing*, *7*(4), 361–376. https://doi.org/10.1007/BF01027724

Füllgrabe, C., & Rosen, S. (2016). On the (un)importance of working memory in speech-in-noise processing for listeners with normal hearing thresholds. *Frontiers in Psychology*, *7*, Article 1268. https://doi.org/10.3389/fpsyg.2016.01268

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110–125. https://doi.org/10.1037/0096-1523.6.1.110

Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language*, *29*(3), 336–360. https://doi.org/10.1016/0749-596X(90)90004-J

Gathercole, S. E., Willis, C., & Baddeley, A. D. (1991). Differentiating phonological memory and awareness of rhyme: Reading and vocabulary development in children. *British Journal of Psychology*, *82*(3), 387–406. https://doi.org/10.1111/j.2044-8295.1991.tb02407.x

Godfrey, J. J., Syrdal-Lasky, A. K., Millay, K. K., & Knox, C. M. (1981). Performance of dyslexic children on speech perception tests. *Journal of Experimental Child Psychology*, *32*(3), 401–424. https://doi.org/10.1016/0022-0965(81)90105-3

Gordon-Salant, S., & Cole, S. S. (2016). Effects of age and working memory capacity on speech recognition performance in noise among listeners with normal hearing. *Ear & Hearing*, *37*(5), 593–602. https://doi.org/10.1097/AUD.0000000000000316

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences*, *15*(1), 3–10. https://doi.org/10.1016/j.tics.2010.10.001

Guang, C., Lefkowitz, E., Dillman-Hasso, N., Brown, V., & Strand, J. (2020). Recall of speech is impaired by subsequent masking noise: A replication of Rabbitt (1968) experiment 2. *Auditory Perception & Cognition*, *3*(3), 158–167. https://doi.org/10.1080/25742442.2021.1896908

Gwilliams, L., Linzen, T., Poeppel, D., & Marantz, A. (2018). In spoken word recognition, the future predicts the past. *Journal of Neuroscience*, *38*(35), 7585–7599. https://doi.org/10.1523/JNEUROSCI.0065-18.2018

Hazan, V., Messaoud-Galusi, S., Rosen, S., Nouwens, S., & 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*(6), 1510–1529. https://doi.org/10.1044/1092-4388(2009/08-0220)

Heald, S., & Nusbaum, H. C. (2014). Speech perception as an active cognitive process. *Frontiers in Systems Neuroscience*, *8*, Article 35. https://doi.org/10.3389/fnsys.2014.00035

Heilman, K. M., Voeller, K., & Alexander, A. W. (1996). Developmental dyslexia: A motor-articulatory feedback hypothesis. *Annals of Neurology*, *39*(3), 407–412. https://doi.org/10.1002/ana.410390323

Hogan, T. P. (2010). A short report: Word-level phonological and lexical characteristics interact to influence phoneme awareness. *Journal of Learning Disabilities*, *43*(4), 346–356. https://doi.org/10.1177/0022219410369083

Hufnagle, D. G., Holt, L. L., & Thiessen, E. D. (2013). Spectral information in nonspeech contexts influences children's categorization of ambiguous speech sounds. *Journal of Experimental Child Psychology*, *116*(3), 728–737. https://doi.org/10.1016/j.jecp.2013.05.008

Hughes, R., & Jones, D. M. (2001). The intrusiveness of sound: Laboratory findings and their implications for noise abatement. *Noise & health*, *4*(13), 51–121.

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin*, *140*(2), 339–373. https://doi.org/10.1037/a0034221

Izura, C., & Playfoot, D. (2012). A normative study of acronyms and acronym naming. *Behavior Research Methods*, *44*(3), 862–889. https://doi.org/10.3758/s13428-011-0175-8

Joanisse, M. F., Manis, F. R., Keating, P., & Seidenberg, M. S. (2000). Language deficits in dyslexic children: Speech perception, phonology, and morphology. *Journal of Experimental Child Psychology*, *77*(1), 30–60. https://doi.org/10.1006/jecp.1999.2553

Joanisse, M. F., & Seidenberg, M. S. (2003). Phonology and syntax in specific language impairment: Evidence from a connectionist model. *Brain and Language*, *86*(1), 40–56. https://doi.org/10.1016/S0093-934X(02)00533-3

