Cognitive Psychology

The Effects of Input Modality, Word Difficulty and Reading Experience on Word Recognition Accuracy

Merel C. Wolf[1], Antje S. Meyer[2], Caroline F. Rowland[2], Florian Hintz[3] [a]

[1] Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands; International Max PlanckResearch School for Language Sciences, Nijmegen, The Netherlands, [2] Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands; Radboud University, Nijmegen, the Netherlands, [3] Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

Language users encounter words in at least two different modalities. Arguably, the most frequent encounters are in spoken or written form. Previous research has shown that – compared to the spoken modality – written language features more difficult words. An important question is whether input modality has effects on word recognition accuracy. In the present study, we investigated whether input modality (spoken, written, or bimodal) affected word recognition accuracy and whether such a modality effect interacted with word difficulty. Moreover, we tested whether the participants' reading experience interacted with word difficulty and whether this interaction was influenced by modality. We re-analyzed data from 48 Dutch university students that were collected in the context of a vocabulary test development to assess in which modality test words should be presented. Participants carried out a word recognition task, where non-words and words of varying difficulty were presented in auditory, visual and audio-visual modalities. In addition, they completed a receptive vocabulary and an author recognition test to measure their exposure to literary texts. Our re-analyses showed that word difficulty interacted with reading experience in that frequent readers (i.e., with more exposure to written texts) were more accurate in recognizing difficult words than individuals who read less frequently. However, there was no evidence for an effect of input modality on word recognition accuracy, nor for interactions with word difficulty or reading experience. Thus, in our study, input modality did not influence word recognition accuracy. We discuss the implications of this finding and describe possibilities for future research involving other groups of participants and/or different languages.

Introduction

With the invention of reading and writing, humans gained the opportunity to use language in another modality than the spoken form. This has an important consequence for the internal representational system of language: two representations –orthographic and phonological– of the same lexical item are stored. As a result of the quality and quantity of modality-specific encounters, these two representations can vary in their level of precision and completeness (i.e., lexical quality, Perfetti, 2007). Moreover, written language differs from spoken language in that written text has been shown to include a larger variety of words than speech does (Cunningham & Stanovich, 1998; Hayes & Ahrens, 1988). Consequently, the mental lexicon of frequent readers probably includes more difficult (i.e., less well-known) words than that of individuals who read less.

The fact that difficult words are encountered most often in the written modality (Cunningham & Stanovich, 1998; Hayes & Ahrens, 1988) is likely to have important consequences for the quality of their orthographic and phonological representations, which in turn may influence word recognition. That is, assuming that difficult words are more often read than heard, accessing word meaning through the written representation may be less error prone compared to hearing the same words. Our current understanding of how word recognition accuracy is affected by the modality in which words are presented is limited. Moreover, it is unclear whether any modality effects on word recognition would be moderated by the words' difficulty and/or individuals' reading experience. Demonstrating effects of input modality on word recognition would have important implications for tools measuring receptive vocabulary size through tests of word recognition. That is, if word recognition accuracy were to differ as a function of modality, researchers developing tests of word recognition would need to consider carefully in which modality to present the test words. If, on the other hand, modality did not show effects

a florian.hintz@mpi.nl

on recognition accuracy, presentation modality would only have to play a minor role when designing new tests.

In the present study, we addressed these questions by re-analyzing a dataset that was collected in the context of developing a Dutch receptive vocabulary test (capitalizing on word recognition ability). Specifically, participants in that experiment had carried out a lexical decision task. They responded to words, ranging substantially in difficulty, presented in three modalities (spoken, written, or bimodal). The goal was to assess in which modality test words in the receptive vocabulary test should be presented. Moreover, there were two groups of participants who received different instructions ("Is this an existing Dutch word?" vs. "Do you know this Dutch word?") to assess potential task instruction effects on word recognition accuracy. Finally, in addition to the main experiment, participants had completed two tests assessing their receptive vocabulary size and exposure to literary texts, respectively. Thus, given the range of word difficulty, the three modality conditions and the additional individual-differences tests, the dataset was well-suited to address the present research questions centering around modality effects on word recognition accuracy and their potential moderators.

Background

Previous studies have reported word recognition benefits for visual and bimodal (simultaneous presentation of orthographic representation and spoken production of the phonological form) modalities compared to the auditory modality using a lexical decision task (Connine et al., 1990; Lopez Zunini et al., 2020; Turner et al., 1998). Responses have been found to be faster and more accurate for words presented in the visual and bimodal (audio-visual) modalities compared to the auditory modality. Note that these findings do not allow for generalizations on how modality affects word recognition accuracy as lexical decision tasks typically use words with a limited difficulty range such that responses (with reaction time as the main measure of interest) are assumed to index the speed with which a lexical entry is accessed. Difficult words are rarely used in lexical decision tasks (see Goldinger, 1996 for a review).

'Megastudies' in which large numbers of participants are tested (often via the internet) are an exception and have used difficult words in their lexical decision tasks. For example, Ferrand et al. (2018) assessed how much of the variance in word recognition accuracy and lexical decision latencies for written and spoken words was explained by word difficulty, operationalized as word frequency. They reported that in the visual modality, 20% of the variance in word recognition accuracy and 45% of the variance in lexical decision latencies was explained by word frequency. These estimates are in line with other reports that focused on the visual modality only. Other studies found that word frequency explained 15% to 49% of the variance in recognition accuracy and 21% to 49% of the variance in lexical decision latencies (Balota et al., 2004; Ferrand et al., 2010; Keuleers et al., 2010; Yap & Balota, 2009). Crucially, in the study by Ferrand et al., word frequency explained only a relatively small portion of variance in the spoken modality (7% and 13% of variance in recognition accuracy and lexical decision latencies, respectively). The strongest predictor of auditory lexical decision times was spoken word duration. One reason for the strong influence of word frequency on word recognition in the visual but not auditory modality could be – as explained above – that written text contains more infrequent words than spoken language (Cunningham & Stanovich, 1998; Hayes & Ahrens, 1988). Thus, language users are more likely to encounter less frequent words in the visual rather than the auditory modality.

Individuals differ substantially in the number and types of words they know (Mainz et al., 2017) and how often they engage in leisure reading (Gallik, 1999; Wift & Ander, 2017). It is likely that differences in receptive vocabulary size and exposure to literary texts influence the interaction between word difficulty and modality on word recognition accuracy.[1] The 'Lexical Quality Hypothesis' (LQH, Perfetti, 2007) holds that word recognition is more efficient, accurate and faster in individuals whose lexical representations are of high quality (Andrews, 2015; Elbro, 1996; Perfetti, 2007, 2011). Such high quality orthographic and phonological representations are precise, fully specified, with strong links between them, allowing for synchronous retrieval. Individuals with much reading experience are assumed to obtain high quality representations through a process called lexical tuning (Castles et al., 1999, 2007). In order to ensure accurate and fast lexical activation in an ever-expanding mental lexicon, lexical representations become more specific and precise, which improves inhibition of lexical competitors during word recognition (Andrews, 1997; Andrews & Hersch, 2010; Perfetti, 1992). Since the mental lexicon of experienced readers contains more and most likely more difficult words than that of inexperienced, infrequent readers, it is likely that their lexical mental representations are of higher quality, especially in the case of difficult words. Thus, experienced readers are likely to show better word recognition accuracy for difficult words compared to individuals with less reading experience. It is important to highlight that an individual's receptive vocabulary comprises multiple aspects, including one's ability to accurately recognize words in different modalities, as well as in-depth semantic knowledge about words. Though one would think that both are correlated (e.g., a person who recognizes many names of dog breeds might also have more in-depth knowledge about differences of dogs), they are not the same. The present work is concerned with word recognition ability.

