



# Phonotactic and lexical factors in talker discrimination and identification

Sandy Abu El Adas<sup>1</sup> · Susannah V. Levi<sup>1</sup>

Accepted: 25 March 2022 / Published online: 31 May 2022  
© The Psychonomic Society, Inc. 2022

## Abstract

Previous research has shown that listeners are better at processing talker information in their native language compared to an unfamiliar language, a phenomenon known as the language familiarity effect. Several studies have explored two mechanisms that support this effect: lexical status and phonological familiarity. Further support for the importance of phonological knowledge comes from studies showing that participants with poorer reading skills perform worse on talker processing tasks. Previous research also suggested that speech perception in individuals with poor reading skills may be task dependent, with poorer performance on identification tasks compared to discrimination tasks. In the current study, we explore talker perception while manipulating lexicality (words, nonwords) and phonotactic probability (high, low) in participants who differ in reading ability and phonological working memory using a talker discrimination task ([Experiment 1](#)) and a talker identification task ([Experiment 2](#)). Results from these experiments revealed an effect of lexical status and phonotactic probability in both the discrimination and the identification tasks. Effects of phonological working memory were found only for the identification task, where participants with higher scores identified more talkers correctly. These results suggest that listeners use both phonological and lexical information when processing talker information. The task-modulated results show that listeners with poorer phonological working memory perform worse on talker identification tasks that tap into long-term memory representations, but not on discrimination tasks that can be completed with more peripheral processing. This may suggest a more general link between phonological working memory and learning talker categories.

**Keywords** Speech perception · Talker processing · Reading ability · Phonological processing

## Introduction

The speech signal carries two types of information: linguistic and talker (often referred to as indexical information). Linguistic information (e.g., phonetics, phonology, morphology) provides information about the message the speaker is trying to convey (*what* is the speaker saying), while talker information provides information about the specific individual (*who* is talking) (Abercrombie, 1967). These two types of information interact bidirectionally during speech perception. Studies show that talker information can affect linguistic processing, such as better word and sentence recognition for familiar talkers (Nygaard & Pisoni, 1998) and faster responses in a speeded phoneme-identification for single- compared to

multiple-talker conditions (Mullennix & Pisoni, 1990). Similarly, linguistic information can affect talker processing, as in the language familiarity effect, where listeners are better at processing information about a talker when listening to speech in their native language compared to an unfamiliar language (for review, see Levi, 2019, and Perrachione, 2019). In the current study, we expand on research that has attempted to uncover the mechanisms that generate the language familiarity effect by simultaneously examining the contribution of lexical status (words vs. nonwords), phonological information (phonotactic probability), and individual differences in either reading ability or phonological working memory while probing both a talker discrimination task and a talker identification task.

✉ Sandy Abu El Adas  
sandyabu@nyu.edu

<sup>1</sup> Department of Communicative Sciences and Disorders, New York University, New York, NY, USA

## Possible mechanisms supporting the language familiarity effect

Several studies have explored the underlying mechanisms that result in better perception of talker information in a listener's

native language and have focused on two possible sources: lexical status and phonological familiarity. To test the contribution of these two types of information, researchers have manipulated various aspects of the stimuli or of the listeners in an attempt to isolate the contributions of these two types of information.

### Controlling for phonological information in the stimuli

Three main types of stimulus manipulations have been used to examine the effects of phonological information: (1) comparing performance in languages that are either phonologically similar or phonologically distinct (Köster & Schiller, 1997; Winters et al., 2008; Zarate et al., 2015); (2) using reversed speech (Fleming et al., 2014; Perrachione et al., 2015); and (3) using phonologically legal nonwords (Perrachione et al., 2015; Xie & Myers, 2015; Zarate et al., 2015). All of these manipulations keep some phonological information while eliminating the involvement of lexical information. The third type of manipulation – using phonologically legal nonwords – provides the best control of phonological familiarity, as nonwords are phonologically legal strings of segments that carry no lexical representation. For this reason, we focus our description below on studies that have used this manipulation.

Studies examining the perception of talker information using nonword stimuli have used a talker-learning task followed by a talker-identification task. In this paradigm, listeners first complete a training phase in which they hear speech, identify the talker, and receive feedback. Following this training, listeners complete a test block without feedback. In Zarate et al. (2015), listeners were provided with sets of disyllabic English words or nonwords that were created by rearranging the syllables from the real-word condition. Zarate et al. (2015) found no difference between the word and nonword conditions, which the authors interpreted as evidence that phonological familiarity supports the language familiarity effect.

In contrast to Zarate et al. (2015), Perrachione et al. (2015) used sentences containing nonwords that were matched to the real word sentences in terms of biphone probability. Unlike Zarate et al. (2015), Perrachione et al. (2015) did find that listeners were more accurate at identifying talkers in the real-word condition than in the nonword condition. This same pattern of results was also found in Xie and Myers (2015), who found better performance for English sentences compared with “Jabberwocky” sentences composed of a combination of real words and nonwords. Unlike Perrachione et al. (2015), the nonword sentences in Xie and Myers (2015) contained some lexical information such as function words (e.g., “More in a tri-campic lingting turress angra the fortune”). The results from these latter two studies suggest that phonological familiarity alone is not sufficient to generate optimal talker processing and that lexical information also contributes to the language familiarity effect. In addition to the nonword manipulation, all

three of these studies included a condition in which listeners performed the same type of task with Mandarin stimuli. In all studies, performance on the Mandarin condition was poorer than on the nonword condition, suggesting that among stimuli that lack lexical representations, phonological familiarity improves performance.

Several possible explanations can account for the disparity between Zarate et al. (2015) and the two other studies mentioned above. One possibility is that listeners in Zarate et al. (2015) were near ceiling (words ~85%, nonwords ~80%), which may have narrowed the differences between the conditions. This high performance may be due to the decision to use the same talkers across conditions; indeed, previous research has shown that the talkers can be identified and discriminated as the same across languages (Wester, 2012; Winters et al., 2008). Second, Zarate et al. (2015) used concatenated sets of two-syllable words or nonwords produced in isolation (e.g., “assume holy heaven deny”) that did not carry prosodic contours, while Perrachione et al. (2015) and Xie and Myers (2015) used sentence-length stimuli that presumably carried prosodic contours (e.g., “The harder he tried the less he got done”). It is unclear how natural the prosodic contours were for the nonsense sentences, though Zarate et al. (2015) intimated that prosody may help listeners in differentiating between talkers across languages. Thus, one possible explanation for the reduced performance in the nonsense sentences is that talkers were not producing natural or consistent prosodic contours that would have supported learning talker categories. Nevertheless, these studies did report that the sentences sounded natural. Related, previous studies have argued that stimuli of different lengths (words vs. sentences) allow listeners to tap into different levels of processing, such as segmental versus prosodic contours (Nygaard & Pisoni, 1998). For these reasons, in the current study, we use monosyllabic words and nonwords as a way to reduce reliance on prosodic information and to control for a naturalness in the production of these tokens.

One type of stimulus manipulation that has not been investigated in previous research on talker processing is phonotactic probability (the probability that two sounds occur as a sequence in a particular word position). It is well established that phonotactic probability affects performance on linguistic and non-linguistic processing tasks (for review, see Buchwald, 2014, and Vitevitch & Luce, 2016). Influences of phonotactic probability have been reported in lexical decision tasks (e.g., Experiment 3 in Vitevitch & Luce, 1999), voice change detection paradigms (e.g., Vitevitch & Donoso, 2011), word recall (e.g., Thorn & Frankish, 2005), word/nonword matching tasks (e.g., Vitevitch & Luce, 1999), speech production (e.g., Vitevitch et al., 2004), and word learning (e.g., Storkel et al., 2006). For example, Vitevitch and Luce (1999) found that listeners responded slower on words and nonwords with high phonotactic probability compared to

words and nonwords with low phonotactic probability using an auditory lexical decision task.

In the talker processing literature, no study has explicitly manipulated phonotactic probability to test its effect on talker processing. Perrachione et al. (2015) controlled mean phonotactic probability across the word and nonword conditions, but the contribution of phonotactic probability was not directly tested. Considering the effects of phonotactic probability on linguistic processing and on voice-change detection, the current study takes the additional step of probing the contribution of phonological familiarity as a mechanism that supports the language familiarity effect by carefully manipulating the phonotactic probability of our stimuli.

Overall, the studies that examined the contribution of lexical information and phonological information on talker processing suggest that both support the perception of talker information. Results showing that listeners perform better on talker processing tasks when listening to real words compared to phonotactically legal nonsense words point to the importance of lexicality. Differences found across types of meaningless utterances (such as better performance in phonologically similar vs. dissimilar languages, better performance for legal nonwords vs. a foreign language) suggests that that phonological familiarity also plays a role. The current study contributes to our understanding of what mechanisms support the language familiarity effect by explicitly controlling phonotactic probability. This will allow us to test whether phonological familiarity incrementally affects how listeners perform on talker processing tasks.