Jones, S. D., & Westermann, G. (2022). Under-resourced or overloaded? Rethinking working memory deficits in developmental language disorder. *Psychological Review*, *129*(6), 1358–1372. https://doi.org/10.1037/rev0000338

Kamhi, A. G., & Catts, H. W. (1986). Toward an understanding of developmental language and reading disorders. *Journal of Speech and Hearing Disorders*, *51*(4), 337–347. https://doi.org/10.1044/jshd.5104.337

Katz, W. F., Curtiss, S., & Tallal, P. (1992). Rapid automatized naming and gesture by normal and language-impaired children. *Brain and Language*, *43*(4), 623–641. https://doi.org/10.1016/0093-934X(92)90087-U

Kawahara, H., Masuda-Katsuse, I., & De Cheveigne, A. (1999). Restructuring speech representations using a pitch-adaptive time–frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds. *Speech Communication*, *27*(3–4), 187–207. https://doi.org/10.1016/S0167-6393(98)00085-5

Kawahara, H., & Morise, M. (2011). Technical foundations of TANDEM-STRAIGHT, a speech analysis, modification and synthesis framework. *Sadhana*, *36*(5), 713–727. https://doi.org/10.1007/s12046-011-0043-3

Kazak, A. E. (2018). Editorial: Journal article reporting standards. *American Psychologist*, *73*(1), 1–2. https://doi.org/10.1037/amp0000263

Ladefoged, P. (1975). *A course in phonetics*. Harcourt Brace Jovanovich.

Lange, K., Kühn, S., & Filevich, E. (2015). "Just Another Tool for Online Studies" (JATOS): An easy solution for setup and management of web servers supporting online studies. *PLoS ONE*, *10*(6), Article e0130834. https://doi.org/10.1371/journal.pone.0130834

Laszlo, S., & Federmeier, K. D. (2007). The acronym superiority effect. *Psychonomic Bulletin & Review*, *14*(6), 1158–1163. https://doi.org/10.3758/BF03193106

Laughery, K. R., & Pinkus, A. L. (1966). Short-term memory: Effects of acoustic similarity, presentation rate and presentation mode. *Psychonomic Science*, *6*(6), 285–286. https://doi.org/10.3758/BF03328069

Lee, M. D., & Wagenmakers, E. J. (2014). *Bayesian Cognitive modeling: A practical course*. Cambridge University Press.

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, *69*(1), 1–33. https://doi.org/10.18637/jss.v069.i01

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, *10*(8), 707–710.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368. https://doi.org/10.1037/h0044417

Luce, P. A., Feustel, T. C., & Pisoni, D. B. (1983). Capacity demands in short-term memory for synthetic and natural speech. *Human Factors*, *25*(1), 17–32. https://doi.org/10.1177/001872088302500102

Magimairaj, B. M., Nagaraj, N. K., & Benafield, N. J. (2018). Children's speech perception in noise: Evidence for dissociation from language and working memory. *Journal of Speech, Language, and Hearing Research*, *61*(5), 1294–1305. https://doi.org/10.1044/2018_JSLHR-H-17-0312

Majerus, S., Amand, P., Boniver, V., Demanez, J.-P., Demanez, L., & Van der Linden, M. (2005). A quantitative and qualitative assessment of verbal short-term memory and phonological processing in 8-year-olds with a history of repetitive otitis media. *Journal of Communication Disorders*, *38*(6), 473–498. https://doi.org/10.1016/j.jcomdis.2005.04.002

Manis, F. R., McBride-Chang, C., Seidenberg, M. S., Keating, P., Doi, L. M., Munson, B., & Petersen, A. (1997). Are speech perception deficits associated with developmental dyslexia? *Journal of Experimental Child Psychology*, *66*(2), 211–235. https://doi.org/10.1006/jecp.1997.2383

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*(7–8), 953–978. https://doi.org/10.1080/01690965.2012.705006

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences*, *10*(8), 363–369. https://doi.org/10.1016/j.tics.2006.06.007