---

1 Note that, based on previous research, we assume that individuals acquire large receptive vocabularies (especially knowledge about difficult words) through reading, as difficult words appear more often in written than spoken language (Cunningham & Stanovich, 1998). We therefore see both measures (receptive vocabulary size and exposure to literary texts) as reflecting an individual's reading experience.

The present study

By conducting the present re-analysis, we aimed to complement and extend previous reports on modality effects in word recognition. Specifically, we investigated (1) whether input modality had an effect on word recognition accuracy, (2) whether such a modality effect interacted with word difficulty, (3) whether there was an interaction between the effects of word difficulty and reading experience on word recognition accuracy, and (4) whether such an interaction was influenced by input modality.

The present dataset was in many respects similar to previous studies that had investigated modality effects on spoken word recognition: in a within-participants design, Dutch university students were presented with words and non-words in three modalities (auditory, visual, and bimodal) and were asked to carry out a binary decision task (e.g., lexical decision). However, there were also important methodological differences: as pointed out above, the words participants responded to varied substantially in word difficulty, which led to many more no-responses than in a typical lexical decision experiment. In a typical lexical decision experiment, researchers are predominantly interested in reaction times for words that are recognized correctly (yes-responses), and errors (i.e., no-responses for existing words) are attributed to momentary lapses of attention rather than lack of knowledge of the words. Thus, words are selected from a limited difficulty range to avoid data loss. The present dataset focused on recognition accuracy rather than speed, and, more importantly, modality effects on accuracy, which required the difficulty range to be much larger than in typical lexical decision tasks. That is, participants were presented with words they knew but also words they did not know or knew less well to avoid ceiling effects.

Relatedly, in contrast to previous studies, word difficulty was approximated using prevalence norms rather than word frequency values. Prevalence norms reflect the degree to which a word is known by the population: the word 'apple' is most likely known by 99+% of the population, whereas the proportion of people knowing the word 'phoneme' is substantially lower. According to Keuleers et al. (2015), prevalence norms provide a more realistic picture of a word's difficulty than frequency does. This is especially true for low-frequency words. For example, while the word 'academia' is probably recognized by the majority of English language users in the US, it rarely occurs in language corpora (i.e., with a frequency of one occurrence per one million words (Brysbaert et al., 2012). Keuleers et al. (2015) reported a medium-sized correlation (r = .35) between prevalence and word frequency (based on data from the Dutch Lexicon Project, Keuleers et al., 2010).

A final methodological difference to earlier studies was that half of the participants had received the standard instruction for a lexical decision task ("Indicate whether this is an existing Dutch word"), and the other half were instructed to "Indicate whether you know the word"—with the latter being a slightly more intuitive task and drawing less on meta-linguistic reasoning. This manipulation (as part of the efforts to develop the receptive vocabulary test) was implemented to test whether word recognition accuracy would vary as a function of task instruction.

In addition to the word recognition task, the participants had completed a receptive vocabulary test (Dutch version of the Peabody Picture Vocabulary Test Dunn & Dunn, 1997; Schlichting, 2005) and the Dutch version of the Author Recognition Test (Brysbaert et al., 2020) to assess exposure to literary texts. It is worth pointing out that even though the participants were university students, one may still expect substantial variation in how frequently individuals engage in literary reading in their leisure time (Acheson et al., 2008). That is, while course reading may contribute to how often students read and to the nature of the texts read, it is by no means the case that all students exhibit the same reading frequency. It was therefore important to include tests that gauge individuals' reading experience.

To re-cap, the present re-analysis investigated modality effects on word recognition accuracy and their potential moderators. Specifically, the first goal was to investigate whether the visual and audio-visual word recognition benefit reported in previous studies would hold when extending the difficulty range of stimulus words. The second goal was to test whether modality interacts with word difficulty such that, as words become more difficult, recognition accuracy is higher in the visual or bimodal compared to the auditory modality. This hypothesis was based on the observation that written text contains more difficult words than speech. The third goal was to test the hypothesis that individuals with larger receptive vocabularies and more exposure to literary texts, show better recognition accuracy of difficult words compared to individuals with less reading experience. Furthermore, as difficult words are more often encountered in written form, individuals with extensive reading experience and larger vocabularies may have a particular advantage when recognizing difficult words in the visual and audio-visual compared to the auditory modality. Thus, the fourth goal of the study was to test whether individuals with more reading experience, reflected in larger receptive vocabularies and more exposure to literary texts, show higher recognition accuracy than individuals with less experience, especially when these words are presented in the visual and audio-visual modality.

To anticipate the main results, none of the predictions concerning the impact of presentation modality was borne out. Instead, we found that word recognition accuracy depended only on the difficulty of the words and the individuals' exposure to literary texts.

Method

Participants

Forty-eight participants (39 female; age: 22.38 years old, SD = 1.78) had contributed to the present dataset. All participants were students at the Radboud University in Nijmegen and were native speakers of Dutch. They had normal or corrected-to-normal vision and hearing, and gave written informed consent prior to testing. Participants were paid for their participation. Half of the participants took part in Experiment 1a, the other half in Experiment 1b. Ethical approval to conduct the study was provided by the ethics committee of the Faculty of Social Sciences at Radboud University.

In addition to the three tests (word recognition experiment, Peabody Picture Vocabulary Test, Dutch Author Recognition Test) described here, all participants had also completed two auditory processing speed tests (Hintz et al., 2020) and Raven's Advanced Progressive Matrices test (Raven et al., 1998) in the context of the receptive vocabulary test development.

Test Materials and Procedure

Word recognition test. On each trial of the word recognition test, participants responded to a target word that was presented either visually, auditorily or bimodally (audio-visual). In Experiment 1a, participants were instructed to decide whether the word was an existing Dutch word or not. In Experiment 1b, participants were instructed to indicate whether they knew the presented target word. Participants were told that 'knowing a word' meant that they had previously encountered the word *and* had a vague idea of its meaning. In both sub-experiments, participants were informed that some of the presented targets were made-up non-words.