### Controlling for phonological processing ability in listeners

Several studies have explored the connection between reading ability and talker processing as another way to examine the role of phonological versus lexical information (Kadam et al., 2016; Perea et al., 2014; Perrachione et al., 2011). Previous studies have demonstrated that reading ability is tied to phonological processing skills and phonological representations (Boets et al., 2011; Bradley & Bryant, 1983; Wagner & Torgesen, 1987; Werker & Tees, 1987; Ziegler & Goswami, 2005). In speech perception, studies using phoneme discrimination and identification tasks often show that poorer readers exhibit shallower categorization curves (Mody et al., 1997; Werker & Tees, 1987) and less consistency of responses to endpoint stimuli (O'Brien et al., 2018), both of which implicate poor phonological representations. However, not all studies have found differences in speech perception in individuals with poorer reading (Hazan et al., 2009; Manis et al., 1997). For example, Hazan et al. (2009) reported that individuals with dyslexia did not perform differently than average readers on a range of speech perception tasks including phoneme identification, phoneme discrimination, and word perception, suggesting that there may be more subtle differences (see also

the section below on the possible contribution of task effects modulating speech perception skills). A study by Manis et al. (1997) found that only a subset (7/25) of children with dyslexia showed poor phoneme identification, suggesting that a minority of children with dyslexia demonstrate poorer speech perception skills.

Because of studies that have found differences in phonological processing in better versus poorer readers, several studies have examined talker processing in people with a range of reading abilities to assess the contribution of the phonology in a non-linguistic task. In addition to improving our understanding of the underlying mechanisms involved in the language familiarity effect, studying talker processing in individuals with poorer reading skills may provide a better understanding of the disorder and provide essential information about how non-linguistic information, such as talker information, is processed in this population.

Perrachione et al. (2011) tested English-speaking individuals with and without dyslexia in a talker identification task in English and Mandarin. Listeners were familiarized with a set of five talkers and then tested on their ability to identify them. The adults with typical reading skills showed the expected language familiarity effect, while the adults with dyslexia were significantly impaired in their ability to recognize talkers in their *native* language and failed to show a language familiarity effect. In addition, they found that performance in the English condition correlated with performance on two different phonological awareness tasks: nonword repetition and elision. Rather than examining processing in individuals with a reading disorder, Kadam et al. (2016) examined the relationship between reading skills and talker processing with unimpaired readers that were median-split into two groups. They used a similar type of training and identification paradigm with talkers producing English and French sentences. They found that better readers learned to identify talker's voices faster than average readers in both language conditions. Though not directly tested, the figures suggest that both the better and the average readers demonstrate a language familiarity effect. Because both studies found that listeners with higher reading skills perform better on talker identification tasks than poorer readers, they argued that the ability to process talker information relies on phonological processing ability.

A study with Spanish-speaking listeners using Spanish and Mandarin stimuli failed to replicate a lack of a language familiarity effect in individuals with dyslexia (Perea et al., 2014). Perea et al. (2014) tested children and adults with and without dyslexia on a talker identification task. Similar to Perrachione et al. (2011), results revealed better performance in both children and adults without dyslexia compared to the children and adults with dyslexia. However, unlike Perrachione et al. (2011), the children and adults with dyslexia *did* exhibit a language familiarity effect with better

performance in Spanish than English. Perea et al. (2014) suggest that differences in the pattern of results may stem from the languages used: Spanish is more orthographically transparent than English. Despite differences in the pattern of results, it is important to note that all three studies support the idea that poor phonological representations in individuals with dyslexia result in some type of poorer performance on a non-linguistic, talker processing task.

### Task effects

These three studies that examined reading ability and the language familiarity effect all used an identification task that requires listeners to store information about a talker's voice. Previous research on the language familiarity effect suggests that the type of task used to probe how listeners perceive information about a talker also affects performance (for review, see Levi, 2019). For example, Winters et al. (2008) found task-dependent effects where listeners exhibited the language familiarity effect in a discrimination task, but not in an identification task with an extensive training period. Fecher and Johnson (2018) found the opposite pattern of effects with adults and children, in which the language familiarity effect was found in a talker identification task but not in a discrimination task.

In addition, studies of linguistic speech perception skills in individuals with reading impairments have also demonstrated some evidence of task effects. Whereas there is consistency across studies showing poorer performance on phoneme identification tasks that require listeners to tap into long-term representations of speech sounds (Lyon et al., 2003; Wagner & Torgesen, 1987), there is variability in discrimination tasks (Serniclaes et al., 2001, 2004). Some studies have reported good or heightened discrimination ability while simultaneously reporting poorer identification in individuals with reading impairments (Serniclaes et al., 2001, 2004). Serniclaes et al. (2001) argued that heightened discrimination performance in individuals with poor reading is due to increased sensitivity to phonologically irrelevant information, and that poor performance on identification tasks illustrates that the main deficit is in labelling speech sounds into categories, what they call the *allophonic theory*. In other words, phonological representations are not properly developed due to an inability to effectively map speech sound into phonemes. Given these task-modulated effects, the current study expands on existing literature examining the effects of reading on talker processing using both a discrimination task and an identification task.

### Current study

The purpose of the current study is to bring together these lines of research, which have examined what type of linguistic

information supports the perception of talker information in a listener's native language. To this end, we will compare performance for word and nonword stimuli and examine individual differences in reading ability and phonological working memory. The current study also adds the additional stimulus manipulation to test phonological familiarity at a more fine-grained level by varying phonotactic probability. Finally, the study probes the relationship between these manipulations and task effects to explore the extent to which listeners differ in how well they process information about talkers.

Based on previous research, we predict better performance for words than nonwords and better performance for items with high phonotactic probability compared to those with low phonotactic probability. Given that some studies have found good within-phoneme discrimination in individuals with reading impairments (Serniclaes et al., 2001, 2004), one possible outcome is that reading ability will not modulate performance on our talker discrimination tasks (Experiments 1a and 1b), because listeners need not tap into long-term phonological or talker representations. In contrast, some studies have found poorer performance on phoneme discrimination tasks in individuals with reading impairments (Mody et al., 1997; Werker & Tees, 1987); thus, if these results demonstrate a more global impairment in perception, we may find that individuals with poorer reading skills perform worse on both the discrimination and the identification tasks. The structure of our statistical analyses allows us to test whether reading ability per se (or instead, phonological working memory) better predicts performance on these tasks. Based on previous research demonstrating poorer performance on talker identification tasks for individuals with poorer reading skills, we expect to find an effect of reading ability or of phonological working memory on the identification task (Experiment 2).

## Experiment 1

Experiment 1 examined talker perception using a discrimination task in which listeners heard two words or two nonwords and had to determine whether the words were spoken by the same talker or different talkers. Two versions of the experiment were run with different sets of participants. Experiment 1a was conducted with a blocked design, with one block for real words and one for nonwords (counterbalanced). Stimuli were blocked by lexicality to increase the likelihood of revealing effects of phonotactic probability (Vitevitch & Luce, 1999). Experiment 1b mixed the real word and nonword trials together, though individual trials were always only words or nonword pairs. These two versions were tested because previous studies have found that the impact of sub-lexical processing differs depending on whether lexical status is blocked or randomized in the experiments (Vitevitch & Luce, 1999).

## Methods

### Participants

Thirty-seven listeners between the ages of 18 and 23 years ( $M = 19.02$ ,  $SD = 1.07$ ) participated in Experiment 1a (29 female, eight male), and 36 listeners between the ages of 21 and 25 years ( $M = 19.59$ ,  $SD = 1.30$ ) participated in Experiment 1b (25 female, 11 male). Listeners were recruited from the psychology research pool at New York University and received partial course credit for participating. Participants were native speakers of American English with no history of speech, language, or hearing disorders, and all passed a pure tone hearing screening at 25 dB HL for octave frequencies of 500, 1,000, 2,000, and 4,000 Hz.

Seven additional participants completed Experiment 1a but were not included in the data analysis for the following reasons: falling asleep during the experiment ( $N = 1$ ), having speech and language therapy ( $N = 1$ ), and errors in administering the experiment ( $N = 5$ ). Fifteen additional participants completed Experiment 1b but were excluded from the data analysis for the following reasons: reported English as their non-dominant language ( $N = 3$ ) and errors in administering the experiment ( $N = 12$ ).

### Materials

**Items** Using Storkel and Hoover's (2010) database of consonant-vowel-consonant (CVC) words and nonwords, 192 monosyllabic items were selected for the current study (see Appendix A). Words were selected from the adult corpus to have a lexical frequency above 100 based on Kučera and Francis (1967). From among words fitting this criterion, 48 words were selected to have high (e.g., "head") and 48 low (e.g., "peace") phonotactic probability. Similarly, nonwords from the adult corpus were selected to have either high (e.g., "var") or low (e.g., "shoose") phonotactic probability. Phonotactic probability was determined by the biphone frequency corresponding to the adult corpus. Independent-samples *t*-tests revealed no statistically significant differences between words and nonwords with high phonotactic probability ( $t(94) = 0.41$ ,  $p = 0.66$ ) or words and nonwords with low phonotactic probability ( $t(94) = 1.38$ ,  $p = 0.17$ ). The mean and range of phonotactic probability across the lexical groups are presented in Table 1, and splits approximately at the mean of .0064 for the adult corpus. Appendix B provides information about the distribution of consonants across the four conditions.

