



**HAL**  
open science

## **MEGALEX: A megastudy of visual and auditory word recognition**

Ludovic Ferrand, Alain Méot, Elsa Spinelli, Boris New, Christophe Pallier, Patrick Bonin, Stéphane Dufau, Sebastiaan Mathôt, Jonathan Grainger

► **To cite this version:**

Ludovic Ferrand, Alain Méot, Elsa Spinelli, Boris New, Christophe Pallier, et al.. MEGALEX: A megastudy of visual and auditory word recognition. *Behavior Research Methods*, 2018, 50 (3), pp.1285-1307. <10.3758/s13428-017-0943-1>. <hal-02324242>

**HAL Id: hal-02324242**

**<https://hal.science/hal-02324242v1>**

Submitted on 29 Oct 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

## **MEGALEX: A Megastudy of Visual and Auditory Word Recognition**

Ludovic Ferrand<sup>1</sup>, Alain Méot<sup>1</sup>, Elsa Spinelli<sup>2</sup>, Boris New<sup>3</sup>, Christophe Pallier<sup>4</sup>, Patrick Bonin<sup>5</sup>, Stéphane Dufau<sup>6</sup>, Sebastiaan Mathôt<sup>6</sup>, and Jonathan Grainger<sup>6</sup>

<sup>1</sup>*Université Clermont Auvergne, CNRS, LAPSCO, F-63000 Clermont-Ferrand, France*

<sup>2</sup>*CNRS and Université Pierre Mendès-France, LPNC UMR 5105, Grenoble, France*

<sup>3</sup>*CNRS and Université Savoie Mont Blanc, LPNC UMR 5105, Chambéry, France*

<sup>4</sup>*INSERM, CEA, U 992, Neurospin, Gif-sur-Yvette, France*

<sup>5</sup>*CNRS and Université de Bourgogne Franche-Comté, LEAD UMR 5022, Dijon, France*

<sup>6</sup>*CNRS and Aix-Marseille University, LPC UMR 7290, France*

Running-head: A megastudy of visual and auditory word recognition

Contact address:

Ludovic FERRAND

CNRS and Université Clermont Auvergne

Laboratoire de Psychologie Sociale et Cognitive (LAPSCO – UMR CNRS 6024)

34, avenue Carnot

63037 Clermont-Ferrand

France

E-mail : ludovic.ferrand@uca.fr

Phone : + 33 473 406 257

Fax : + 33 473 406 114

**Abstract**

Using the megastudy approach, we report a new database (MEGALEX) of visual and auditory lexical decision times and accuracy rates for tens of thousands of words. We collected visual lexical decision data for 28,466 French words and the same number of pseudowords, and auditory lexical decision data for 17,876 French words and the same number of pseudowords (synthesized tokens were used for the auditory modality). This constitutes the first large-scale database for auditory lexical decision, and the first database to enable a direct comparison of word recognition in different modalities. Different regression analyses were conducted to illustrate potential ways to exploit this megastudy database. First, we compared the proportion of variance accounted for by five word frequency measures. Second, we conducted item-level regression analyses to examine the relative importance of lexical variables influencing performance in the different modalities (visual and auditory). Finally, we compared the similarities and differences between the two modalities. All data are freely available on our website (<https://sedufau.shinyapps.io/megalex/>) and are searchable on <http://www.lexique.org> inside the Open Lexique search engine.

## 1. Introduction

Understanding the cognitive processes involved in visual and auditory word recognition remains a major challenge in cognitive psychology and psycholinguistics. In recent years, researchers have focused on lexical variables affecting the speed and accuracy with which words are processed (e.g., Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Baayen, Feldman, & Schreuder, 2006; Brysbaert, Warriner, & Kuperman, 2014; Ernestus & Cutler, 2015; New, Ferrand, Brysbaert, & Pallier, 2006; Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012; Yap & Balota, 2009). In the current study, we focus on ways in which word processing times converge or diverge in the lexical decision task across two modalities: visual and auditory. Visual and auditory word recognition have rarely been studied together. Indeed, to our knowledge, research on spoken-word recognition and reading has developed independently, with little attempt to draw parallels between the two<sup>1</sup>. The main goal of the present study was to apply the megastudy approach to compare processing times for a very large number of words presented in visual and auditory lexical decision. In this study, we focus (1) on lexical decision (rather than naming), and (2) on very large databases (with more than 10,000 words tested) (see Table 1).

### 1.1. The Megastudy Approach

Megastudies are studies in which word processing times are gathered for a large number of words (typically between 10,000 and 40,000 words; see Balota, Yap, Hutchison, & Cortese, 2013; Keuleers & Balota, 2015, for reviews). The findings are typically explored with multiple regression analyses. In the visual modality using the lexical decision task (see Table 1 for a review), the first megastudy was published by Balota et al. (2007)<sup>2</sup>. In the so-called

---

<sup>1</sup> Note that we are talking about reading and not the use of visual cues in speech perception (e.g., lip reading). There is indeed a large literature on the integration of audio-visual information during speech perception.

<sup>2</sup> Before that, other large-scale studies had been conducted, testing a smaller number of (monosyllabic) words, typically between 1,000 and 3,000. For instance, Balota, Cortese, Sergent-Marshall, Spieler, and Yap (2004) collected naming times and lexical decision times for 2,902 monosyllabic English words. Before them,

English Lexicon Project (ELP), lexical decision and naming data were collected for 40,481 words from several hundreds of participants (each responding to 3,400 stimuli in lexical decision). Ferrand et al. (2010) published a similarly collected set of lexical decision times for 38,840 French words (the French Lexicon Project, FLP) from 975 participants (each responding to 2,000 stimuli). Yap, Rickard Liow, Jalil, and Faizal (2010) collected lexical decision and naming data for 9,592 Malay words (the Malay Lexicon Project, MLP) from 44 participants (each responding to 1,020 stimuli in lexical decision). Keuleers, Diependaele and Brysbaert (2010) published a set of lexical decision times for 14,000 Dutch words (the Dutch Lexicon Project, DLP) with a new approach since 39 participants responded to all stimuli (see also Brysbaert et al., 2016, for the Dutch Lexicon Project 2, DLP2). A similar approach was adopted for the British Lexicon Project (Keuleers, Lacey, Rastle, & Brysbaert, 2012), in which data on 28,000 English words were collected by testing 78 participants, each of them responding to one of the two lists of 14,000 words in lexical decision. Dufau, Duñabeita, Moret-Tatay, McGonigal, Peeters, Alario, Balota et al. (2011) collected lexical decision times for thousands of words in seven languages (English, Basque, Catalan, Dutch, French, Malay and Spanish) by using smartphone technology (each participant responding to 50, 100 or 140 stimuli). Finally, Tse, Yap, Chan, Sze, Shaoul, and Lin (2016) collected lexical decision data for 25,286 Chinese words (the Chinese Lexicon Project, CLP) from 594 participants (each responding to 2,810 stimuli).