The selection of words was based on the prevalence database provided by Keuleers et al. (2015). This database contains prevalence measures for approximately 54,000 Dutch words, approximating to what extent each of these words is known to the whole population (i.e. ranging from < 5% to > 99%). Keuleers and colleagues established the prevalence values in a large-scale online study involving more than 360,000 unique participants. The participants performed an untimed lexical decision task on a randomly selected set of 100 words. The words were presented visually. The authors established item difficulty (i.e., prevalence) by applying item-response theory (i.e., fitting a Rasch model, Doran et al., 2007). Using these prevalence values, we selected 240 target words from the database by Keuleers et al. (2015). The mean prevalence for these words was 0.75 (SD = 0.09, range = 0.6 – 0.91).

The words for the present study were selected to have similar prevalence values across males and females and different age groups (younger adults, middle-aged individuals, older citizens). Plural forms, past tense forms of verbs, first person singular forms of verbs, and loanwords were not selected. The 240 words were divided evenly into three groups in a way that mean prevalence and range were matched precisely across groups (*M* = 0.75, range = 0.6 – 0.9). Furthermore, we selected 48 non-words, which were generated in Wuggy, a multilingual pseudoword generator (Keuleers & Brysbaert, 2010) and used in the mega-study by Keuleers et al. (2015). All of these non-words had an average accuracy (i.e., correct rejection rate) of at least 90%.

As for the words, we divided the selected non-words into three equal groups. Each group of 80 words was complemented with 16 non-words. The 96 targets in each group were rotated across the three modalities such that each participant was presented with each target only once. Trial presentation was blocked by modality. The order of word and non-word trials within each block was pseudo-randomized prior to the experiment. We counterbalanced the order of blocks across participants. Rotating each target across the three modalities and counterbalancing the order of modal-

ity blocks resulted in six experimental lists. Participants were randomly assigned to one list; each participant was presented with all 288 targets (240 words, 48 non-words, 96 per modality) on a given list.

Each trial started with a central fixation cross. Participants advanced by pressing a button. Following their button press, they either saw a visually presented target, heard an auditorily presented target or, on bimodal trials, saw and heard a target (visual and auditory presentation coincided). To parallel the visual trials, participants could listen to targets on auditory and bimodal trials as often as they wanted, just as they could look at the written target for as long as they wanted. They used the right control button on the keyboard to provide a 'this is a Dutch word/I know this word' response and the left control button to give a 'non-word/ I don't know this word' response. The task was untimed and participants could take short pauses between the modality blocks.

The dependent variable was word recognition accuracy (1 vs. 0). Our analyses, based on participants' average word recognition accuracy, showed that the data were neither skewed (-0.29) nor kurtotic (-0.59).

Peabody Picture Vocabulary Test (PPVT). Participants' receptive vocabulary size was assessed using a digitized version of the Dutch PPVT (Dunn & Dunn, 1997; Dutch translation by Schlichting, 2005). On each trial, participants first previewed four numbered line drawings on their screen. When they were ready, they pressed the Return key on their keyboard to hear the probe. They had to indicate which of the pictures best corresponded to the meaning of the spoken word by typing the corresponding number (1, 2, 3, or 4). Following the standard protocol for the test, items were presented in blocks of twelve items, with blocks increasing in difficulty. The starting level was 13, the best level participants could attain was 17. The test ended when a participant made nine or more errors within one block. Participants took, on average, twelve minutes to complete the test (range: 8 to 15 minutes). The participants' score was their raw score, that is, the serial number of their last item minus the number of errors made during the test. The maximum score was 204. Analyses, including participants from both sub-experiments, showed that the distribution of scores was neither skewed (-0.23), nor kurtotic (-0.11).

Dutch Author Recognition Test (DART). We used a pen-and-paper version of the Dutch Author Recognition Test, developed by Brysbaert et al. (2020), to measure reading frequency. The Author Recognition Test is a validated, recognized proxy measure of reading frequency (Acheson et al., 2008; Dabrowska, 2018; James et al., 2018; Mar & Rain, 2015; Payne et al., 2012; Stanovich & West, 1989). The underlying assumption is that the awareness level of authors' names increases as individuals read more often. In the test, participants were provided with a list of 132 names, divided into three columns of 44 names each. The 132 names were 90 names of Dutch and international fiction authors and 42 foils (names of non-authors). Brysbaert et al. (2020) had established the suitability of the material in multiple pretests, starting from a list of almost 15.000 fiction (book) authors. The final selection of 90 author names covers the whole difficulty spectrum, ranging from authors that are likely to be known by a large proportion of individuals to

authors that are likely to be known only by frequent readers of fiction. The order of author and foil names was random and was the same for each participant. Participants' task was to indicate which of the listed names were authors. Participants' score was the proportion of correctly identified author names minus the proportion of incorrectly selected foils. The maximum score was 1. Analyses, including participants from both sub-experiments, showed that the distribution of DART scores was moderately skewed (1.16) and kurtotic (1.28). Overall, the scores were on the lower end of the performance spectrum suggesting that the test was fairly difficult.

Results

Table 1 summarizes participants' scores on the PPVT and DART. Means, standard deviations (SDs) and ranges were very similar in Experiment 1a and 1b. Importantly, SDs and ranges suggested quite some variability across participants. PPVT and DART were moderately correlated ($r = .56$) such that participants with larger receptive vocabularies were also frequent readers (i.e., knew more authors).

Word Recognition Test

False alarm rate (i.e., the proportion of 'Yes-responses' to non-words) was, on average, 8% ($SD = 16\%$, range = 2%-100%; M Experiment 1a = 5%, M, Experiment 1b = 11%). One participant from Experiment 1b was excluded from all analyses because they had a false alarm rate of 100%, which means they responded "Yes, I know this word" to all non-words. This suggested that they did not take the test seriously or had not understood the task. With the removal of that participant, the false alarm rate dropped to 5% ($SD = 4\%$, range = 2%-19%). Overall, participants found it easy to recognize the non-existing words (high correct rejection and low false alarm rates). This was the case for all three modality conditions (see also Figure 1).

Table 2 depicts the mean word recognition accuracy by modality condition. Overall word recognition accuracy was 49%. The means suggest there was little difference between auditory, visual and audio-visual modalities. Participants in Experiment 1b were numerically slightly less accurate than participants in Experiment 1a. Figure 2 plots word recognition accuracy as a function of word difficulty. It is important to highlight that the prevalence scores denoted on the x-axis of Figure 2 are not equivalent to 'word recognition accuracy'. Instead, these values were obtained by Keuleers et al. (2015) by applying item-response theory (i.e., a Rasch model). Though the recognition scores were overall lower than expected on the basis of the norming data, the figures shows that there was a strong relationship between the two data sets. The correlations of the recognition scores in the three modality with the prevalence values were r = .66 (bimodal), r = .62 (audio-only) and r =.69 (visual-only), respectively.