**Speakers** Four female native speakers of American English were recorded producing the 192 items. Speakers were between the ages of 19 and 24 years from the New York City region (New Jersey and New York) with no reported history

**Table 1** Phonotactic probability range (mean in parentheses) for the four categories of items

	High	Low
Word	.006-.019 (.010)	.0008-.005 (.003)
Nonword	.006-.019 (.009)	.0008-.005 (.002)

of speech and language disorders (see Table 2). The recordings took place in a sound-attenuated IAC booth at the Communicative Sciences and Disorders Department at New York University using a SHURE 10A head-mounted microphone and a Fostex FR-2LE recorder with a 44.1-kHz sampling rate.

Items were presented orthographically in capital letters in E-Prime 2.0 (Schneider & Zuccoloto, 2007). Words and nonwords were presented separately and included four practice trials prior to the presentation of the test targets. The real words appeared for 1,000 ms. Words were presented twice in random order. For the nonwords, the experimenter controlled the duration that the item appeared on the screen so as not to rush the talkers. In addition to an orthographic representation, a rhyming word was included to facilitate pronunciation. In cases where the pronunciation of the vowel or consonant was ambiguous, additional information was provided (e.g., "th like thin" or "th like that"). Each nonword appeared twice in random order except nonwords that started an interdental fricative, which appeared four times in total to ensure an accurate production. For both words and nonwords, the second repetition of the target was selected unless that production had an error. All sound files were amplitude normalized to have the same average amplitude using Praat (Boersma & Weenink, 2008) and presented to listeners at a comfortable listening level.

### Procedure

**AX talker discrimination** Participants first completed the AX talker discrimination task in which they heard two different words or nonwords and were then asked to decide whether these items were produced by the same talker or different talkers. The experiment was conducted in a quiet testing room in the Department of Communicative Sciences and Disorders

**Table 2** Speaker's demographic and acoustic information

Speaker	Age	Mean duration (ms)	Mean F0 (Hz)
Female 1	21	599 ms	179 Hz
Female 2	22	510 ms	199 Hz
Female 3	19	593 ms	206 Hz
Female 4	24	720 ms	193 Hz

at New York University. Stimuli were presented over Sennheiser HD-280 circumaural headphones via a desktop PC running Windows 7 and E-Prime 2.0 (Schneider & Zuccoloto, 2007). Responses were collected using a Chronos response box. “Same” and “Different” labels were placed on the leftmost and rightmost buttons and counterbalanced across participants. The participants were instructed to respond as quickly and as accurately as possible and were encouraged to make a decision even when they were unsure.

Prior to the experimental trials, the participants completed a practice block with feedback that contained novel words produced by two additional talkers not included in the test trials. In Experiment 1a, the experimental trials were blocked by lexical status (either words or nonwords) and counterbalanced across participants. In Experiment 1b, word and nonword trials were mixed together with a break halfway through (after 48 trials). Even though the trials were mixed, each individual trial only contained either words or nonwords. For both Experiment 1a and Experiment 1b, trials contained two items that were drawn from the same bin, of either high or low phonotactic probability. Half of the trials were same-talker pairs and half were different-talker pairs. In the same-talker trials, each talker occurred in the same number of trials, and in the different-talker trials, the combinations of talkers were balanced such that each talker was paired with one of the other talkers and in the same frequency of occurrence.

The two items in each trial were separated by a 500-ms interstimulus interval. Prior to each trial, a fixation (+) appeared for 750 ms followed by a blank screen for 100 ms. After participants made their response, there was an additional 250 ms before the start of the next trial. The AX experiment took approximately 10 min and was always administered first.

**Tests of reading subskills** Following the talker AX task, participants completed standardized tests of reading subskills. Following Kadam et al. (2016), we administered the same ten reading subtests to the participants in Experiment 1a. Test order was counterbalanced across participants. For each participant, a composite reading score was calculated as the mean of the percentile rank across the ten subtests.

Two subtests of the Test of Word Reading Efficiency (TOWRE; Torgesen et al., 1999) were administered: (1) Sight Word Reading and (2) Phonemic Decoding Efficiency. In the Sight Word Reading subtest, participants have 45 seconds to read as many real words as possible from a list. The Phonemic Decoding Efficiency is the same, except the list is comprised of nonwords (e.g., “ip”). Two subtests of the Woodcock Reading Mastery Test-III (WRMT; Woodcock, 2011) were administered: (1) Word Identification and (2) Word Attack. These tests are similar to the TOWRE, but do not have a time limit. In the Word Identification subtest, participants read real words and in the Word Attack subtest, participants read nonwords (e.g.,

“laip”). Participants were recorded while completing these subtests. Two trained research assistants listened to the recordings and transcribed the items as correct or incorrect. In the case where there was disagreement, a third trained research assistant completed the transcription.

Six subtests from the Comprehensive Test of Phonological Processing (CTOPP; Wagner et al., 1999) were used: (1) Blending, (2) Elision, (3) Nonword Repetition, (4) Rapid Digit Naming, (5) Rapid Letter Naming, and (6) Rapid Object Naming. In Elision, participants first repeat a word and then produce it while excluding a sound or sounds (e.g., “say time without saying /m/”). In Blending, participants are instructed to put sounds together to form a word (e.g., “*what word do these sounds make: can-dy?*”). The Nonword Repetition task contains prerecorded nonwords that range from one to six syllables. In the three rapid naming tasks, participants see objects, numbers, or letters on a page and are asked to name them as fast as possible and are timed. Responses on the nonword repetition task and on the rapid naming tasks were recorded. Nonword repetition responses were coded by two trained research assistants and any differences were adjudicated by a third transcriber. For the rapid naming tasks, duration measures were taken in Praat and accuracy was coded by two trained research assistants.

To reduce the time of the experiment, only the two subtests from the TOWRE and the nonword repetition task from the CTOPP were administered to participants in Experiment 1b. As will be seen in the results section for Experiment 1a, the nonword repetition – a test of phonological working memory – resulted in the best model fit, which is why this subtest was also administered to the participants in Experiment 1b. The two TOWRE subtests were selected from among all of the possible subtests because the combined composite from these two subtests had the highest correlation with the overall composite composed of all ten subtests of reading skills. Correlations were calculated from the participants in Experiment 1a and an additional 20 participants who completed these ten subtests as part of a different experiment in our lab. The TOWRE composite score showed the strongest positive correlation with the overall composite reading score (average of ten subtests) ( $r = .85$ ,  $p < 0.01$ ). The correlation matrix is shown in [Appendix C](#). Thus, the participants in Experiment 1b only completed the nonword repetition task and the two subtests from the TOWRE.

### Statistical analyses

For each set of data, a series of linear mixed-effect models were fit to the data using the lme4 package (Bates et al., 2015) in RStudio (Team, 2017). Separate models were constructed for the three outcome measures: sensitivity, bias, and reaction time.

Because listeners often show a bias response on AX tasks, we conducted an analysis based on signal detection theory,

with nonparametric measure of sensitivity ( $A'$ ) (see Equations 1.a and 1.b; Grier, 1971) and bias ( $B_D''$ ) (see Equations 2.a and 2.b; Donaldson, 1992). The sensitivity measure accounts for performance on both same and different trials, and is based on the proportion of hits and false alarms. In this analysis, hits corresponded to responses of *different* on different-talker trials and false alarms corresponded to responses of *different* on same-talker trials.  $A'$  measures the listener's sensitivity to detect differences between talkers and ranges from 0.5 to 1.0, where a higher value indicates better sensitivity to detect differences. In the bias measure, positive values correspond to a bias towards responding same-talker, while negative values correspond to a bias responding different-talker. Measures of  $A'$  and  $B_D''$  were generated for each of the four conditions (two levels of lexicality by two levels of phonotactic probability).

$$A' = 0.5 + \frac{(\text{PHIT}-\text{PFA}) \cdot (1 + \text{PHIT}-\text{PFA})}{4 \cdot \text{PHIT} \cdot (1-\text{PFA})} \text{ when } H \geq F \text{ (1.a)}$$

$$A' = 0.5 - \frac{(\text{PFA}-\text{PHIT}) \cdot (1 + \text{PFA}-\text{PHIT})}{4 \cdot \text{PFA} \cdot (1-\text{PHIT})} \text{ when } H < F \text{ (1.b)}$$

$$B_D'' = \frac{\text{PHIT} \cdot (1-\text{PHIT}) - \text{PFA} \cdot (1-\text{PFA})}{\text{PHIT} \cdot (1-\text{PHIT}) + \text{PFA} \cdot (1-\text{PFA})} \text{ when } H \geq F \text{ (2.a)}$$

$$B_D'' = 0.5 + \frac{\text{PFA} \cdot (1-\text{PFA}) - \text{PHIT} \cdot (1-\text{PHIT})}{\text{PFA} \cdot (1-\text{PFA}) + \text{PHIT} \cdot (1-\text{PHIT})} \text{ when } H \geq F \text{ (2.b)}$$

For each outcome measure, two models were considered that differed on which individual differences measure was included (reading composite or nonword repetition). In Experiment 1a the first model used the ten-subtest, overall composite score as the measure of reading ability (following Kadam et al., 2016), and the second model used performance on the nonword repetition task, a measure of phonological working memory (following Perrachione et al., 2011). In Experiment 1b the first model used the two subtests of the TOWRE as the reading measure (because the correlations indicated that this measure was most highly correlated with the overall ten-subtest composite), and the second model used nonword repetition. The nonword repetition task is normed to 10 with a standard deviation of 3 and the TOWRE composite score is normed to 100 with a standard deviation of 15.