McCreery, R. W., Spratford, M., Kirby, B., & Brennan, M. (2017). Individual differences in language and working memory affect children's speech recognition in noise. *International Journal of Audiology*, *56*(5), 306–315. https://doi.org/10.1080/14992027.2016.1266703

McQueen, J. M., Tyler, M. D., & Cutler, A. (2012). Lexical retuning of children's speech perception: Evidence for knowledge about words' component sounds. *Language Learning and Development*, *8*(4), 317–339. https://doi.org/10.1080/15475441.2011.641887

Melby-Lervåg, M., & Lervåg, A. (2012). Oral language skills moderate nonword repetition skills in children with dyslexia: A meta-analysis of the role of nonword repetition skills in dyslexia. *Scientific Studies of Reading*, *16*(1), 1–34. https://doi.org/10.1080/10888438.2010.537715

Messaoud-Galusi, S., Hazan, V., & Rosen, S. (2007, August 6–10). *Clarifying the speech perception deficit in dyslexic children* [Conference session] Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken.

Moreno, R., & Martínez, R. J. (2008). Causality as validity: Some implications for the social sciences. *Quality & Quantity*, *42*(5), 597–604. https://doi.org/10.1007/s11135-006-9060-4

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Morton, J., & Chambers, S. M. (1976). Some evidence for 'speech' as an acoustic feature. *British Journal of Psychology*, *67*(1), 31–45. https://doi.org/10.1111/bjop.1976.67.issue-1

Nelson, J. M., Lindstrom, J. H., Lindstrom, W., & Denis, D. (2012). The structure of phonological processing and its relationship to basic reading. *Exceptionality*, *20*(3), 179–196. https://doi.org/10.1080/09362835.2012.694612

Nicolson, R. I., & Fawcett, A. J. (2007). Procedural learning difficulties: Reuniting the developmental disorders? *TRENDS in Neurosciences*, *30*(4), 135–141. https://doi.org/10.1016/j.tins.2007.02.003

Nicolson, R. I., & Fawcett, A. J. (2011). Dyslexia, dysgraphia, procedural learning and the cerebellum. *Cortex*, *47*(1), 117–127. https://doi.org/10.1016/j.cortex.2009.08.016

Noordenbos, M. W., & Serniclaes, W. (2015). The categorical perception deficit in dyslexia: A meta-analysis. *Scientific Studies of Reading*, *19*(5), 340–359. https://doi.org/10.1080/10888438.2015.1052455

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *Journal of Child Psychology and Psychiatry*, *57*(11), 1247–1257. https://doi.org/10.1111/jcpp.12573

Obleser, J., & Kotz, S. A. (2011). Multiple brain signatures of integration in the comprehension of degraded speech. *Neuroimage*, *55*(2), 713–723. https://doi.org/10.1016/j.neuroimage.2010.12.020

Oganian, Y., & Ahissar, M. (2012). Poor anchoring limits dyslexics' perceptual, memory, and reading skills. *Neuropsychologia*, *50*(8), 1895–1905. https://doi.org/10.1016/j.neuropsychologia.2012.04.014

Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and Hearing*, *39*(2), 204–214. https://doi.org/10.1097/AUD.0000000000000494

Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition*, *101*(2), 385–413. https://doi.org/10.1016/j.cognition.2006.04.008

Pennington, B. F., Van Orden, G. C., Smith, S. D., Green, P. A., & Haith, M. M. (1990). Phonological processing skills and deficits in adult dyslexics. *Child Development*, *61*(6), 1753–1778. https://doi.org/10.2307/1130836

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing*, *37*(1), 5S–27S. https://doi.org/10.1097/AUD.0000000000000312

Rabbitt, P. M. (1968). Channel-capacity, intelligibility and immediate memory. *The Quarterly Journal of Experimental Psychology*, *20*(3), 241–248. https://doi.org/10.1080/14640746808400158

Ramus, F., & Ahissar, M. (2012). Developmental dyslexia: The difficulties of interpreting poor performance, and the importance of normal performance. *Cognitive Neuropsychology*, *29*(1–2), 104–122. https://doi.org/10.1080/02643294.2012.677420

Ramus, F., Marshall, C. R., Rosen, S., & Van Der Lely, H. K. (2013). Phonological deficits in specific language impairment and developmental dyslexia: Towards a multidimensional model. *Brain*, *136*(2), 630–645. https://doi.org/10.1093/brain/aws356