Recognition accuracy was analyzed using Bayesian logistic mixed-effects modelling in R (R Development Core
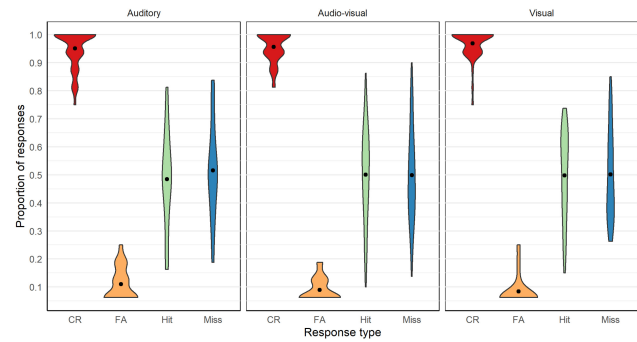
Figure 1. Proportion of correct rejections and false alarms for nonwords, and hits and misses for words by modality

CR = correct rejection, FA = false alarm.

Team, 2008, version 3.6.2, 2019), using the brms package (Burkner, 2017). Analyses were conducted on responses to words. Bayesian analyses are concerned with the likely magnitude of effects rather than statistical significance. Effects were considered meaningful when the 95% Credible Intervals (CI) did not contain zero, which indicates that the parameter has a non-zero effect with high certainty. Moreover, effects were considered meaningful if the point estimate was about twice the size of its error, indicating that the estimated effect is large compared to the uncertainty around it. The posterior probability is reported for these effects, which indicates the proportion of samples with a value equal to or more extreme than the estimate. In addition, Bayes Factors (BF10, BF01) were calculated for all effects, which give an indication of the relative evidence for the alternative hypothesis (H1) compared to the null hypothesis (H0) or vice versa. Our interpretation of the Bayes Factors followed the guidelines by Jeffreys (1961), where a BF of 1-3 can be interpreted as anecdotal evidence, a BF of 3-10 as substantial evidence and a BF of >10 of strong evidence for or against the null/alternative hypothesis. Note that BF10 indicates a Bayes factor that favors H1 over H0, and BF01 indicates a Bayes factor in favor of H0 over H1. The model had four chains of 8000 iterations each, with the first half representing a warm-up period. A weak prior (Cauchy distribution with center 0 and scale 2.5 using a sampling algorithm) was used, as is appropriate for non-hierarchical logistic regression models (Gelman et al., 2008). Models were run until the R-hat value for each parameter was 1.00, indicating full convergence. Modality was contrast-coded based on simple contrasts, with the auditory modality being the reference level in the first model, and the audio-visual modality being the reference in the second model. With simple contrast coding, the reference level is always coded as −1/3, and the level that it is compared to is coded as 2/3. This way of coding is similar to treatment contrast coding, but has the advantage that the intercept corresponds to the grand mean instead of corresponding to the mean of the reference level. Moreover, factors outside of interactions can be interpreted as main effects.

Table 1. Participants' test results on PPVT and DART in Experiment 1a and Experiment 1b

Task	Experiment 1a					Experiment 1b				
	Mean	SD	Range	Skewness	Kurtosis	Mean	SD	Range	Skewness	Kurtosis
PPVT	178.54	9.49	160-198	0.19	-0.60	178.13	8.82	155-191	-0.76	0.14
DART	26.79	14.19	10-63	1.04	0.38	21.91	9.50	7-41	0.53	-0.78

Note. Range is rounded up for brevity. PPVT = Peabody Picture Vocabulary Test, DART = Dutch Author Recognition Test.

Table 2. Word recognition accuracy by experiment and modality

|	Overall	Modality		
		Auditory	Audio-visual	Written
Experiment 1a	.52 (.17)	.52 (.17)	.53 (.17)	.52 (.17)
Experiment 1b	.46 (.17)	45 (.16)	.47 (.16)	.48 (.16)
Average	.49 (.16)	.48 (.17)	.50 .(17)	.50 (.16)

Note. Standard deviation displayed in brackets.

The models contained Modality (auditory vs. visual vs. audio-visual) as a fixed factor. Word Difficulty was scaled and centered and added to the model as continuous predictor. Participants' PPVT and DART scores were centered and scaled and added to the model as continuous predictors. Because both sub-experiments differed in task, we included Task Version as a fixed factor to model the difference in task instruction between the participants. Based on our hypotheses, interactions between Modality, Word Difficulty and PPVT/DART were added to the model. Furthermore, we added interactions between Task Version, Modality and Word Difficulty to test whether Task Version affected the modality effect or the interaction effect between input modality and word difficulty. The random effect structure included random intercepts by word and participant and random slopes for modality by word and participant. The model formula was thus: brm(Correct ~ (Modality * cWord_Difficulty) * (Task_Version + cPPVT + cDART) + (1 + modality | PP_nr) + (1 + modality | Word), family = bernoulli, data = all_Data, chains=4, cores = 2, iter = 8000, prior = Pr1).

The full model output for the model with the spoken modality as the reference level is displayed in Table 3 and the model output for the model with the audio-visual modality as the reference level is displayed in Table 4. As to be expected, we observed strong evidence for a main effect of Word Difficulty with easier (i.e., more prevalent) words leading to more correct responses than difficult words. We observed no evidence for a main effect of Modality. In fact, the Bayes factors suggested substantial evidence in favor of the null hypothesis (BF01 > 10). Similarly, we did not observe main effects of Task Version (BF01 = 6.67 – 7.69). None of the interactions involving modality showed a significant effect—all of which showed strong evidence in favor of the null hypothesis. Furthermore, the models revealed that Word Difficulty interacted with Task version, PPVT, and DART. However, the Bayes factors showed that there was substantial evidence only for the last mentioned interaction. It suggests that frequent readers performed better than less frequent ones in particular for difficult words (Figure 3).

Discussion

The present study investigated whether input modality had an effect on word recognition accuracy, whether this modality effect interacted with word difficulty, whether there was an interaction between word difficulty and reading experience on word recognition accuracy, and whether these interactions were influenced by input modality. To
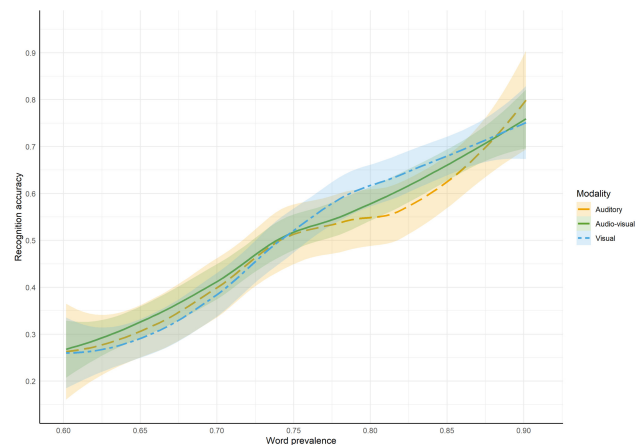
Figure 2. Smoothed (using loess regression) word recognition accuracy split out by modality and word difficulty