All models included fixed effects for lexical status (word/nonword), phonotactic probability (high/low), the scaled individual differences measure, and all two-way interactions. In Experiment 1a, models also included a fixed effect for block order (words first, nonwords first). The models of reaction time included a fixed effect of trial-type (same trials, different trials). All models included random intercepts for participants and all categorical fixed effects were sum-coded.

The two models that differed in which individual differences measure was included were compared using Bayesian Information Criterion (BIC) (Schwarz, 1978) to select the best fitting model. BIC is used for model selection criteria that measures the goodness of fit of the model to the data. The model with the lowest BIC is chosen. To determine whether the lowest BIC value indicated a meaningful difference from the competing model, we calculated Bayes factors for each model comparisons based on Wagenmakers (2007) and Raftery (1995).

## Results

### Sensitivity: Experiment 1a

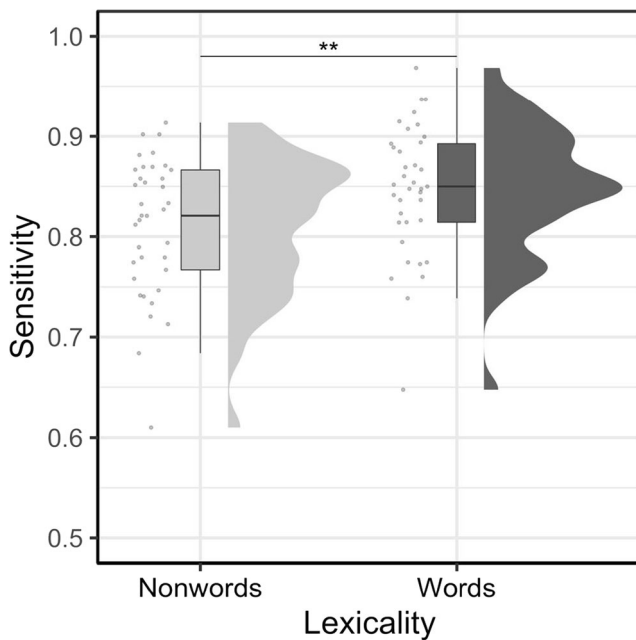
The model using the nonword repetition task as the individual differences measure had the lower BIC value (-283.9 vs. -282.5). Bayes factors suggested weak evidence for this selected model with the nonword repetition task over the model with the composite score ( $BF_{01} = 2.01$ ,  $Pr = .67$ ). Results revealed a significant effect of lexical status on sensitivity ( $p = 0.002$ ), where listeners had higher sensitivity to talker differences when listening to words compared to nonwords (Fig. 1). No other effects reached significance. The full output of the selected model can be found in [Appendix D](#).

### Bias: Experiment 1a

The lowest BIC value was found for the model that used the nonword repetition task as the individual differences measure (202.8 vs. 203.5). Bayes factors suggested weak evidence for this model ( $BF_{01} = 1.42$ ,  $Pr = .59$ ). The analysis of bias revealed only a significant effect of lexical status ( $p < 0.001$ ), where listeners were more biased to respond *same talker* when they heard words compared to nonwords (Fig. 2). No other effects reached significance. The full output of the selected model can be found in [Appendix D](#).

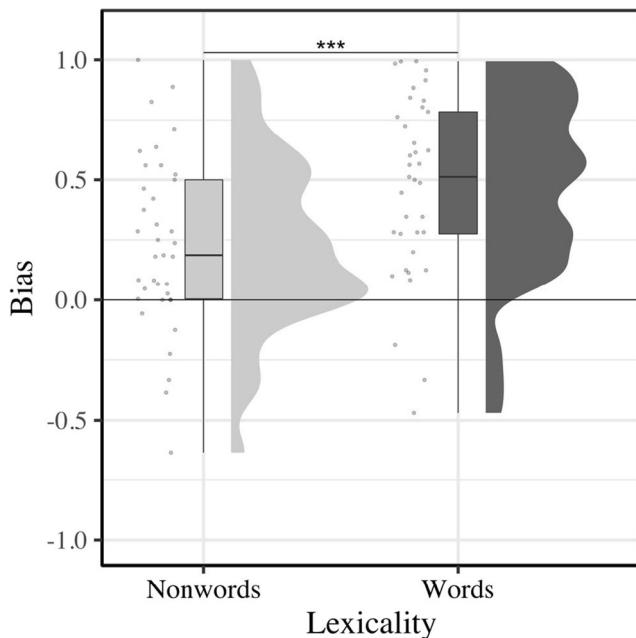
### Reaction time: Experiment 1a

First, incorrect responses were eliminated from the analysis (911/3552 trials, 25.64%). From the remaining correct trials, trials with responses longer than 2,000 ms and shorter than 200 ms from the onset of the second item were removed from the analysis (53/2641, 2%). All reaction times were log transformed for modeling. The lowest BIC value was found for the model that used the nonword repetition as the individual differences measure (166.7 vs. 169.0). Bayes factors suggested positive evidence for this model ( $BF_{01} = 3.16$ ,  $Pr = .76$ ). The selected model revealed a significant effect of lexical status ( $p < 0.001$ )



**Fig. 1** Sensitivity ( $A'$ ) by lexical status for Experiment 1a. Higher values indicate better talker discrimination

where listeners responded faster on the word trials than on the nonword trials. In addition, there was a significant interaction between lexical status and phonotactic probability ( $p = 0.003$ ). An emmeans analysis (Lenth, 2019) revealed a significant effect of phonotactic probability in the word condition only, as shown in Table 3 and in Fig. 3. These findings indicate that when listeners were responding to real words, they were faster for words with



**Fig. 2** Bias ( $B_D$ ) by lexical status in Experiment 1a. Positive values indicate a bias towards responding “same,” while negative values indicate a bias towards responding “different”

high phonotactic probability than those with low phonotactic probability. Although the pairwise comparison for the nonwords did not reach significance, it shows a trend in the opposite direction with faster reaction times for nonwords with low phonotactic probability than those with high phonotactic probability. No other effects reached significance. The full output of the selected model can be found in Appendix D.

**Sensitivity: Experiment 1b**

The model using the TOWRE composite score as the individual differences measure resulted in the lowest BIC value (-225.9 vs. -224.9). Bayes factors suggested weak evidence for this model ( $BF_{01} = 1.65$ ,  $Pr = .63$ ). In this version of the experiment where word pairs and nonword pairs were not blocked, no effects reached significance. The full model output is in Appendix E.

**Bias: Experiment 1b**

The lowest BIC value was found for the model that included the nonword repetition as the individual differences measure (230.9 vs. 232.2). Bayes factors suggested weak evidence for this model ( $BF_{01} = 1.92$ ,  $Pr = .66$ ). The selected model revealed a significant interaction between lexical status and phonotactic probability ( $p = 0.032$ ). An emmeans analysis (Lenth, 2019) revealed a significant effect of phonotactic probability in the nonword condition only (see Table 4 and Fig. 4). This finding indicates that when listeners heard nonwords, they were more biased to respond *same talker* for items with high phonotactic probability compared to those with low phonotactic probability. No other effects reached significance. The full model output is in Appendix E.

**Reaction time: Experiment 1b**

First, incorrect responses were eliminated from the analysis (900/3360 trials, 26.78%). Then, from the remaining correct trials, trials with responses longer than 2,000 ms and shorter than 200 ms from the onset of the second word were removed from the analysis (103/2460, 4.18%). All remaining reaction times were log transformed for modeling. The lowest BIC value was found for the model with nonword repetition as the individual differences measure (352.6 vs. 356.0). Bayes factors suggested positive evidence for the selected model ( $BF_{01} = 5.47$ ,  $Pr = .85$ ). Despite positive evidence for this model over the other, no effects reached significance. The full model output is in Appendix E.