---

Seidenberg and Waters (1989) collected naming times for 2,897 monosyllabic English words. Treiman, Mullenix, Bijeljac-Babic, and Richmond-Welty (1995) collected naming times for 1,327 monosyllabic English words. Spieler and Balota (1997) collected naming data for 2,820 monosyllabic English words. Later on, Ferrand, Brysbaert, Keuleers, New, Bonin, Méot, Augustinova, and Pallier (2011) collected naming, lexical decision and progressive demasking data for 1,482 monosyllabic French words. More recently, Adelman, Marquis, Sabatos-DeVito, and Estes (2013) collected naming times from four people who read each 2,820 English words 50 times each, and even more recently, Dufau, Grainger, Midgley, and Holcomb (2015) measured event-related potentials for 1,000 English words in a go/no lexical decision task. Very recently, Schröter and Schroeder (2017) collected naming and lexical decision data for 1,152 German words in seven different age groups (Grades 1 to 6, as well as two groups of younger and older adults).

**Table 1.** List of word recognition megastudies using the lexical decision task (adapted from Keuleers et al., 2012)

Source	Name	Modality	Material	Participants
Balota et al. (2007)	The English Lexicon Project (ELP)	Visual	40,481 English words	816 students
Ferrand et al. (2010)	The French Lexicon Project (FLP)	Visual	38,840 French words	975 students
Keuleers et al. (2010)	The Dutch Lexicon Project (DLP)	Visual	14,037 Dutch words	39 students and university staff
Yap et al. (2010)	The Malay Lexicon Project (MLP)	Visual	9,592 Malay words	44 students
Dufau et al. (2011)	Science XL	Visual	thousands of words in seven languages (English, Basque, Catalan, Dutch, French, Malay, and Spanish)	> 4,000 people
Keuleers et al. (2012)	The British Lexicon Project (BLP)	Visual	28,730 British words	78 students and university staff
Ernestus and Cutler (2015)	BALDEY	Auditory	2,780 Dutch words	20 students
Brysbaert et al. (2016)	The Dutch Lexicon Project 2 (DLP2)	Visual	30,016 Dutch words	81 students and university staff
Tze et al. (2016)	The Chinese Lexicon Project (CLP)	Visual	25,286 Chinese words	594 students
Ferrand et al. (present study)	MEGALEX	Visual and Auditory	28,466 French words in the visual modality and 17,876 French words in the auditory modality	197 students and university staff

In the auditory modality, megastudies are scarce, possibly because generating auditory stimuli entails a great deal more effort than generating visual stimuli. To our knowledge, only

one published megastudy used the auditory lexical decision task, BALDEY ("biggest auditory lexical decision experiment yet"). In this megastudy, Ernestus and Cutler (2015) collected lexical decision times on 2,780 Dutch words from 20 participants (see Table 1).

As pointed out by Balota et al. (2007, 2013) and Brysbaert et al. (2016, see their Table 1 page 442), megastudies have a number of important advantages over the more commonly used factorial designs. First, megastudies use multiple regression techniques applied to the data obtained with very large numbers of words, and one is no longer limited to analysing a small number of words since all words (polysyllabic, polymorphemic, inflected forms such as feminine, plural, and verbal forms) are included in the analysis. This avoids the potential biases that may occur when one selects specific items (see e.g., Forster, 2000). Second, variables (such as word frequency for instance) can be tested in a continuous way along their entire range. Third, megastudies identify the unique predictive power of a large set of targeted variables. Fourth, megastudies have proven valuable for comparing competing metrics of word frequency (e.g., Brysbaert & New, 2009; Ferrand et al., 2010), evaluating the impact of novel psycholinguistic variables (e.g., Bonin, Méot, Ferrand, & Bugaïska, 2015; Brysbaert et al., 2016; Juhasz & Yap, 2013; Yarkoni, Balota, & Yap, 2008), and exploring potential nonlinear functional relationships between lexical variables and word recognition performance (e.g., Baayen et al., 2006; New et al., 2006).

Because of these advantages, an increasing number of large-scale psycholinguistic studies have recently endorsed the megastudy approach using regression designs (see Table 1; see also Keuleers & Balota, 2015, and Brysbaert et al., 2016, for reviews).

## **1.2. Factors Influencing Visual and Auditory Lexical Decision Performance**

As Brysbaert et al. (2016, p. 442) put it, "One of the main goals of megastudies is to examine which variables affect word recognition and what their relative importance is". A survey of the megastudy literature reveals that word frequency is the most important variable in visual

lexical decision (accounting for at least 40% unique variance in ELP; Balota et al., 2007; Brysbaert et al., 2016; 34% in DLP1; Keuleers et al., 2010; 33% in DLP2; Brysbaert et al., 2016; and 38% in FLP; Ferrand et al., 2010)<sup>3</sup>. A second important variable accounting for a decent part of variance in lexical decision times is word length in letters (accounting for 16% of variance in ELP; and 7% in the DLP1 and the FLP)<sup>4</sup>. Lexical decision latencies are generally longer for long words; but New, Ferrand, Pallier and Brysbaert (2006) re-examined the length effect within the ELP and they found a quadratic effect: the effect of number of letters was facilitatory for words of 3–5 letters, null for words of 5–8 letters, and inhibitory for words of 8–13 letters. This quadratic effect was replicated in French (Ferrand et al., 2010). A third important variable that matters is the orthographic similarity to other words (OLD 20; Yarkoni, Balota, & Yap, 2008). Lexical decisions are faster to words that are more orthographically similar to other words (accounting for 20% of variance in ELP, 5% in DLP and 6.5% in FLP; again see footnote 4).