Bands indicate 95% confidence bands around the predicted values of the loess regression.
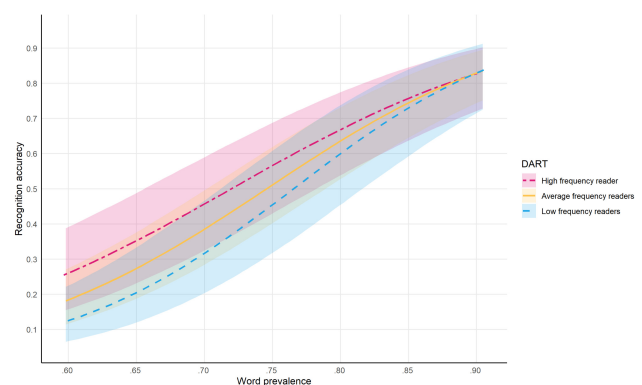
Figure 3. Predicted effect of word difficulty (prevalence) and DART scores on recognition accuracy

The shaded areas represents the 95% credible intervals.

address these questions, we re-analyzed a dataset collected in the context of the development of a vocabulary test.

Our first goal was to examine how word recognition accuracy would be affected by the modality of word presentation. We hypothesized, in line with previous literature on modality effects in word recognition (Connine et al., 1990; Lopez Zunini et al., 2020; Turner et al., 1998), that word

Table 3. Full model output for the model with the spoken modality as reference level

Predictor	Estimate	*SE*	95% credible interval	BF10	BF01
(Intercept)	0.08	0.20	-0.30, 0.47	0.02	50
Modality: audio-visual	0.07	0.12	-0.16, 0.30	0.04	25
Modality: written	-0.01	0.13	-0.28, 0.24	0.04	25
Word Difficulty	**0.92**	**0.06**	**0.80, 1.03**	**1.53e+13**	**6.54e-09**
Task Version	-0.25	0.27	-0.78, 0.29	0.15	6.67
PPVT	0.00	0.16	-0.32, 0.32	0.05	20
DART	0.20	0.16	-0.13, 0.52	0.11	9.09
Modality: audio-visual * Word Difficulty	-0.02	0.09	-0.20, 0.16	0.03	33.33
Modality: written * Word Difficulty	0.02	0.09	-0.16, 0.21	0.03	33.33
Modality: audio-visual * Task Version	0.05	0.16	-0.27, 0.36	0.05	20
Modality: written * Task Version	0.19	0.19	-0.17, 0.56	0.10	10
Modality: audio-visual * PPVT	0.11	0.09	-0.08, 0.29	0.06	16.67
Modality: written PPVT	0.00	0.11	-0.23, 0.21	0.03	33.33
Modality: audio-visual * DART	-0.01	0.10	-0.20, 0.18	0.03	33.33
Modality: written * DART	-0.01	0.11	-0.23, 0.21	0.04	25
Word Difficulty * Task Version	**-0.12**	**0.05**	**-0.21, -0.02**	**0.22**	**4.55**
Word Difficulty * PPVT	**0.09**	**0.03**	**0.03, 0.15**	**1.08**	**0.93**
Word Difficulty * DART	**-0.11**	**0.03**	**-0.17, -0.05**	**3.78**	**0.26**
Modality: audio-visual * Word Difficulty * Task version	-0.06	0.12	-0.29, 0.18	0.04	25
Modality: written * Word Difficulty * Task version	0.00	0.12	-0.24, 0.23	0.04	25
Modality: audio-visual * Word Difficulty * PPVT	0.07	0.07	-0.07, 0.21	0.04	25
Modality: written * Word Difficulty * PPVT	0.07	0.07	-0.07, 0.21	0.04	25
Modality: audio-visual * Word Difficulty * DART	-0.02	0.07	-0.16, 0.13	0.02	50
Modality: written * Word Difficulty * DART	0.10	0.07	-0.04, 0.24	0.06	16.67

Note. Meaningful effects are displayed in bold. PPVT = Peabody Picture Vocabulary Test. DART = Dutch Author Recognition Test.

recognition accuracy would be higher when words are presented in the visual or bimodal compared to the auditory modality. Our Bayesian analyses did not confirm this hypothesis. An explanation for this may lie in methodological differences between the present dataset and previous experiments. For example, in order to avoid ceiling effects in accuracy, the words in the present dataset varied much more in word difficulty than the stimulus words selected for standard lexical decision tasks. Extremely difficult words are typically avoided to reduce loss of data due to high error rates. Consequently, errors in traditional word recognition paradigms mostly indicate momentary failures of attention when participants respond to words that they are expected to know. By contrast, in our study, errors most likely indicated that the participants did not know the word. Moreover, unlike in standard lexical decision tasks, responses in the present study were untimed. Time-pressure might be crucial for seeing modality effects. In the visual modality, the entire word is immediately available to the cognitive processing systems (c.f. Coltheart et al., 2001), whereas in the auditory modality the same information becomes available in an incremental fashion (c.f. Marslen-Wilson &

Tyler, 1980; McClelland & Elman, 1986; Norris et al., 2000). The fact that a word's constituents are available all at once in the visual modality might have led to modality effects in traditional, timed lexical decision tasks where participants respond as quickly as possible. Our results suggest that modality is of less importance in untimed lexical decision tasks, where participants are instructed to consider carefully whether they know the target word or not. Our study was conducted in Dutch, and some of the results discussed here may be language-specific. However, this conclusion – that timed responses might be more sensitive to modality effects than untimed ones – should hold for other languages as well.

Our second goal was to investigate the interaction of modality and word difficulty on word recognition accuracy. We hypothesized that, as words became more difficult, recognition accuracy would be increasingly higher in the visual and bimodal modality than the auditory modality. Arguably, difficult words are more often encountered in the written form and consequently orthographic representations were predicted to be of higher quality than the phonological representations of the same words. However, this

Table 4. Full model output for the model with the audio-visual modality as reference level