**Table 3** Pairwise comparison (emmeans) of lexical status and phonotactic probability effects on predicted estimates of listener reaction time (log) based on the mixed-effect linear regression model

Lexical status	Phonotactic probability comparison	Estimate	SE	<i>t</i>	<i>p</i> *
Nonword	High phonotactic probability vs. low phonotactic probability	0.026	0.013	1.94	0.052
Word	High phonotactic probability vs. low phonotactic probability	-0.029	0.013	-2.19	<b>0.028</b>

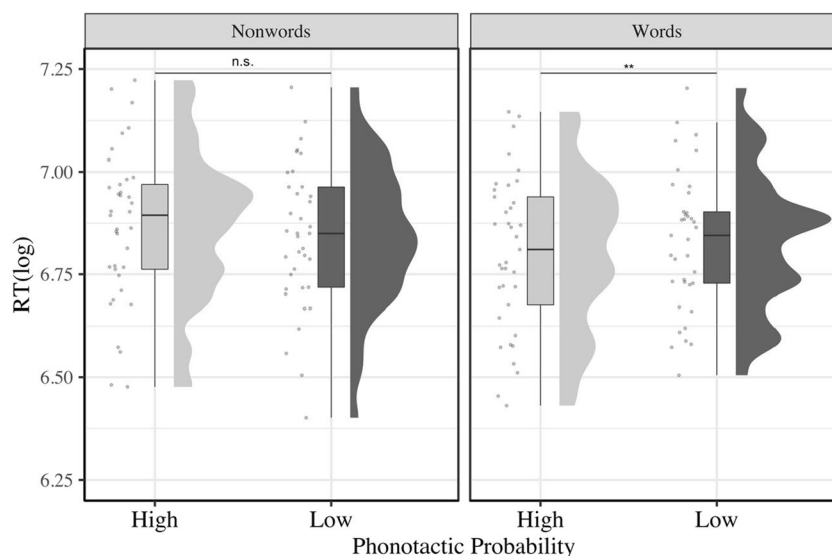
*p* values are adjusted using Holm's method and indicated as *p*\*

## Discussion

The results from Experiments 1a and 1b showed evidence for a contribution of both lexical information and phonological information, though no effects of reading ability or phonological working memory were found. When words and nonwords were blocked (1a), participants were more sensitive to detecting talker differences for words than for nonwords, similar to studies that have found better performance for talker identification with word or nonword sentences (Perrachione et al., 2015; Xie & Myers, 2015). In addition, participants were more biased to respond *same talker* for words than nonwords (in the blocked version, 1a). This may be part of a continuum of perceptual response bias; previous studies have found that when the two items on a talker discrimination task co-occur (as in the case with lexical compounds such as “day-dream”), listeners are more likely to respond that the items are spoken by the same person (Narayan et al., 2017; Quinto et al., 2020). The data in Experiment 1 with words versus nonwords may demonstrate the extreme of this continuum, where nonwords have no co-occurrence by virtue of having a lexical frequency of zero. These effects of lexical status only appeared when the items were blocked by lexical status.

The effect of phonotactic probability was subtler. In Experiment 1a when items were blocked, participants were faster for real words that were more word-like (high phonotactic probability) than real words that were less word-like (low phonotactic probability). In Experiment 1b, there was no overall effect of lexical status on response bias; however, the less word-like a nonword pair was (low phonotactic probability), the less biased participants were to respond *same talker*.

Experiment 1 found no evidence for an effect of reading ability or phonological working memory (nonword repetition) on talker discrimination. These results are in contrast with studies that found an effect of reading ability and phonological working memory (as a measure of phonological processing skills) on talker processing (Kadam et al., 2016; Perea et al., 2014; Perrachione et al., 2011). One major difference between these previous studies and Experiments 1a and 1b is that our experiments used a discrimination task, whereas the previous studies used talker identification tasks following a brief training. Given that previous studies with individuals with reading impairments have found task-modulated effects on classic tests of speech perception, one possibility is that the contribution of a participant's reading ability or phonological



**Fig. 3** Reaction time (RT) by lexical status and phonotactic probability in Experiment 1a. Larger values indicate slower RTs

**Table 4** Pairwise comparison (emmeans) of lexical status and phonotactic probability on predicted estimates of listener bias on the mixed-effect linear regression model

Lexical status	Phonotactic probability comparison	Estimate	SE	<i>t</i>	<i>p</i> *
Nonword	High phonotactic probability vs. low phonotactic probability	0.197	0.098	2.00	<b>0.047</b>
Word	High phonotactic probability vs. low phonotactic probability	-0.097	0.098	-0.98	0.325

*p* values are adjusted using Holm's method and indicated as *p*\*

processing ability would only occur for tasks that tap into long-term memory and require access to stored phonological representations of words. This hypothesis stems from differences in phoneme perception tasks (poorer performance on phoneme identification tasks and possibly intact performance on phoneme discrimination tasks) that have been attributed to impaired phonological representations (Serniclaes et al., 2001, 2004). Identification tasks require that listeners retrieve a label from long-term memory (Bricker & Pruzansky, 1976). If that stored representation is impaired, tasks that require mapping incoming signals to that representation will also be impaired. In contrast, discrimination tasks can be performed for items that have no stored representation. Thus, for talker processing tasks, one possible explanation for the lack of an effect of reading ability or phonological processing ability on the talker discrimination task could be that listeners do not need to access stored representations of talker, lexical, or phoneme information.

Experiment 2, which uses a training-identification paradigm, allows us to examine whether individuals with poorer reading or phonological processing skills also have difficulty creating an auditory category that is not directly linguistic in

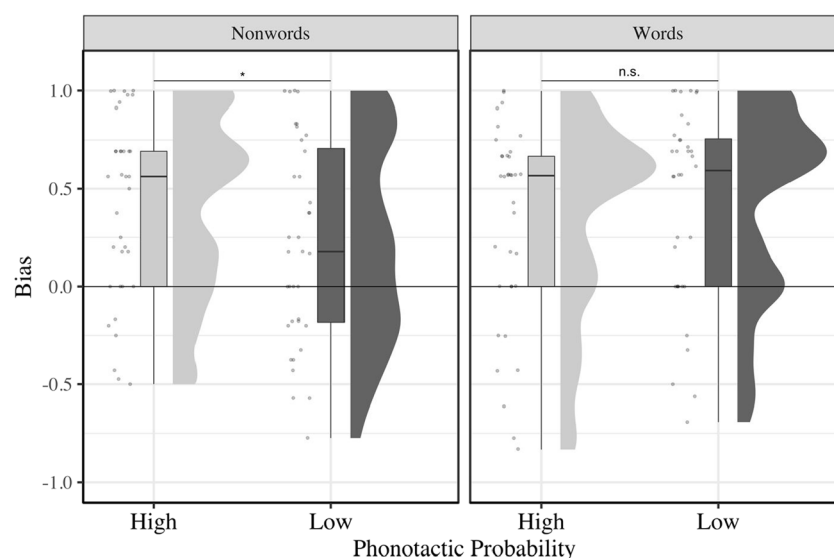
nature. If individuals with poorer reading or phonological processing skills are impaired in other types of auditory categorization that goes beyond phonological categories, it is possible that these individual differences effects will be present when using a talker identification task even though no effect was found for the talker discrimination task. The purpose of Experiment 2 was to test whether individual differences measures will emerge as a predictor of performance on a task that requires labeling and access to long-term memory representations.

## Experiment 2

### Methods

#### Participants

Thirty-one listeners between the age of 19 and 22 years ( $M = 19.74$ ,  $SD = 1.15$ ) participated in the study (23 females, eight males). Listeners were recruited from the psychology research pool at New York University and received partial course



**Fig. 4** Bias ( $B_D$ ) by lexical status and phonotactic probability in Experiment 1b. Positive values indicate a bias towards responding “same,” while negative values indicate a bias towards responding “different”

credit for participating. Participants were native speakers of American English with no history of speech, language, or hearing disorders and all passed a pure tone hearing screening at 25 dB HL for octave frequencies of 500, 1,000, 2,000, and 4,000 Hz. Eight additional participants completed [Experiment 2](#) but were not included in the data analysis for the following reasons: having speech and language therapy ( $N = 1$ ) and errors in administering the experiment ( $N = 7$ ).

## Materials

The same 192 items from [Experiment 1](#) were used for Experiment 2.

## Procedure

**Talker identification task** Listeners completed talker training with feedback followed by a talker identification test without feedback. In the training phase, listeners learned to identify the voices of four talkers (Fig. 5). Participants were trained on either words or nonwords, but all were tested on both words and nonwords. The training phase started with a familiarization in which they heard each talker say four words or nonwords (two high phonotactic probability, two low phonotactic probability) while the image of only that talker was displayed on the screen. This familiarization was repeated twice. Listeners then completed a training phase with four blocks of 52 trials (24 high phonotactic probability and 24 low phonotactic probability  $\times$  4 talkers = 208 trials total) presented in random order. Listeners were asked to decide which talker produced the target item by clicking a button on a Chronos response box. The two rightmost and two leftmost buttons that corresponded to the position of the characters on the screen were used. After each response, listeners received two forms of feedback: one indicating accuracy (correct or incorrect), and another playing the

same target item with only the image of the correct talker on the screen.

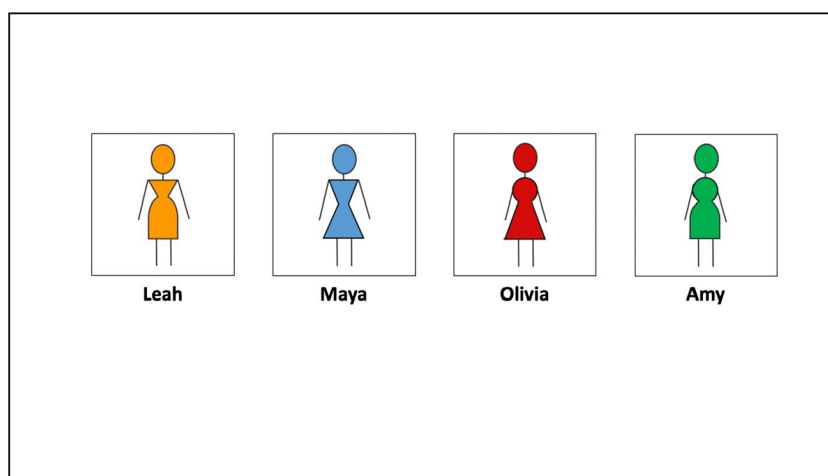
Following the training phase, listeners completed a talker identification test similar to training except no feedback was provided. Testing contained 80 trials: five items were selected from each of the four categories (high and low phonotactic probability, word and nonword). These 20 items were presented from each of the four talkers. None of the items from the test phase were used during training. A break screen appears halfway through the test phase.