Other lexical variables have also been identified even if they only account for a small (but often significant) percentage of variance (around 1-2% of additional variance). This is the case of (1) semantic variables (such as imageability, sensory experience ratings, concreteness, number of semantic features, number of associates, to name a few; e.g., Balota et al., 2004; Bonin, Méot, Ferrand, & Bugaïska, 2015; Ferrand et al., 2011; Juhasz & Yap, 2013; Juhasz, Yap, Dicke, Taylor, & Gullik, 2011); (2) morphological variables (such as morphological family size, plural word forms, etc., - e.g., Baayen et al., 2006; Gímenes, Brysbaert, & New, 2016); and (3) affective variables (such as valence and arousal; e.g., Kuperman, Estes, Brysbaert, & Warriner, 2014).

---

<sup>3</sup> Rated age-of-acquisition (AoA), the age at which the word was learned first, explains about 5% of additional variance after word frequency has been partialled out (see Brysbaert et al., 2016; Ferrand et al., 2011).

<sup>4</sup> One likely reason for this difference is that in ELP, nonwords were generated by changing one letter in a corresponding target word. This way of constructing nonwords confounded nonword length and word likeness, whereas in DLP and FLP, nonwords were generated in such a way that their orthographic features mimicked the orthographic features of the words.

In auditory lexical decision, the only megastudy available (Ernestus & Cutler, 2015) reported that word duration was an important variable and that it was a better predictor of response time than different measures of the uniqueness point (i.e. the first moment at which a word differs from all other words)<sup>5</sup>. This result suggests that listeners tended to wait until they had heard the last phoneme in the word before making their decision. Word frequency was also an important variable, with the Dutch subtitle-based word frequency (Keuleers, Brysbaert, & New, 2010) better accounting for the auditory lexical decision data than other frequency measures (such as CELEX for instance; Baayen, Piepenbrock, & Gulikers, 1995).

Standard research using the auditory lexical decision task in factorial experiments has mainly focused on the effects of lexical variables such as word frequency and word-form similarity. In a review of the auditory lexical decision task published twenty years ago, Goldinger (1996) mentioned one variable that surely influenced lexical decision performance (word frequency), one that probably affected it (neighborhood density) and one that should be controlled (total stimulus duration). So far, researchers have reported clear frequency effects (e.g., Cleland, Gaskell, Quinlan, & Tamminen, 2006; Connine, Mullennix, Shernoff, & Yelen, 1990; Luce & Pisoni, 1998; Petrova, Gaskell, & Ferrand, 2011; Taft & Hambly, 1986), as well as effects of phonological neighbors that are inhibitory in nature (e.g., Goh, Suarez, Yap, & Tan, 2009; Goldinger, Luce, & Pisoni, 1989; Luce & Pisoni, 1998; Vitevitch & Luce, 1999; Ziegler, Muneaux, & Grainger, 2003), and effects of cohorts (e.g., Taft & Hambly, 1986; Marslen-Wilson, 1990) or uniqueness point (e.g., Goodman & Huttenlocher, 1988).

More recently, Goh, Yap, Lau, Ng, and Tan (2016) collected auditory lexical decision (and semantic categorization) data for 514 nouns from 80 participants. The goal of their study

---

<sup>5</sup> However, due to the analysis method used by Ernestus and Cutler (2015, pp. 1477-1478), they could not provide the percentage of unique variance explained by this variable, but only reveal which variable was the best predictor.

was to determine the unique contribution of semantic richness variables (such as concreteness, number of semantic features, and emotional valence), above and beyond the contribution of lexical variables (such as word duration, frequency, structural properties, and number of morphemes) to auditory word recognition. Their results showed that the lexical variables collectively accounted for 44.5% of the variance in lexical decision RT. There were significant positive relationships between RT on the one hand, and word duration and number of morphemes on the other hand, such that words that had longer duration and more morphemes were associated with slower RTs. Semantic richness variables (concreteness, valence, and number of semantic features) collectively accounted for an additional 3% of unique variance in lexical decision RT, with faster responses for spoken words that were concrete, emotionally valenced, and with a high number of semantic features, suggesting that words with richer semantic representations are recognized faster.

### **1.3. The present study**

The main goal of the present study was to investigate and compare visual and auditory lexical decision times using the megastudy approach. The data collected in the MEGALEX project can be used to perform different types of analyses. As an example, here we outline the different central questions that we had specifically in mind when we designed MEGALEX: Which word frequency measure best predicts response times in the visual and auditory lexical decision tasks? What is the relative importance of the different lexical variables in both modalities? A specific question is when, during stimulus presentation, participants in the auditory lexical decision task decide whether a stimulus is a word? A more general question concerns the similarities and differences of the different variables involved in the visual and auditory lexical decision task. Finally, one methodological goal was to compare the power of the design used in MEGALEX (in which each participants responded to half of the full set of

stimuli) compared to the design used in the French Lexicon Project (in which participants only responded to 2,000 of about 40,000 stimuli).

## **2. Experiment**

### **2.1. Method**

#### **2.1.1. Participants**

A total of 197 right-handed participants finished the experiment (96 in the visual LDT and 101 in the auditory LDT). Five other participants started the visual or auditory experiment but did not finish it. Participants were students or employees coming from three different universities: University Blaise Pascal (Clermont-Ferrand), University Pierre-Mendès France (Grenoble) and Aix-Marseille University (Marseille). Their age ranged between 17 and 52 years (mean = 25.76, SD = 5.71). All were native speakers of French, and had normal or corrected-to-normal vision. None of them had hearing problems. Each participant participated in either the visual LDT or the auditory LDT. Participants were informed that successful completion of the (visual or auditory) experiment would take about 20 h, for which they would receive a payment of 300 euros. They were also informed that they would be excluded if their accuracy dropped below 80% in three successive blocks (as in Keuleers et al., 2012). They were also informed that if they dropped out the experiment or if their accuracy fell consistently below the 80% benchmark, they would be paid 5 euros per hour completed.