Predictor	Estimate	SE	95% credible interval	BF10	BF01
(Intercept)	0.08	0.19	-0.29, 0.45	0.02	50
Modality: spoken	-0.07	0.11	-0.29, 0.16	0.05	20
Modality: written	-0.09	0.13	-0.34, 0.17	0.05	20
Word Difficulty	**0.92**	**0.06**	**0.80, 1.03**	**4.00e+14**	**2.50e-15**
Task Version	-0.25	0.27	-0.79, 0.28	0.13	7.69
PPVT	0.00	0.16	-0.31, 0.31	0.06	16.67
DART	0.20	0.16	-0.12, 0.52	0.12	8.33
Modality: spoken * Word Difficulty	0.03	0.09	-0.15, 0.21	0.03	33.33
Modality: written * Word Difficulty	0.05	0.09	-0.12, 0.21	0.03	33.33
Modality: spoken * Task Version	-0.06	0.15	-0.35, 0.25	0.05	20
Modality: written * Task Version	0.14	0.19	-0.23, 0.51	0.08	12.5
Modality: spoken * PPVT	-0.11	0.09	-0.29, 0.08	0.06	16.67
Modality: written PPVT	-0.11	0.11	-0.34, 0,12	0.06	16.67
Modality: spoken * DART	0.02	0.09	-0.17, 0.20	0.03	33.33
Modality: written * DART	0.00	0.11	-0.22, 0.23	0.03	33.33
Word Difficulty * Task Version	**-0.12**	**0.05**	**-0.21, -0.02**	**0.25**	**4**
Word Difficulty * PPVT	**0.09**	**0.03**	**0.04, 0.15**	**1.68**	**0.6**
Word Difficulty * DART	**-0.11**	**0.03**	**-0.17, -0.05**	**3.75**	**0.27**
Modality: spoken * Word Difficulty * Task version	0.05	0.12	-0.18, 0.28	0.04	25
Modality: written * Word Difficulty * Task version	0.06	0.12	-0.17, 0.29	0.04	25
Modality: spoken * Word Difficulty * PPVT	-0.07	0.07	-0.21, 0.07	0.04	25
Modality: written * Word Difficulty * PPVT	0.00	0.07	-0.14, 0.14	0.04	25
Modality: spoken * Word Difficulty * DART	0.01	0.07	-0.13, 0.16	0.02	50
Modality: written * Word Difficulty * DART	0.11	0.07	-0.03, 0.25	0.08	12.5

Note. Meaningful effects are displayed in bold. PPVT = Peabody Picture Vocabulary Test. DART = Dutch Author Recognition Test.

hypothesis was not supported by our findings: There was no significant interaction between difficulty and modality. This may indicate that, even though difficult words are most likely to be encountered in the written modality, their phonological representations might be just as precise and complete as those of easier words. Theories of reading aloud (Coltheart et al., 2001) and reading acquisition (Ehri, 1995; Shankweiler, 1999; Share, 1995) propose a mechanism that describes how phonological representations are created from written input. During recoding, readers mentally recode the graphemes into phonemes upon a written encounter with a novel word, thereby creating both an orthographic and phonological representation of the novel word. Such a mechanism might work specifically well in transparent languages, such as Dutch where graphemes generally map one-to-one onto a phonemes (Seymour et al., 2003). It is conceivable that recoding is less efficient in opaque languages, such as English, where grapheme-phoneme correspondences are more unreliable. Moreover, this explanation may also especially apply to the sample tested in the present study. Our participants were university students with no deficiencies in the linguistic domain. Our findings may

not generalize to individuals with language or reading disabilities, or individuals with weak grapheme-phoneme correspondences. For these groups, one might find a general advantage of auditory or audio-visual over written presentation or a specific advantage for harder words.

The third goal of the study was to investigate the interaction between word difficulty and individual differences in receptive vocabulary size and exposure to literacy texts on word recognition accuracy. We expected, and found, that the indicators of vocabulary size, the PPVT score, and of reading experience, the DART score, were correlated (r =.56). This correlation most likely arose because written texts are likely to use a varied vocabulary, including low-prevalence words. Thus, frequent reading enriches a person's vocabulary. We predicted that both variables, PPVT and DART should predict word recognition scores, especially for low-prevalence scores. This is because the high-prevalence words should be included in most individuals' vocabularies, whereas the low-prevalence words should be more likely to be included in the vocabularies of individuals with larger receptive vocabularies and more exposure to literary texts.

With respect to the PPVT scores, our prediction that individuals with high PPVT scores would show an accuracy advantage for difficult words over individuals with low PPVT scores was not borne out. The models revealed a statistically significant interaction between PPVT and word difficulty (participants with larger PPVT scores recognized easier words more accurately than participants with lower scores), however, the Bayes factors suggested that there was at best anecdotal evidence for this effect. Given that the Dutch version of the PPVT has been shown to predict adults' word recognition performance in other studies (e.g., Hintz et al., 2020), this result is unexpected and so far unexplained.

For the DART scores, we obtained evidence for the expected interaction. We indeed observed that participants who read frequently (i.e., knew more authors) recognized more difficult words than participants who read less often. This finding corroborates the idea that increased exposure to novel words fine-tunes lexical representations (Castles et al., 1999, 2007) and that these high quality representations improve word recognition (Perfetti, 2007) by increasing the speed and accuracy of word recognition (Andrews, 1997; Andrews & Hersch, 2010; Perfetti, 1992).

The fourth goal of the present study was to investigate whether the interaction between word difficulty and reading experience on word recognition accuracy was influenced by modality. We predicted that experienced readers, compared to individuals, who are less experienced, would show increased word recognition accuracy of difficult words, especially when these words are presented in the visual and bimodal modality, as difficult words are most often encountered in the written form. Our results did not provide any evidence for this three-way interaction. A possible explanation may be that, as discussed above, it is possible to create phonological representations of difficult words that are sufficiently precise and accurate to recognize this word in its spoken form efficiently, regardless of reading experience. This explanation, however, may only pertain to transparent languages, such as Dutch, and populations similar to the sample in the present study, which consisted of highly literate university students without any language or reading disabilities. Investigating the modality effect in other languages and in samples with a larger range of language and reading abilities may be important avenues for future research.

A final goal of the study was to explore the effects of different instructions on the participants' word recognition scores. We found that asking participants "Is this an existing word?" versus "Do you know this word?" had no significant influence on their word recognition accuracy.

Though the primary goals of the study concerned the effects of presentation modality, it also offers the opportunity to explore the merit of using prevalence, rather than frequency, as an indicator of word difficulty. We opted for varying prevalence because recent studies had shown that prevalence explained about 7% of additional variance on top of the variance explained by frequency in word recognition tasks. Moreover, criticism has been expressed about the validity of frequency norms for difficult words (Brysbaert et al., 2016; Keuleers et al., 2015). That is, some words with a low frequency of occurrence may not be difficult to recog-

nize, as they are known to a large part of the population. Our data confirmed that prevalence indeed predicted word recognition accuracy, especially for low prevalence words. Therefore, the present study may also be seen as a small-scale validation of the prevalence norms, as it demonstrated the predictive value of the norms in different modalities.