**Tests of reading subskills** Participants in Experiment 2 completed the same three standardized subtests as the participants in Experiment 1b.

## Statistical analyses

Two data sets were included for analysis: training and testing. For each data set, two models with accuracy as the outcome measure were considered: one with the TOWRE composite score and one with the nonword repetition task as the measure of individual differences.

Both models included fixed effects for lexical status (word/nonword), phonotactic probability (high/low), scaled individual differences (TOWRE composite score or nonword repetition), condition (trained on words, trained on nonwords), talker (1–4), and all two-way interactions between lexicality, phonotactic probability, condition, and individual differences measures. The analysis of training also included block and an interaction between block and condition. Block was defined as a split between the first and second halves of the training portion. All models included random intercepts for participant. Similar to Experiments 1a and 1b, the models were compared using BIC values and evaluated for the differences between them using Bayes factors.



**Fig. 5** The four characters from Experiment 2 as they appear on the response screen for the training and testing tasks

## Results

### Training

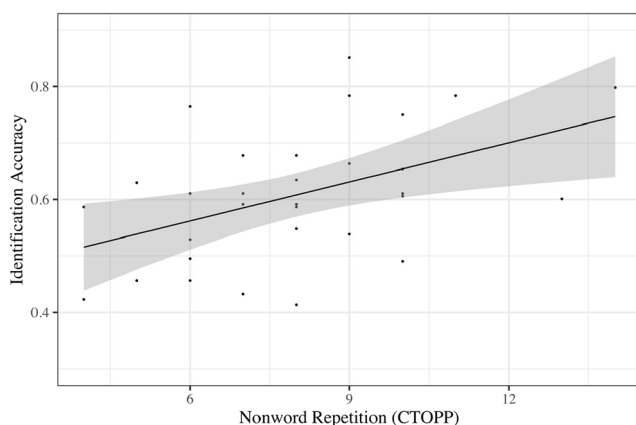
The lowest BIC value was found for the model that included the nonword repetition (8327.3 vs. 8333.0). Bayes factors suggested positive evidence for this model ( $BF_{01} = 17.29$ ,  $Pr = .94$ ). The model showed a significant effect of nonword repetition ( $p = .002$ ), where listeners with higher nonword repetition scores were more accurate on talker identification than those with lower scores (see Fig. 6), and a significant effect of block ( $p < .0001$ ) where accuracy was greater in the later blocks compared to earlier blocks.

In addition, there was a significant interaction between block and condition ( $p = .031$ ). The emmeans analysis in Table 5 of this interaction revealed no difference in performance between those trained on words and those trained on nonwords in the first half of training, but by the second half of training, those trained with words performed significantly better than those trained with nonwords (see Fig. 7), suggesting a benefit for items with lexical representations.

Finally, the model revealed a significant effect of talker, where some talkers were more identifiable than others. The emmeans analysis revealed that listeners performed significantly better when listening to talker 4 compared to the other talkers (see Appendix F). No other effects reached significance.

### Testing

The lowest BIC value was found for the model that included nonword repetition (3197.7 vs. 3203.6). Bayes factors suggested positive evidence for this model ( $BF_{01} = 19.11$ ,  $Pr = .95$ ). This model revealed that nonword repetition significantly predicted accuracy ( $p = .001$ ). Listeners with higher nonword repetition scores were more accurate on the talker



**Fig. 6** Relationship between the nonword repetition task (scaled score) on identification accuracy (proportion correct) during the training phase for Experiment 2

identification task (see Fig. 8). In addition, there was a significant effect of phonotactic probability ( $p = .036$ ) where accuracy was greater for items with *low* phonotactic probability than high phonotactic probability (Fig. 9).

The model also revealed a significant interaction between lexical status and condition ( $p = 0.001$ ), with greater accuracy for items that match the lexical status they were trained on. The emmeans analysis in Table 6 of this interaction shows that listeners trained on nonwords performed significantly better with new nonwords than with real words (see Fig. 10). Although the effect of matching training and test items did not reach significance for the participants trained on real words, the pattern of results is in the expected direction.

Similar to training, the model revealed a significant effect of talker, where some talkers were more identifiable than others. The emmeans analysis revealed that listeners performed significantly differently when listening to talker 4 compared to the other talkers (see Appendix F). No other effects reached significance.

## Discussion

The results from Experiment 2 show that listeners can learn talker's voices during the learning phase, but also that learning talker's voices is affected by properties of items. In particular, during learning, listeners who were trained with words performed better than those trained with nonwords. This finding suggests a benefit of lexical information on talker processing. These results are in line with Perrachione et al. (2015) and Xie and Myers (2015), who found better talker identification for sentences composed of real words compared to those composed of nonsense words. Our findings here are important because prior to our study, the only other study with word-length stimuli (Zarate et al., 2015) had not found a benefit of real words over nonsense words on talker-identification accuracy. Thus, our findings suggest that the lack of a difference in Zarate et al.'s (2015) study is not due to differences in the length of the stimuli, but to some other aspect of the study, such as using a mixture of native and non-native talkers or using the same talkers throughout the conditions. Our results also suggest that these differences may wash out with additional training, because we only found a benefit of words over nonwords in the training phase, but not in the testing phase. We return to this in the *General discussion*.

Our results revealed an effect of phonotactic probability, although the directionality of this effect was counter to our hypothesis where listeners were more accurate at identifying talkers when they heard words and nonwords with *low* phonotactic probability than high phonotactic probability.

Our results also revealed an effect of phonological working memory on talker identification in both training

**Table 5** Pairwise comparison (emmeans) of block and condition on predicted estimates of listener accuracy on the mixed-effect linear regression model

Block	Condition comparison	Estimate	SE	z- ratio	<i>p</i> *
Early block	Trained on nonword vs. word	-0.116	0.172	-0.67	0.498
Late block	Trained on nonword vs. word	-0.351	0.173	-2.02	<b>0.042</b>

*p* values are adjusted using Holm's method and indicated as *p*\*

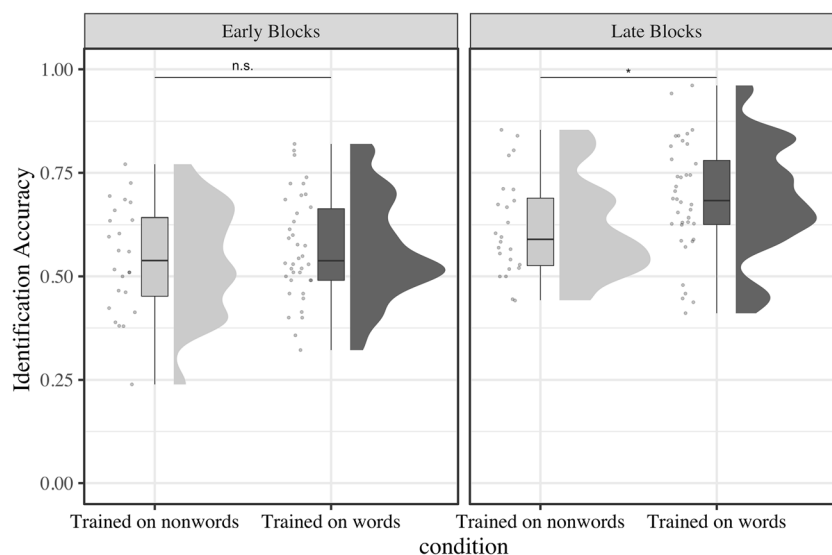
and testing where listeners with higher nonword repetition scores were more accurate to identify talkers. These results are similar to studies that have found an effect of reading ability in individuals with reading impairments (Perea et al., 2014; Perrachione et al., 2011) or typical readers (Kadam et al., 2016). Recall that Perrachione et al. (2011) also found evidence for nonword repetition as a predictor of performance. Because the effect of individual differences measures (here, nonword repetition) was found in Experiment 2 but not in Experiment 1, this supports the idea that individuals with poorer phonological processing skills have a deficit in generating categories that extends beyond linguistic processing and impacts their ability to create a talker category. That said, it is not possible to determine whether the poorer phonological processing skills result in poorer talker identification or whether a different underlying mechanism causes both of these reduced skills.