#### **2.1.2. Ethical statement**

This study was approved by both the Clermont-Ferrand Sud-Est VI Statutory Ethics Committee (Comité de Protection des Personnes (CPP) Sud-Est 6, France; Authorization #AU 1058) and the French National Agency for Medicines and Health Products Safety (Agence Nationale de Sécurité du Médicament et des Produits de Santé (ANSM), France;

Authorization #2013-A00591-44 and #B-130783-81) according to the articles of law L. 1121-1-2 and R 1121-3.

### **2.1.3. Apparatus**

The experimental software (OpenSesame; Mathôt, Schreij, & Theeuwes, 2012) and testing apparatus were identical at each of the three testing sites (Clermont-Ferrand, Grenoble, and Marseille). All the participants were tested on the same platform.

For the visual LDT, the stimuli were presented on a 22-in Dell LCD monitor (P2212H) with a refresh rate of 60 MHz and a resolution of 1,920 x 1,080 pixels, placed at a distance of about 60 cm from the participants. The monitor was controlled by a PC (Dell Precision T3600). The stimuli were presented in lowercase in Inconsolata (12-point font size), and they appeared on the screen as dark characters on a grey background. The participants responded on a Logitech Gamepad F310, which is used for superfast computer games and does not have time delays with keyboards (see, e.g., Shimizu, 2002).

For the auditory LDT, the apparatus was identical to the one used in the visual LDT, except that the stimuli were played binaurally over headphones (Sennheiser HD 25-1 II) through a high-quality audio soundcard (Asus Xonar DX PCIe). Timing calibration was performed using the Blackbox Toolkit version 2 (<http://www.blackboxtoolkit.com/>).

### **2.1.4. Word and pseudoword stimuli**

The stimuli consisted of a list of 28,466 words. The words were taken from the French Lexicon Project (Ferrand et al., 2010; see Table 2 for the descriptive statistics of the words and their behavioral measures). For each word, a matched pseudoword was generated using the *Lexique* toolbox (available from

<http://www.lexique.org/toolbox/toolbox.pub/index.php?page=trigramme>). In particular, we used the tool that creates pseudowords based on randomization of bigrams or trigrams coming from real words. The advantage of this generator is that it creates pseudowords having characteristics very close to the category of words chosen. For instance when one enters all the words having 6 letters, 2 syllables, being a noun, the toolbox will generate pseudowords having 6 letters, 2 syllables and looking like possible nouns. One can therefore generate words resembling plausible French nouns or plausible French inflected verbs, etc. For instance, for 3 letter words, we entered all our words having 3 letters and we generated many 3 letter pseudowords. Then we screened the pseudowords in order to exclude real words and in order to end up with as many 3 letter pseudowords as we had 3 letter words.

<Insert Table 2 about here: See table at the end of the MS>

**Visual stimuli.** In each session, the participants received 260 words and 260 pseudowords. Each session contained two blocks of 130 words and 130 pseudowords, with a short break between the two blocks. Each session lasted about 20 minutes. Overall, each participant received 55 sessions (or 109 blocks), corresponding to about 20 hours of testing.

**Auditory stimuli.** They were identical to those used in the visual lexical decision task, except that homophonous forms were suppressed, giving a total amount of 17,876 words and 17,876 pseudowords. The auditory stimuli were synthesized using Apple's Text-To-Speech accessibility feature (TTS's "say" command; Apple, Cupertino, Ca, USA)<sup>6</sup>. For instance, shell scripts containing "say" commands synthesized all the words and pseudowords of the visual experiment at regular speed in a female French voice (called "Audrey") and produced 22100 Hz stereo wav files (each text stimulus had its own associated wav file). Pair-wise auditory signal comparison was then carried-on to detect identical wav files (hence

---

<sup>6</sup> Initially, we compared the auditory output of two different text-to-speech software programs, Acapela (<http://www.acapela-group.com/>) and TTS (from Apple), on a few hundred words and nonwords selected randomly. We chose the software that according to five judges (LF, BN, SD, ES and CP) produced the best quality.

homophonous forms). Such forms were then removed from stimulus lists. Finally, leading and trailing non-informative zero values in wav files were removed. In each session, the participants received 178 words and 178 pseudowords. Each session contained two blocks of 89 words and 89 pseudowords, with a short break between the two blocks. Each session lasted about 20 minutes. Overall, each participant received 50 sessions (or 100 blocks), corresponding to about 20 hours of testing.

#### **2.1.5. Creation of lists for participants**

Both in the visual and in the auditory experiments, each participant was presented with exactly half of the experimental items. This was accomplished in the following way. For each successive pair of participants, we shuffled the words and pseudowords lists separately, then split each of them in half. The items in the first halves were attributed to the first participant and the items in the second half were attributed to the second participant. The words and pseudowords were then mixed and their order shuffled again for each participant. Finally, the obtained lists were split in blocs of 260 stimuli in the visual modality or 178 stimuli in the auditory modality to create individual experimental runs.

#### **2.1.6. Procedure**

Each participant started with a one-hour startup session at one of the three universities and was assigned to the visual or the auditory LDT. During that session, participants received information about the experiment and signed the approval forms. Participants then completed practice session of 40 trials (20 words and 20 pseudowords), allowing us to demonstrate the main features of the experiment. This was immediately followed by a real session of 20 min (Session 1, consisting of two blocks). After this first session, participants were free to enter the different labs during office hours and to go through the experiment at their own pace,

using a booking system to reserve time slots. After reservation, participants could sit at any of the seven computers specifically devoted to the study. After entering their registration code in the experiment system, they were presented with their next block of trials. After each completed block, participants could choose whether to continue or to stop the session. The only advice given to the participants was to limit their participation to 2 h per half day. The fastest participant finished the experiment in 15 days' time, the slowest took 58 days (on average, participants took 32 days,  $SD=15.40$ ).

**Visual LDT.** There were 10 practice trials before each experimental block. The participants saw words and pseudowords presented on the center of a screen and they had to indicate as rapidly and as accurately as possible whether the presented letter string was a French word or a pseudoword. The participants responded using response buttons on a Logitech Gamepad F310. They answered "yes" by pressing the button corresponding to the index finger of their right hand and "no" by pressing the button corresponding to the index finger of their left hand.