An obvious question is whether prevalence was a better predictor of word recognition accuracy than word frequency. It is important to stress that our study was not designed with this question in mind. Nonetheless, we performed several complementary analyses to explore this issue. We used Google Books to establish the word frequencies for our materials. Search options were set to occurrences in the Dutch language, within Dutch internet pages and restricted to a time window of January 1 1995 to January 1 2020. The words had a mean frequency of 1468 raw occurrences in Google Books (SD = 1836, range = 4 – 12700 occurrences). Frequencies were log-transformed, and correlated with the prevalence values. We found no significant correlation between prevalence and Google frequency (Pearson correlation: n = 240, r = 0.06, p = .36, Spearman rank correlation: n = 240, r = 0.07, p = .30). This is unexpected as Keuleers et al. (2015) reported a medium-sized correlation (n ~ 14.000, r = .35, based on data from the Dutch lexicon project, Keuleers et al., 2010) of prevalence and frequency. Note, however, that their correlation was based on a different prevalence database than the one we used for the present study. Moreover, we only used a small subset (n = 240) of the 54.000 words listed in Keuleers et al. (2015). More importantly, recognition accuracy did not correlate with Google frequency (r = 0.06, p = .35). This contrasts with the strong correlation between recognition accuracy and prevalence (r = 0.73, p < .001). We re-ran the Bayesian models described above (Tables 3, 4), replacing prevalence with Google frequency. There was no evidence for a main effect of Google frequency (estimate = 0.12, SE = 0.08, 95% CI = [-0.03, 0.28]), nor any interaction effects with the other predictors, except anecdotal evidence for an interaction with Task Version. A comparison of the models (using the WA and LOO information criteria) showed that replacing prevalence with frequency decreased model fit as reflected in larger LOOIC and WAIC values (model with prevalence predictor: LOOIC = 12410.34, WAIC = 12408.22; model with frequency predictor: LOOIC = 12452.45, WAIC = 12449.98).

Thus, in our study word recognition accuracy was predicted by prevalence, but not by Google frequency. To reiterate, our study was not designed to assess the effects of word frequency and we do not wish to claim that frequency can never have an impact on word recognition. There is, of course, a large body of work clearly demonstrating the influence of word frequency on the speed and accuracy of lexical access in word comprehension tasks (Brysbaert et al., 2018 for review). However, it is not known how influential prevalence would be in the same tasks. Important goals for further research would be to develop prevalence norms for other languages than Dutch and to explore and contrast the impact of prevalence and frequency in different linguistic tasks (see Brysbaert et al., 2019, for prevalence norms for 62,000 English words). Frequency and prevalence norms provide complementary information, one telling us how well represented words are in a corpus, the other

telling us how well they are represented in the minds of a panel of speakers of the language. High prevalence words are probably recognized by many because they appear often in written and spoken language. Low prevalence words (often technical, political terms), on the other hand, are more likely to be acquired through reading. Each way of garnering information, from corpora or via meta-linguistic judgements, has advantages and disadvantages, and consequently the usefulness of the information will depend on the investigator's research goals.

In sum, we found no evidence that the modality of input affected word recognition in Dutch. This held regardless of word difficulty and participants' reading experience. This lack of a modality effect suggests that word knowledge, more specifically individuals' ability to recognize words, can be assessed equally well in the written and spoken modality. However, we wish to stress again that we tested speakers of an orthographically highly transparent language, and that the participants were university students. We cannot rule out that input modality matters for assessments of word recognition ability in less transparent languages and, perhaps more importantly, for assessments of participants with overall lower levels of reading experience or skills.

Acknowledgements

Funding

Competing interests

The authors declare no competing interests.

Contributions

Substantial contributions to conception and design: ASM, FH; Acquisition of data: FH; Analysis and interpretation of data: MW, FH; Drafting the article or revising it critically for important intellectual content: MW, ASM, CFR, FH; Final approval of the version to be published: MW, ASM, CFR, FH.

Data accessibility statement

The data underlying the results presented in the study are available at the Archive of the Max Planck Institute for Psycholinguistics (Nijmegen, NL): https://hdl.handle.net/1839/cf88a63e-e413-49a1-8c00-3988df63e9b8. By signing the consent form, our participants explicitly agreed that their anonymized data may only be shared with other academics and for academic purposes only. This aspect of the consent form was a requirement of the board that approved the ethics application. That is, the present study was covered by an 'umbrella' that provided ethical approval for a larger research program involving the collection and analysis of genetic material from participants. To comply with the requirements, we included the above restriction, which applies to sharing all data acquired within the larger research program. Interested researchers need to create an account with the Archive of the Max Planck Institute for Psycholinguistics providing a user name, email address, their full name and affiliation. Alternatively, in case their institution is part of one of the supported Identity Federations (Shibboleth), which is the case for many academic/research institutions, interested individuals may simply use their own institutional account to log in.

REFERENCES

Acheson, D. J., Wellu, J. B., & MacDonald, M. C. (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior Research Methods*, *40*(1), 278–289. https://doi.org/10.3758/brm.40.1.278

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, *4*(4), 439–461. https://doi.org/10.3758/Bf03214334

Andrews, S. (2015). Individual differences among skilled readers: The role of lexical quality. In A. Pollatsek & R. Treiman (Eds.), *The Oxford handbook of reading.* (pp. 129–148). Oxford University Press.

Andrews, S., & Hersch, J. (2010). Lexical precision in skilled readers: Individual differences in masked neighbor priming. *Journal of Experimental Psychology: General*, *139*(2), 299–318. https://doi.org/10.1037/a0018366

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*(2), 283–316. https://doi.org/10.1037/0096-3445.133.2.283

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, *27*(1), 45–50. https://doi.org/10.1177/0963721417727521

Brysbaert, M., Mandera, P., McCormick, S. F., & Keuleers, E. (2019). Word prevalence norms for 62,000 English lemmas. *Behavior Research Methods*, *51*(2), 467–479. https://doi.org/10.3758/s13428-018-1077-9

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods*, *44*(4), 991–997. https://doi.org/10.3758/s13428-012-0190-4

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: Evidence from the Dutch Lexicon Project 2. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(3), 441–458. https://doi.org/10.1037/xhp0000159

Brysbaert, M., Sui, L., Dirix, N., & Hintz, F. (2020). Dutch Author Recognition Test. *J Cogn*, *3*(1), 6. https://doi.org/10.5334/joc.95

Burkner, P. C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Castles, A., Davis, C., Cavalot, P., & Forster, K. (2007). Tracking the acquisition of orthographic skills in developing readers: masked priming effects. *Journal of Experimental Child Psychology*, *97*(3), 165–182. https://doi.org/10.1016/j.jecp.2007.01.006

Castles, A., Davis, C., & Letcher, T. (1999). Neighbourhood effects on masked form priming in developing readers. *Language and Cognitive Processes*, *14*(2), 201–224. https://doi.org/10.1080/016909699386347

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256. https://doi.org/10.1037/0033-295x.108.1.204

Connine, C. M., Mullennix, J., Shernoff, E., & Yelen, J. (1990). Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(6), 1084–1096. https://doi.org/10.1037//0278-7393.16.6.1084

Cunningham, A. E., & Stanovich, K. E. (1998). What reading does for the mind. *American Educator*, *22*, 8–17.