Finally, Experiment 2 found an effect of matching between training and testing, where listeners performed better when an item's lexical status matched what they were trained on; listeners trained on nonwords performed significantly better with new nonwords than with words, and listeners trained on words

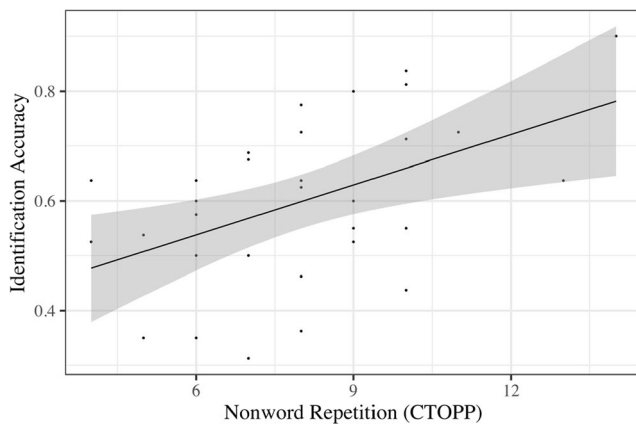
performed better on words than nonwords. These results are in line with research on encoding specificity, where individuals access, recall, and retrieve information learned more easily when the conditions of learning and testing match (Tulving & Thomson, 1973).

## General discussion

The goal of the current study was to simultaneously examine the contribution of lexical information, phonological familiarity, and individual differences measures that relate to phonology on how listeners process information about who is talking. In addition to examining these properties of the materials and of the participants, we also tested talker processing in two different tasks, one that does not require any stored information (talker discrimination) and one that does (talker identification). Using both tasks is important especially in light of studies that found good performance on phoneme discrimination tasks in individuals with reading impairments, yet poor performance on phoneme identification tasks (Serniclaes et al., 2001, 2004). The impact of these manipulations



**Fig. 7** Identification accuracy (proportion correct) by block and condition in the training phase for Experiment 2

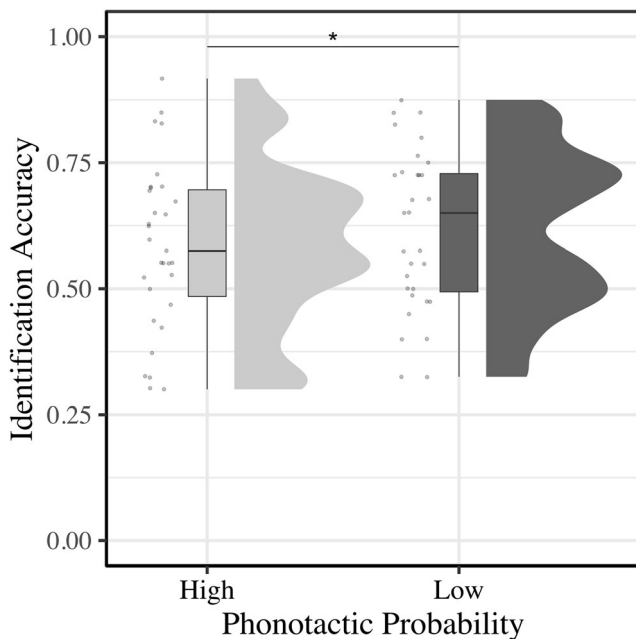


**Fig. 8** Relationship between the nonword repetition task (scaled score) on identification accuracy (proportion correct) during the testing phase for Experiment 2

on how listeners process talker information is discussed in more detail below.

### Support for lexical information

The current study found some effects of lexical information on talker processing. In the blocked version of the discrimination task (Experiment 1a), listeners were faster at discriminating between talkers in the word condition compared to the nonword condition, though this was washed out in the version where items were mixed (Experiment 1b). In Experiment 2, we did not find a



**Fig. 9** Identification accuracy (proportion correct) by phonotactic probability during the testing phase for Experiment 2

general effect of lexicality although lexical status appeared as: (1) a benefit in training where listeners trained on words showed greater improvement compared to those trained on nonwords, and (2) a benefit of matching between training and testing, as has been found in other studies of encoding specificity (Tulving & Thomson, 1973). These results are in line with previous studies (Perrachione et al., 2015; Xie & Myers, 2015; Zarate et al., 2015) that have demonstrated the effect of lexical information on talker processing using a training and identification task with sentence-length stimuli.

### Support for phonological familiarity

The contribution of the phonological familiarity was probed using two manipulations. First, words and nonwords were all phonologically legal, but differed in phonotactic probability. Tokens with high phonotactic probability contain more frequent combinations of sounds and should thus be more familiar in terms of their phonology. Second, individual differences in the listeners were examined as a predictor of performance using either a broad measure of reading (various composite scores) or a more direct measure of phonological processing ability (phonological working memory measured by the nonword repetition task).

Our results show a nuanced effect of phonotactic probability in the discrimination task (Experiment 1) and in the identification task (Experiment 2). Listeners were faster to discriminate real words with high phonotactic probability than real words with low phonotactic probability when words and nonwords were blocked, and less likely to respond *same talker* on nonwords with low phonotactic probability when words and nonwords were mixed. In contrast, there were no differences between high phonotactic probability and low phonotactic probability in the identification task; however, indirect support for phonological familiarity did emerge from the individual differences results as detailed below. It is worth pointing out that phonotactic probability is often elusive and highly dependent on the task (e.g., lexical decision, same/different) and also on whether items are blocked or interleaved (Luce & Large, 2001; Vitevitch & Luce, 1999, 2005). One thing that is not captured by this biphone frequency measure is familiarity or lexical frequency of the item; thus, an infrequent sequence may occur in a word with very high lexical frequency (e.g., “the”). As such, the phonotactic probability manipulation may not be the optimal way to assess phonological familiarity.

### Task effects and individual differences measures

Results from the identification task in Experiment 2 show that listeners with higher phonological processing skills

**Table 6** Pairwise comparison (emmeans) of lexical status and condition on predicted estimates of listener accuracy on the mixed-effect linear regression model

Condition	Lexical status comparison	Estimate	SE	t- ratio	<i>p</i> *
Trained on nonwords	Nonword vs. word	0.359	0.141	2.54	<b>0.010</b>
Trained on words	Nonword vs. word	-0.197	0.111	-1.77	0.076

*p* values are adjusted using Holm's method and indicated as *p*\*

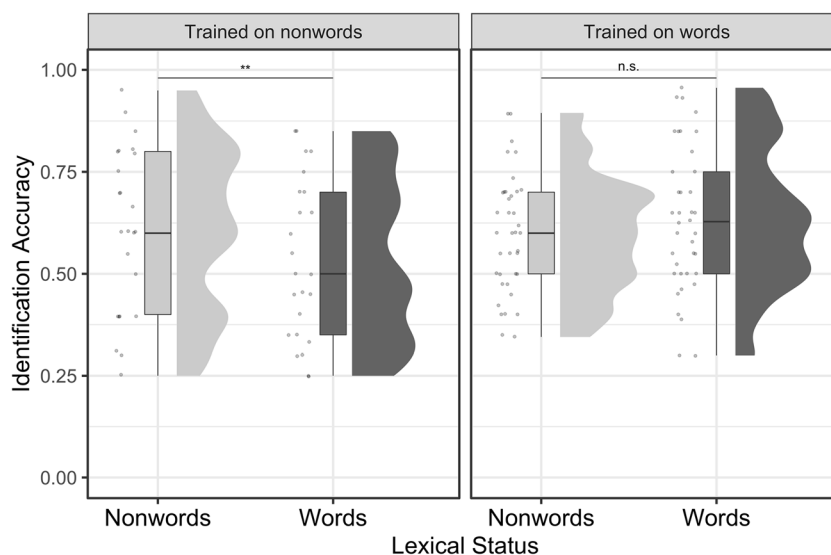
(better nonword repetition scores) were more accurate at identifying talkers. These results provide additional support for a listener's phonological processing skills, which has been tested previously either through measures of reading ability or directly with the nonword repetition task (Kadam et al., 2016; Perea et al., 2014; Perrachione et al., 2011). Two studies found an effect of reading ability in individuals with reading impairments (Perea et al., 2014; Perrachione et al., 2011) and one study found an effect of reading with typical readers (Kadam et al., 2016). These results support the idea that phonological familiarity and phonological processing skills affect how listeners perceive talker information.

The fact that the effects of phonological working memory were only present in the identification task illustrates that individuals with poorer phonological working memory are impaired in their ability to cluster information around a talker category. This idea fits well with the previous literature on the perception of phoneme categories, where there is evidence that phonological processing ability affects performance on labeling tasks and tap into long-term memory representations, but not necessarily

on tasks that can be completed with more peripheral, acoustic-matching processes, as can be done in a discrimination task (Serniclaes et al., 2001, 2004).

The type of task used to assess talker processing may thus have an effect on performance, as each task can activate or tap into different types of processing skills or resources (Levi, 2019). The cognitive demands these two tasks tap into have been described in Bricker and Pruzansky (1976). In an identification task, the listener is required to encode and store information about a talker in the long-term memory and then access it, whereas in a discrimination task, similar to the one in Experiment 1, the listener needs to only compare acoustic information from two tokens on each trial. Although it may be true that over the course of a discrimination experiment the listeners learn to identify the talker, they are nevertheless only required to make a decision on each individual trial. Therefore, while an identification task activates higher cognitive processing skills, a discrimination task requires lower levels of processing considering that the listener isn't required to store any talker information.

The model comparisons in Experiment 1 and 2 revealed that the nonword repetition generated better fitting



**Fig. 10** Accuracy (proportion correct) by training condition and lexical status in Experiment 2

models than using the composite scores for reading ability.<sup>1</sup> Thus, the results of previous talker processing studies that have found differences in performance based on reading ability may not be due to reading ability per se, but instead to concomitant difficulties with phonological processing (given previous research showing a relationship between reading ability and nonword repetition; Snowling et al., 1986; Snowling, 1981). Indeed, in Perrachione et al. (2011), performance on a talker identification task in English was correlated with performance on the nonword repetition task.