The sequence of events was as follows (for one trial): (1) a fixation point (".") appeared in the center of the screen for 250 msec; (2) a blank screen appeared for 100 msec; (3) a stimulus (word or pseudoword) was presented in the center of the screen; (4) the participant made the response as fast and as accurately as possible; (5) the stimulus was erased from the screen. The stimulus remained on the screen until a manual response was detected or for 2 sec maximum if no response was made. At the end of each trial, there was a 1,000 msec inter-trial interval with a blank, grey screen. A feedback was provided during each session, one at the end of each block (with mean correct reaction times and percentage of errors for words).

**Auditory LDT.** The procedure was similar to the one used in the visual modality. There was 10 practice trials before each experimental block. The sequence of events was as follows (for one trial): (1) a fixation point (".") appeared in the center of the screen for 250 msec,

announcing the auditory stimulus; (2) a stimulus (word or pseudoword) was then played binaurally over headphones; (2) the participant made the response as fast and as accurately as possible. The response was triggered as soon as possible, even if the auditory stimulus was still playing. There was a 2 sec time-out if the participant had not given any response. At the end of each trial, there was a 1,000 msec intertrial interval with a blank, grey screen. Feedback – mean correct reaction times and percentage of errors – was provided at the end of each block.

## 2.2. Results

**Trimming procedure for the visual lexical decision.** First, 3,939 trials with technical problems were eliminated, which left us with 2,596,095 trials (1,311,832 words).

As in Ferrand et al. (2010), all response latencies shorter than 300 msec or longer than 2,000 msec were then identified as outliers (resulting in the elimination of 8,658 trials (4,383 words), i.e., 0.33% of the data). Nearly all participants had a mean accuracy higher than 80%. The data for 3 participants who did not fulfill these criteria were dropped (73,272 trials; 37,071 for words). Eight blocks with accuracies lower than 75% were also eliminated (1,981 trials; 1,009 words). The resulting 2,512,184 trials (1,269,369 words) are freely available on our website (<https://sedufau.shinyapps.io/megalex/>) and are searchable on <http://www.lexique.org> inside the Open Lexique search engine.

For the regression analyses on reaction times (RTs) and accuracy presented below, 2,690 words (119,684 trials) with accuracies below 67% were removed (as in Balota et al., 2004; Ferrand et al., 2010, 2011). Moreover, for the remaining correct RTs, the means and SDs were computed within blocks\*participants, and all RTs less than three SDs below the mean or greater than three SDs above it were considered as outliers as well. This resulted in the

rejection of 1.7% (19,229) of the RTs on correct trials. The mean RTs for the trimmed correct word trials, as well as the mean accuracy rates are presented in Table 2.

**Trimming procedure for the auditory lexical decision.** We used the same procedure for the RTs for correct responses as applied in the visual modality. The elimination of 16,878 trials with technical problems led first to a set of 1,669,585 trials (845,078 words). Response latencies shorter than 600 msec or longer than 2,500 msec were then identified as outliers (resulting in the elimination of 1,084 trials (557 words), i.e., 0.06% of the data). These threshold values are slightly different from the ones used for the visual modality. These were chosen after inspection of the cumulative frequency of the distribution of RTs for correct responses: 600 msec corresponds to the value for which the slope began to change significantly and for which eliminations seemed not to concern particular items. As the decrease of the slope was very monotonic at the right of the distribution, the upper bound was fixed at 2,500 msec. One participant (161,16 trials; 8,168 words) with a global accuracy below 80% and 17 blocks (2,738 trials; 1,396 words) with accuracies lower than 75% were then dropped. The resulting 1,649,647 trials (834,957 words) are freely available on our website.

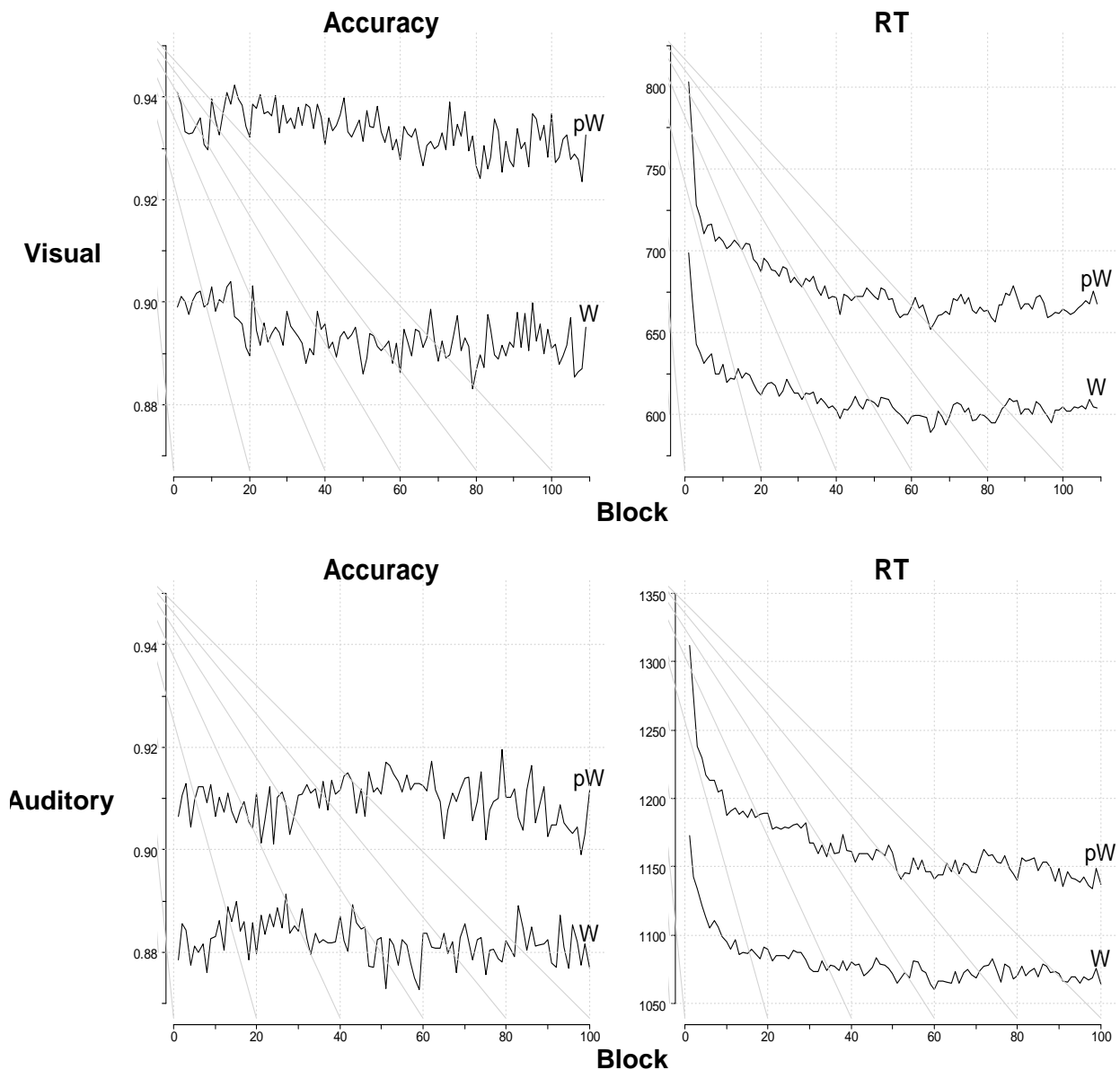
For the regression analyses, 2,034 words (94,197 trials) with accuracies below 67% were then removed (as in Balota et al., 2004; Ferrand et al., 2010, 2011). Moreover, for the remaining correct RTs, all RTs less than three SDs below the mean of the within blocks\*participants or greater than three SDs above this mean were considered outliers as well. This resulted in the rejection of 1.3% (9,580) of the RTs on the correct trials. The mean RTs for the trimmed correct word trials, as well as the mean accuracy rates are presented in Table 2.