Ramus, F., & Szenkovits, G. (2008). What phonological deficit? *Quarterly Journal of Experimental Psychology*, *61*(1), 129–141. https://doi.org/10.1080/17470210701508822

R Core Team. (2022). *R: A language and environment for statistical computing* (Version 4.1.3) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Renz, T., Leistner, P., & Liebl, A. (2018). Auditory distraction by speech: Sound masking with speech-shaped stationary noise outperforms −5 dB per octave shaped noise. *The Journal of the Acoustical Society of America*, *143*(3), EL212–EL217. https://doi.org/10.1121/1.5027765

Robertson, E. K., Joanisse, M. F., Desroches, A. S., & Ng, S. (2009). Categorical speech perception deficits distinguish language and reading impairments in children. *Developmental Science*, *12*(5), 753–767. https://doi.org/10.1111/j.1467-7687.2009.00806.x

Rogers, J. C., & Davis, M. H. (2017). Inferior frontal cortex contributions to the recognition of spoken words and their constituent speech sounds. *Journal of Cognitive Neuroscience*, *29*(5), 919–936. https://doi.org/10.1162/jocn_a_01096

Roman, A. S., Pisoni, D. B., Kronenberger, W. G., & Faulkner, K. F. (2017). Some neurocognitive correlates of noise-vocoded speech perception in children with normal hearing: A replication and extension of. *Ear and Hearing*, *38*(3), 344–356. https://doi.org/10.1097/AUD.0000000000000393

Rönnberg, J., Rudner, M., Lunner, T., & Zekveld, A. A. (2010). When cognition kicks in: Working memory and speech understanding in noise. *Noise and Health*, *12*(49), 263–299. https://doi.org/10.4103/1463-1741.70505

Rönnberg, N., Rudner, M., Lunner, T., & Stenfelt, S. (2014). Assessing listening effort by measuring short-term memory storage and processing of speech in noise. *Speech, Language and Hearing*, *17*(3), 123–132. https://doi.org/10.1179/2050572813Y.0000000033

Rowe, E. J., & Rowe, W. G. (1976). Stimulus suffix effects with speech and nonspeech sounds. *Memory & Cognition*, *4*(2), 128–131. https://doi.org/10.3758/BF03213153

Saltzman, D., Luthra, S., Myers, E. B., & Magnuson, J. S. (2021). Attention, task demands, and multitalker processing costs in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *47*(12), 1673–1680. https://doi.org/10.1037/xhp0000963

Schlittmeier, S. J., Weißgerber, T., Kerber, S., Fastl, H., & Hellbrück, J. (2012). Algorithmic modeling of the irrelevant sound effect (ISE) by the hearing sensation fluctuation strength. *Attention, Perception, & Psychophysics*, *74*(1), 194–203. https://doi.org/10.3758/s13414-011-0230-7

Scott, S. (1993). *P-centres in speech-an acoustic analysis* [PhD thesis]. University of College London.

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, *12*(6), 1123–1128. https://doi.org/10.1177/1745691617708630

Smith, H. J. (2023). *Evaluating perceptual deficits as a causal explanation of phonological dysfunction: Behavioural and neural evidence* [Doctoral dissertation]. University of Cambridge.

Smither, J. A. A. (1993). Short term memory demands in processing synthetic speech by old and young adults. *Behaviour & Information Technology*, *12*(6), 330–335. https://doi.org/10.1080/01449299308924397

Smith-Spark, J. H., & Fisk, J. E. (2007). Working memory functioning in developmental dyslexia. *Memory*, *15*(1), 34–56. https://doi.org/10.1080/09658210601043384

Snowling, M. (1998). Dyslexia as a phonological deficit: Evidence and implications. *Child Psychology and Psychiatry Review*, *3*(1), 4–11. https://doi.org/10.1017/S1360641797001366

Snowling, M. J., Lervåg, A., Nash, H. M., & Hulme, C. (2019). Longitudinal relationships between speech perception, phonological skills and reading in children at high-risk of dyslexia. *Developmental Science*, *22*(1), Article e12723. https://doi.org/10.1111/desc.12723