Dabrowska, E. (2018). Experience, aptitude and individual differences in native language ultimate attainment. *Cognition*, *178*, 222–235. https://doi.org/10.1016/j.cognition.2018.05.018

Doran, H., Bates, D., Bliese, P., & Dowling, M. (2007). Estimating the multilevel Rasch model: With the lme4 package. *Journal of Statistical Software*, *20*(2), 1–18. https://doi.org/10.18637/jss.v020.i02

Dunn, L., & Dunn, L. (1997). *Peabody Picture Vocabulary Test (3rd ed.)*. American Guidance Service.

Ehri, L. C. (1995). Phases of development in learning to read words by sight. *Journal of Research in Reading*, *18*(2), 116–125. https://doi.org/10.1111/j.1467-9817.1995.tb00077.x

Elbro, C. (1996). Early linguistic abilities and reading development: A review and a hypothesis. *Reading and Writing*, *8*(6), 453–485. https://doi.org/10.1007/Bf00577023

Ferrand, L., Meot, A., Spinelli, E., New, B., Pallier, C., Bonin, P., Dufau, S., Mathot, S., & Grainger, J. (2018). MEGALEX: A megastudy of visual and auditory word recognition. *Behavior Research Methods*, *50*(3), 1285–1307. https://doi.org/10.3758/s13428-017-0943-1

Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Meot, A., Augustinova, M., & Pallier, C. (2010). The French Lexicon Project: lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods*, *42*(2), 488–496. https://doi.org/10.3758/BRM.42.2.488

Gallik, J. D. (1999). Do They Read for Pleasure? Recreational Reading Habits of College Students. *Journal of Adolescent & Adult Literacy*, *42*(6), 480–488.

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y. S. (2008). A Weakly Informative Default Prior Distribution for Logistic and Other Regression Models. *Annals of Applied Statistics*, *2*(4), 1360–1383. https://doi.org/10.1214/08-Aoas191

Goldinger, S. D. (1996). Auditory lexical decision. *Language and Cognitive Processes*, *11*(6), 559–568.

Hayes, D. P., & Ahrens, M. G. (1988). Vocabulary simplification for children: a special case of "motherese"? *Journal of Child Language*, *15*(2), 395–410. https://doi.org/10.1017/s0305000900012411

Hintz, F., Jongman, S. R., Dijkhuis, M., van 't Hoff, V., McQueen, J. M., & Meyer, A. S. (2020). Shared lexical access processes in speaking and listening? An individual differences study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(6), 1048–1063. https://doi.org/10.1037/xlm0000768

James, A. N., Fraundorf, S. H., Lee, E. K., & Watson, D. G. (2018). Individual differences in syntactic processing: Is there evidence for reader-text interactions? *J Mem Lang*, *102*, 155–181. https://doi.org/10.1016/j.jml.2018.05.006

Jeffreys, H. (1961). *Theory of Probability* (3rd ed.). Oxford University Press.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: a multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627–633. https://doi.org/10.3758/BRM.42.3.627

Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice effects in large-scale visual word recognition studies: a lexical decision study on 14,000 dutch mono- and disyllabic words and nonwords. *Frontiers in Psychology*, *1*(174), 174. https://doi.org/10.3389/fpsyg.2010.00174

Keuleers, E., Stevens, M., Mandera, P., & Brysbaert, M. (2015). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *Quarterly Journal of Experimental Psychology*, *68*(8), 1665–1692. https://doi.org/10.1080/17470218.2015.1022560

Lopez Zunini, R. A., Baart, M., Samuel, A. G., & Armstrong, B. C. (2020). Lexical access versus lexical decision processes for auditory, visual, and audiovisual items: Insights from behavioral and neural measures. *Neuropsychologia*, *137*, 107305. https://doi.org/10.1016/j.neuropsychologia.2019.107305

Mainz, N., Shao, Z., Brysbaert, M., & Meyer, A. S. (2017). Vocabulary Knowledge Predicts Lexical Processing: Evidence from a Group of Participants with Diverse Educational Backgrounds. *Frontiers in Psychology*, *8*(1164), 1164. https://doi.org/10.3389/fpsyg.2017.01164

Mar, R. A., & Rain, M. (2015). Narrative Fiction and Expository Nonfiction Differentially Predict Verbal Ability. *Scientific Studies of Reading*, *19*(6), 419–433. https://doi.org/10.1080/10888438.2015.1069296

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*(1), 1–71. https://doi.org/10.1016/0010-0277(80)90015-3

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*, 299–325.

Payne, B. R., Gao, X., Noh, S. R., Anderson, C. J., & Stine-Morrow, E. A. (2012). The effects of print exposure on sentence processing and memory in older adults: Evidence for efficiency and reserve. *Neuropsychology, Development, and Cognition. Section B: Aging, Neuropsychology and Cognition*, *19*(1–2), 122–149. https://doi.org/10.1080/13825585.2011.628376

Perfetti, C. (1992). The representation problem in reading acquisition. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 145–174). Lawrence Erlbaum.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*(4), 357–383. https://doi.org/10.1080/10888430701530730

Perfetti, C. (2011). Phonology Is Critical in Reading. In S. A. Brady, D. Braze, & C. A. Fowler (Eds.), *Explaining Individual Differences in Reading: Theory and Evidence* (pp. 153–175). Psychology Press.

R Development Core Team. (2008). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Raven, J. C., Court, J. H., & Raven, J. E. (1998). *Raven's coloured progressive matrices*. Harcourt Assessment.

Schlichting, J. E. P. T. (2005). *Peabody picture vocabulary test-III-NL*. Harcourt Test Publishers.

Seymour, P. H., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*(Pt 2), 143–174. https://doi.org/10.1348/000712603321661859

Shankweiler, D. (1999). Words to Meanings. *Scientific Studies of Reading*, *3*(2), 112–127. https://doi.org/10.1207/s1532799xssr0302_2

Share, D. L. (1995). Phonological recoding and self-teaching: sine qua non of reading acquisition. *Cognition*, *55*(2), 151–218; discussion 219-26. https://doi.org/10.1016/0010-0277(94)00645-2

Stanovich, K. E., & West, R. F. (1989). Exposure to Print and Orthographic Processing. *Reading Research Quarterly*, *24*(4), 402–433. https://doi.org/10.2307/747605

Turner, J. E., Valentine, T., & Ellis, A. W. (1998). Contrasting effects of age of acquisition and word frequency on auditory and visual lexical decision. *Memory and Cognition*, *26*(6), 1282–1291. https://doi.org/10.3758/bf03201200

Wift, A., & Ander, S. (2017). *Rumors of the Demise of Books Greatly Exaggerated*. Gallup. https://news.gallup.com/poll/201644/rumors-demise-books-greatly-exaggerated.aspx

Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language*, *60*(4), 502–529. https://doi.org/10.1016/j.jml.2009.02.001

SUPPLEMENTARY MATERIALS

Peer review history

Download: https://collabra.scholasticahq.com/article/24919-the-effects-of-input-modality-word-difficulty-and-reading-experience-on-word-recognition-accuracy/attachment/63354.docx?auth_token=dx33Pqx-WhGX43mSXv-k