The trimming procedure left us with 25,776 words (out of 28,466) in the visual modality and 15,842 words (out of 17,876) in the auditory modality. In both modalities, raw RTs were transformed into standardized  $z$  scores per participant and per block to remove effects due to block differences and variability/speed between participants (as recommended by Faust, Balota, Spieler, & Ferraro, 1999; see also Ferrand et al., 2010 for such a procedure). By items RTs and  $z$ RTs means and accuracies for the independent variables used in the regression analyses are freely available on our website (<https://sedufau.shinyapps.io/megalex/>) and are searchable on <http://www.lexique.org> inside the Open Lexique search engine.

### **2.2.1. Reaction Times and Accuracy in MEGALEX**

Figure 1 displays the effects of practice on RT and accuracy over the total duration of the experiment in both modalities (i.e., 109 blocks in the visual modality and 100 blocks in the auditory modality).

In the visual modality, the practice effect is about 100 ms difference in RT for words and 120 ms for pseudowords. This is similar to the effect observed in the BLP (Keuleers et al., 2012) where it was also 100 ms. In the auditory modality, the practice effect is also around 100 ms difference in RT for words and 140 ms for pseudowords (see Figure 1).



**Figure 1.** Practice effects in MEGALEX for accuracy (left panel) and reaction times (right panel) for the visual modality (upper panel) and the auditory modality (lower panel). pW: pseudowords; W: words. RTs are given only for correct responses.

### 2.2.2. Reliability of the MEGALEX Measures

As in previous megastudies (Ferrand et al., 2010; Keuleers et al., 2010; Tse et al., 2016), we determined the reliability of accuracy and RTs in order to evaluate how useful they might be. For each word, participants were split in two groups each containing half of the participants having seen that word. RT means, zRT means, and accuracies were calculated for each half.

Correlations were then calculated between item scores obtained in the first and second halves. Spearman-Brown corrections are also reported as  $r_{\text{corr}} (2r/(1+r))$ . In this way, the correlation between (1) accuracy calculated on the first half of the participants who saw a given word and (2) accuracy calculated on the remaining participants who saw the same word was .66 ( $r_{\text{corr}} = .8$ ) in the visual modality and .68 ( $r_{\text{corr}} = .81$ ) in the auditory modality. For RTs in the visual modality,  $r = .70$  ( $r_{\text{corr}} = .82$ ) and for  $z\text{RT}$ ,  $r = .82$  ( $r_{\text{corr}} = .9$ ). For the RTs in the auditory modality,  $r = .82$  ( $r_{\text{corr}} = .9$ ) and for  $z\text{RT}$ ,  $r = .89$  ( $r_{\text{corr}} = .94$ ). The fact that the reliability of  $z\text{RT}$  is higher than the reliability of the raw RTs confirms that taking away differences in overall RT and variability between participants and blocks removes noise from the data and does not artificially reduce the variability of the items.

### 2.2.3. Predictors of the MEGALEX Lexical Decision Times

We considered the following predictors for the by-item analyses:

**Word frequency.** We used the following corpora to measure word frequency: books (New et al., 2001, 2004); subtitles (New et al., 2007); and blogs, Twitter and newspapers (Gimenes & New, 2016).

**Word length (in letters).** The number of letters in the word.

**Word length (in phonemes).** The number of phonemes in the word.

**OLD20.** The orthographic similarity to other words (Yarkoni et al., 2008), calculated on the basis of the *Lexique* vocabulary (Ferrand et al., 2010).

**PLD20.** The phonological similarity to other words (Yarkoni et al., 2008), calculated on the basis of the *Lexique* vocabulary (Ferrand et al., 2010).

**Orthographic uniqueness Point.** The point in the letter string, proceeding from the first letter to the last, at which only one orthographic candidate remains in the cohort of possible words (calculated on the basis of *Lexique* vocabulary: New et al., 2004).

**Phonological uniqueness point.** The point in the phoneme string at which only one phonological candidate remains in the cohort of possible words (calculated on the basis of *Lexique* vocabulary: New et al., 2004).

**Stimulus duration.** The total duration of the spoken item, from the onset of the word to the end of the word, which corresponded to the duration of the edited soundfiles.

#### **2.2.4. Which word frequency measure best predicts response times in the visual and auditory lexical decision task?**

As Tse et al. (2016, p. 7) put it, “Word frequency has consistently been shown to be one of the strongest predictors of visual word recognition in various languages. Given that this is a central lexical variable to manipulate and control, it is important to determine which word frequency count accounts for the largest proportion of variance in lexical decision performance”.

To find out which word frequency measure best predicts response times in the visual and auditory lexical decision task, we have considered five frequency measures that are publicly accessible for the current megastudy.