Snowling, M. J., Nash, H. M., Gooch, D. C., Hayiou-Thomas, M. E., Hulme, C., & Wellcome Language and Reading Project Team. (2019). Developmental outcomes for children at high risk of dyslexia and children with developmental language disorder. *Child Development*, *90*(5), e548–e564. https://doi.org/10.1111/cdev.13216

Stanovich, K. E. (1986). Cognitive processes and the reading problems of learning disabled children: Evaluating the assumption of specificity. In J. Torgesen & B. Wong (Eds.), *Psychological and educational perspectives on learning disabilities* (pp. 87–131). Academic Press.

Stenbäck, V., Hällgren, M., Lyxell, B., & Larsby, B. (2015). The Swedish Hayling task, and its relation to working memory, verbal ability, and speech-recognition-in-noise. *Scandinavian Journal of Psychology*, *56*(3), 264–272. https://doi.org/10.1111/sjop.12206

Surprenant, A. M. (1999). The effect of noise on memory for spoken syllables. *International Journal of Psychology*, *34*(5–6), 328–333. https://doi.org/10.1080/002075999399648

Swan, D., & Goswami, U. (1997). Phonological awareness deficits in developmental dyslexia and the phonological representations hypothesis. *Journal of Experimental Child Psychology*, *66*(1), 18–41. https://doi.org/10.1006/jecp.1997.2375

Szenkovits, G., & Ramus, F. (2005). Exploring dyslexics' phonological deficit I: Lexical versus sub-lexical and input vs output processes. *Dyslexia*, *11*(4), 253–268. https://doi.org/10.1002/dys.308

Tallal, P., Allard, L., Miller, S., & Curtiss, S. (1997). Academic outcomes of language impaired children. In C. Hulme & M. J. Snowling (Eds.), *Dyslexia: Biology, cognition and intervention* (pp. 167–181). Whurr.

Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex*, *41*(3), 399–433. https://doi.org/10.1016/S0010-9452(08)70276-4

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, *45*(1), 2–40. https://doi.org/10.1046/j.0021-9630.2003.00305.x

Veneman, C. E., Gordon-Salant, S., Matthews, L. J., & Dubno, J. R. (2013). Age and measurement time-of-day effects on speech recognition in noise. *Ear and Hearing*, *34*(3), 288–299. https://doi.org/10.1097/AUD.0b013e31826d0b81

Vidyasagar, T. R., & Pammer, K. (2010). Dyslexia: A deficit in visuo-spatial attention, not in phonological processing. *Trends in Cognitive Sciences*, *14*(2), 57–63. https://doi.org/10.1016/j.tics.2009.12.003

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, *101*(2), 192–212. https://doi.org/10.1037/0033-2909.101.2.192

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., Donahue, J., & Garon, T. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology*, *33*(3), 468–479. https://doi.org/10.1037/0012-1649.33.3.468

Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology*, *91*(3), 415–438. https://doi.org/10.1037/0022-0663.91.3.415

Yang, L., Li, C., Li, X., Zhai, M., An, Q., Zhang, Y., Zhao, J., & Weng, X. (2022). Prevalence of developmental dyslexia in primary school children: A systematic review and meta-analysis. *Brain Sciences*, *12*(2), Article 240. https://doi.org/10.3390/brainsci12020240

Yu, T. J., & Schlauch, R. S. (2019). Diagnostic precision of open-set versus closed-set word recognition testing. *Journal of Speech, Language, and Hearing Research*, *62*(6), 2035–2047. https://doi.org/10.1044/2019_JSLHR-H-18-0317

Ziegler, J. C. (2008). Better to lose the anchor than the whole ship. *Trends in Cognitive Sciences*, *12*(7), 244–245. https://doi.org/10.1016/j.tics.2008.04.001

Ziegler, J. C., Pech-Georgel, C., George, F., Alario, F. X., & Lorenzi, C. (2005). Deficits in speech perception predict language learning impairment. *Proceedings of the National Academy of Sciences*, *102*(39), 14110–14115. https://doi.org/10.1073/pnas.0504446102

Received February 28, 2023
Revision received September 25, 2023
Accepted October 9, 2023 ■