One question following this study is whether the deficit in creating an auditory category extends beyond speech sounds and talker processing. More research looking at how listeners with a range of phonological processing skills generate other non-linguistic categories is needed to answer these questions.

## Conclusion

Taken together, the current study shows an effect of both lexical and phonological information on talker processing using a talker discrimination task and a talker identification task. Listeners were more sensitive to talker differences when the tokens were real words compared to nonwords, and effects of phonological processing ability were only found for the identification task; listeners with higher phonological processing skills were more accurate in identifying talkers than listeners with lower scores. This supports the idea that phonological processing ability influences performance on tasks that require access to long-term phonological representation or tasks that require labelling categories. The effects of phonological processing skills on talker processing also supports the idea that the phonological make up of items matters for talker processing tasks and that poor phonological processing skills may be tied to a listener's ability to cluster information around a category.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.3758/s13414-022-02485-4>.

**Acknowledgements** We thank the research assistants Gretchen Go, Ashley Quinto, Ned Dana, Emily Maher, Mikala Miller, Maya Novik, Serena Piol, Janie Rist, Alison Vinokur, Jasleen Nair, Deanna Goudelias, Isabella Gadinis, and Nandita Karthikeyan who assisted in transcription and coding, and the participants who made this study possible.

## References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh University.

<sup>1</sup> This is based on the models that showed a positive evidence based on Bayes factors.

- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. *R package version, 1*(1–7), 2014.
- Boersma, P., & Weenink, D. (2008). Praat: Doing phonetics by computer. Version 5.0.06 [software].
- Boets, B., Vandermosten, M., Poelmans, H., Luts, H., Wouters, J., & Ghesquière, P. (2011). Preschool impairments in auditory processing and speech perception uniquely predict future reading problems. *Research in Developmental Disabilities, 32*(2), 560–570.
- Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read—A causal connection. *Nature, 301*(5899), 419–421.
- Bricker, P. D., & Pruzansky, S. (1976). Speaker recognition. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 295–326). Academic Press.
- Buchwald, A. (2014). Phonetic processing. *The Oxford Handbook of Language Production, 245–258*.
- Donaldson, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General, 121*(3), 275.
- Fecher, N., & Johnson, E. K. (2018). Effects of language experience and task demands on talker recognition by children and adults. *The Journal of the Acoustical Society of America, 143*(4), 2409–2418.
- Fleming, D., Giordano, B. L., Caldara, R., & Belin, P. (2014). A language-familiarity effect for speaker discrimination without comprehension. *Proceedings of the National Academy of Sciences, 111*(38), 13795–13798.
- Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin, 75*(6), 424.
- 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?*
- Kadam, M. A., Orena, A. J., Theodore, R. M., & Polka, L. (2016). Reading ability influences native and non-native voice recognition, even for unimpaired readers. *The Journal of the Acoustical Society of America, 139*(1), EL6–EL12.
- Köster, O., & Schiller, N. O. (1997). Different influences of the native language of a listener on speaker recognition. *Forensic Linguistics. The International Journal of Speech, Language and the Law, 4*, 11.
- Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Brown University Press.
- Lenth, R. (2019). *Emmeans: Estimated marginal means, aka least-squares means (version R package version (Vol. 1.3.4), p. (1.5.2-1))*. [Computer software].
- Levi, S. V. (2019). Methodological considerations for interpreting the language familiarity effect in talker processing. *Wiley Interdisciplinary Reviews: Cognitive Science, 10*(2), e1483.
- Luce, P. A., & Large, N. R. (2001). Phonotactics, density, and entropy in spoken word recognition. *Language and Cognitive Processes, 16*(5–6), 565–581.
- Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. *Annals of Dyslexia, 53*(1), 1–14.
- 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.
- Mody, M., Studdert-Kennedy, M., & Brady, S. (1997). Speech perception deficits in poor readers: Auditory processing or phonological coding? *Journal of Experimental Child Psychology, 64*(2), 199–231.
- Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics, 47*(4), 379–390.
- Narayan, C. R., Mak, L., & Bialystok, E. (2017). Words get in the way: Linguistic effects on talker discrimination. *Cognitive Science, 41*(5), 1361–1376.
- Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics, 60*(3), 355–376.



- O'Brien, G. E., McCloy, D. R., Kubota, E. C., & Yeatman, J. D. (2018). Reading ability and phoneme categorization. *Scientific Reports*, 8(1), 1–17.
- Perea, M., Jiménez, M., Suárez-Coalla, P., Fernández, N., Viña, C., & Cuetos, F. (2014). Ability for voice recognition is a marker for dyslexia in children. *Experimental Psychology*.
- Perrachione, T. K. (2019). Recognizing speakers across languages. In S. Frühholz & P. Belin (Eds.), *The Oxford handbook of voice perception* (pp. 514–538). Oxford University Press.
- Perrachione, T. K., Del Tufo, S. N., & Gabrieli, J. D. (2011). Human voice recognition depends on language ability. *Science*, 333(6042), 595–595.
- Perrachione, T. K., Dougherty, S. C., McLaughlin, D. E., & Lember, R. A. (2015). The effects of speech perception and speech comprehension on talker identification. *18th International Congress of Phonetic Sciences*.
- Quinto, A., Abu El Adas, S., & Levi, S. V. (2020). Re-examining the effect of top-down linguistic information on speaker-voice discrimination. *Cognitive Science*, 44(10), e12902.
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 111–163.
- Schneider, E., & Zuccoloto, A. (2007). *E-prime 2.0 [computer software]*. Pittsburg, PA: Psychological Software Tools.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 461–464.
- Serniclaes, W., Heghe, S. V., Mousty, P., Carré, R., & Sprenger-Charolles, L. (2004). Allophonic mode of speech perception in dyslexia. *Journal of Experimental Child Psychology*, 87(4), 336–361.
- Serniclaes, W., Sprenger-Charolles, L., Carré, R., & Demonet, J.-F. (2001). Perceptual discrimination of speech sounds in developmental dyslexia. *Journal of Speech, Language, and Hearing Research*, 44(2), 384–399.
- Snowling, M., Goulandris, N., Bowlby, M., & Howell, P. (1986). Segmentation and speech perception in relation to reading skill: A developmental analysis. *Journal of Experimental Child Psychology*, 41(3), 489–507.
- Snowling, M. J. (1981). Phonemic deficits in developmental dyslexia. *Psychological Research*, 43(2), 219–234.
- Storkel, H. L., Armbrüster, J., & Hogan, T. P. (2006). Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing Research*.
- Storkel, H. L., & Hoover, J. R. (2010). An online calculator to compute phonotactic probability and neighborhood density on the basis of child corpora of spoken American English. *Behavior Research Methods*, 42(2), 497–506.
- Team, R. C. (2017). *R: A language and environment for statistical computing* (p. 2016). Vienna, Austria: R Foundation for Statistical Computing.
- Thorn, A. S., & Frankish, C. R. (2005). Long-term knowledge effects on serial recall of nonwords are not exclusively lexical. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(4), 729.
- Torgesen, J. K., Rashotte, C. A., & Wagner, R. K. (1999). *TOWRE: Test of word reading efficiency*. TX: Pro-ed Austin.
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80(5), 352.
- Vitevitch, M. S., Armbrüster, J., & Chu, S. (2004). Sublexical and lexical representations in speech production: Effects of Phonotactic probability and onset density. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(2), 514–529.
- Vitevitch, M. S., & Donoso, A. (2011). Processing of indexical information requires time: Evidence from change deafness. *Quarterly Journal of Experimental Psychology*, 64(8), 1484–1493.
- Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, 40(3), 374–408.
- Vitevitch, M. S., & Luce, P. A. (2005). Increases in phonotactic probability facilitate spoken nonword repetition. *Journal of Memory and Language*, 52(2), 193–204.
- Vitevitch, M. S., & Luce, P. A. (2016). Phonological neighborhood effects in spoken word perception and production. *Annual Review of Linguistics*, 2, 75–94.
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14(5), 779–804.
- 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.
- Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive test of phonological processing: CTOPP*. Austin, TX: Pro-Ed.
- Werker, J. F., & Tees, R. C. (1987). Speech perception in severely disabled and average reading children. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, 41(1), 48.
- Wester, M. (2012). Talker discrimination across languages. *Speech Communication*, 54(6), 781–790.
- Winters, S. J., Levi, S. V., & Pisoni, D. B. (2008). Identification and discrimination of bilingual talkers across languages. *The Journal of the Acoustical Society of America*, 123(6), 4524–4538.
- Woodcock, R. W. (2011). *Woodcock Reading mastery tests: WRMT-III*. Pearson.
- Xie, X., & Myers, E. B. (2015). General language ability predicts talker identification. *CogSci*.
- Zarate, J., Tian, X., Woods, K. J. P., & Poeppel, D. (2015). Multiple levels of linguistic and paralinguistic features contribute to voice recognition. *Scientific Reports*, 5(1).
- Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, 131(1), 3.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.