First, we used two measures of word frequencies from *Lexique* (New, Pallier, Ferrand, & Matos, 2001 ; New, Pallier, Brysbaert, & Ferrand, 2004): (1) the book-based frequency (Frantext), and (2) the subtitle-based frequency (Subtlex-FR). The subtitle-based frequencies were first computed in French by New, Brysbaert, Véronis, and Pallier (2007). They found that the subtitle-based frequencies were a better predictor of lexical decision times than the book-based frequencies (see also Brysbaert & New, 2009, for a similar result in English). Ferrand et al. (2010) replicated this finding in the French Lexicon Project, the subtitle-based frequencies explaining 5%-6% more of the variance in the percent errors and reaction times than did the book-based frequencies. Furthermore, Ferrand et al. (2010; see also Brysbaert &

New, 2009) showed that the predictive power of the frequencies further increased when the averages of the subtitle and book frequencies were used. We investigated whether this is also true for auditory lexical decision.

Second, we used three new measures of word frequencies based on (1) Twitter, (2) blog posts, and (3) newspapers (taken from *Worldlex* by Gimenes & New, 2016). Based on lexical decision times for 35,658 words taken from the French Lexicon Project (Ferrand et al., 2010), Gimenes and New (2016) showed that the three new frequencies averaged (Twitter + Blogs + Newspapers) explained slightly more variance (48.46%) than the two classic frequencies -- book-based and subtitle-based averaged --(47.56%). The same result was obtained for the English Lexicon Project (Balota et al., 2007) and the Dutch Lexicon Project (Keuleers et al., 2010).

The frequency measures were log transformed (see Baayen et al., 2006). Also, because Balota et al. (2004; see also Ferrand et al., 2011) reported a nonlinear relationship between log frequency and lexical decision times, we report regression analyses both for  $\log(\text{frequency})$  and  $\log(\text{frequency}) + \log^2(\text{frequency})$ .

Table 3 shows the percentages of variance in reaction times and accuracy explained by the different frequency measures for all words in each modality (visual and auditory) and also for common words in both modalities. There are several interesting results. First, in line with previous findings (Ferrand et al., 2010; New et al., 2007), the subtitle frequency measure outperforms the book frequency measure in the visual modality. Note however that this was not the case in the auditory modality, for which both frequencies were roughly similar (with a slight advantage for the book frequency); this goes against Ernestus and Cutler's (2015) finding that the subtitle frequency measure outperformed the book frequency measure as in

the visual modality<sup>7</sup>. Second, the proportion of variance explained by frequency is relatively low in the auditory modality compared to the visual modality (for common words, 11.3% vs. 45.1%); this is much lower than the estimates previously reported for visual lexical decision performance<sup>8</sup>. Third, as in the French Lexicon Project (Ferrand et al., 2010), more variance is explained when the averages of subtitle and book frequencies were used than when these two frequencies are taken separately; this was the case in both modalities, suggesting that a combination of spoken and written frequencies may be the way forward. Fourth, this frequency measure combining subtitle and book frequencies roughly shows similar results to the three new frequency measures combining Twitter, blogs, and newspapers when the square of the log frequency was not added (similarly to what Gimenes & New, 2016, found; see Table 3). However, when the square of the log frequency was added, the frequency measure combining subtitle and book frequencies outperforms the three new frequency measures combining Twitter, blogs, and newspapers; this was the case in both modalities. Fifth, these analyses confirm that less noise is present in the  $z$ RT variable than in the raw RT variable.

Given that the combination of subtitle and book frequencies is unequivocally the best measure, both for visual and auditory word processing (see Table 3), we will use this measure in the remainder of this article.

---

<sup>7</sup> Ernestus and Cutler (2015) also tested a spoken frequency measure (CGN, containing recordings of completely spontaneous casual speech), but the subtitle frequency also outperformed this spoken frequency measure. Note however that when lemma frequencies (instead of form frequencies) were considered, the book frequency measure outperformed the subtitle frequency measure, suggesting that this subtitle frequency measure (based on CELEX) better reflects participants' knowledge of word lemmas. Interestingly, Ernestus and Cutler (2015, p. 1483) reported that “the subtitle and the spoken word form frequencies were better predictors than the corresponding lemma frequencies, suggesting that participants' recognition of morphologically complex words was based on the word form themselves rather than their stems”.

<sup>8</sup> Due to their analysis method used, Ernestus and Cutler (2015) cannot provide the percentage of unique variance explained by this variable, but they can only reveal which variable is the best predictor. However, in an unpublished study available at the following web address (<http://crr.ugent.be/members/marc-brybaert#pu6>), Yap and Brybaert (2009) also reported low proportions of variance explained by frequency around 10%. These estimates were obtained by regression analyses based on the large-scale study of Luce and Pisoni (1998) involving auditory lexical decision times for 918 monosyllabic English words.

**Table 3**

*Percentages of Variance Explained in the MEGALEX data by Five Frequency Measures Available in Lexique (Based on Subtitles and on Books) and Worldlex (based on Twitter, Blogs, and Newspapers) and Their Combination*

	Visual $R^2$ (all words)			Auditory $R^2$ (all words)		
	Acc	RT	$z$ RT	Acc	RT	$z$ RT
Log(FreqBlogs)	12.4	31.5	33.1	4.2	7.8	7.8
Log(FreqBlogs) + Log <sup>2</sup>	16.1	34.8	36.6	7.0	7.9	8.0
Log(FreqTwitter)	9.7	28.5	30.0	2.9	7.5	7.6
Log(FreqTwitter) + Log <sup>2</sup>	13.3	32.6	34.4	5.2	7.7	7.8
Log(FreqNews)	10.4	24.8	26.2	3.6	4.9	4.9
Log(FreqNews) + Log <sup>2</sup>	13.1	26.8	28.4	5.8	5.0	5.0
Log(Blogs + Twitter + News)	13.5	32.7	34.4	4.7	7.3	7.3
Log(Blogs + Twitter + News) + Log <sup>2</sup>	16.3	34.4	36.3	6.9	7.3	7.4
Log(FreqSubtitle)	11.3	30.4	32.2	3.2	9.6	9.7
Log(FreqSubtitle) + Log <sup>2</sup>	16.0	35.3	37.6	5.7	9.9	10.0
Log(FreqBook)	10.4	27.5	29.0	3.4	11.0	11.2
Log(FreqBook) + Log <sup>2</sup>	13.1	29.8	31.5	5.8	11.5	11.7
Log(FreqSubtitle + FreqBook)	13.5	33.9	35.8	4.2	11.9	12.1
Log(FreqSubtitle + FreqBook) + Log <sup>2</sup>	18.4	38.1	40.4	7.2	12.5	12.6

	Visual $R^2$ (common words)			Auditory $R^2$ (common words)		
	Acc	RT	$z$ RT	Acc	RT	$z$ RT
Log(FreqBlogs)	14.5	35.7	37.5	2.5	6.8	6.9
Log(FreqBlogs) + Log <sup>2</sup>	17.7	38.2	40.2	4.5	6.9	6.9
Log(FreqTwitter)	12.1	33.4	35.1	1.6	6.7	6.8
Log(FreqTwitter) + Log <sup>2</sup>	15.3	36.7	38.6	3.2	6.8	6.9
Log(FreqNews)	11.9	28.1	29.6	2.1	4.1	4.1
Log(FreqNews) + Log <sup>2</sup>	14.2	29.8	31.4	3.8	4.2	4.2
Log(Blogs + Twitter + News)	15.3	36.4	38.2	2.7	6.3	6.3
Log(Blogs + Twitter + News) + Log <sup>2</sup>	17.1	37.2	39.0	4.3	6.3	6.3
Log(FreqSubtitle)	12.9	34.3	36.3	1.6	8.7	8.8
Log(FreqSubtitle) + Log <sup>2</sup>	16.7	38.2	40.4	3.3	8.9	9.0
Log(FreqBook)	12.7	32.9	34.8	1.9	10.0	10.2
Log(FreqBook) + Log <sup>2</sup>	15.4	35.1	37.1	3.7	10.4	10.6
Log(FreqSubtitle + FreqBook)	15.7	38.9	41.2	2.2	10.8	11.0
Log(FreqSubtitle + FreqBook) + Log <sup>2</sup>	20.1	42.6	45.1	4.3	11.2	11.3

Note. All words: 25,776 in the visual modality; 15,842 in the auditory modality. Common words: 14,868 in both modalities. Acc, accuracy; RT, reaction times;  $z$ RT, standardized reaction times. Note that 14,868 words were identical in the visual and auditory modalities (see lower half of Table 3 for the results concerning these words).

### 2.2.5. Comparison of MEGALEX with the French Lexicon Project

A comparison of the visual lexical decision times of MEGALEX and the FLP (25,122 words in common) reveals a clear superiority of MEGALEX over the FLP (in terms of percentage of variance accounted for). For word frequency (based on a combination of the book and the subtitle frequencies), 8.6% more variance is explained by MEGALEX (40% in MEGALEX vs. 31.4% in FLP; see Table 4). Note however that the shape of the frequency effect is very similar in both megastudies (see Figure 2)<sup>9</sup>. There is also a slight advantage of MEGALEX over FLP for word length (in number of letters) and orthographic similarity (see Table 5).

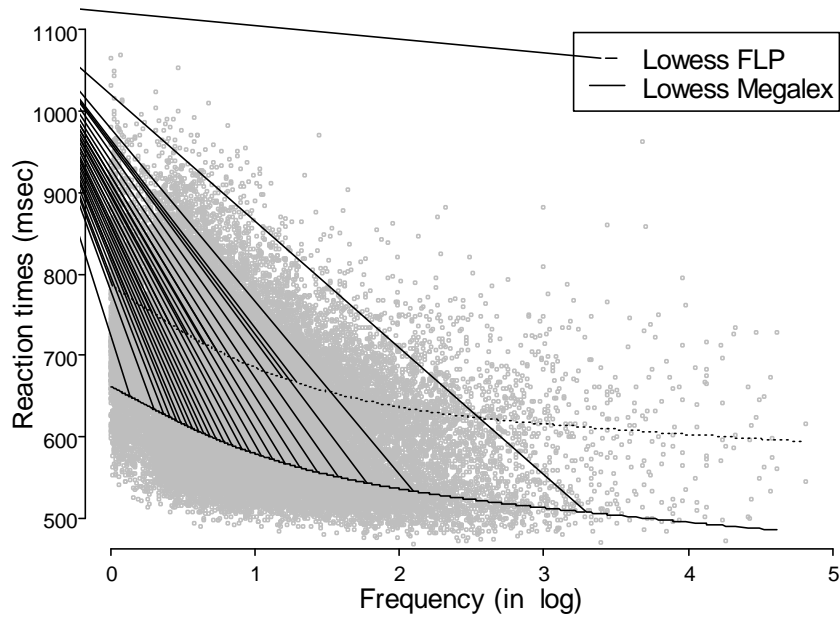
**Table 4**

*Percentages of Variance Explained in MEGALEX and The French Lexicon Project (FLP) for 25,122 Common Words and for Two Frequency Measures Available in Lexique (Based on Subtitles and on Books) and Their Combination*

	MEGALEX $R^2$			FLP $R^2$		
	Acc	RT	zRT	Acc	RT	zRT
Log(FreqSubtitle)	10.9	30.5	32.3	6.4	21.0	22.9
Log(FreqSubtitle) + Log <sup>2</sup>	15.4	35.3	37.5	9.3	25.3	27.8
Log(FreqBook)	9.7	27.3	28.8	6.7	20.7	22.6
Log(FreqBook) + Log <sup>2</sup>	12.1	29.3	31.0	8.6	23.3	25.5
Log(FreqSubtitle + FreqBook)	12.8	33.7	35.6	8.0	24.5	26.8
Log(FreqSubtitle + FreqBook) + Log <sup>2</sup>	17.4	37.8	40.0	11.1	28.7	31.4

Note. Acc, accuracy; RT, reaction times; zRT, standardized reaction times.

<sup>9</sup> As suggested by one reviewer, MEGALEX RTs are consistently faster than FLP RTs (see Figures 2 and 4). This might be due to practice effects, as individual participants received many more trials in MEGALEX than FLP.



**Figure 2.** The shape of the word frequency effect in MEGALEX (solid line) and The French Lexicon Project (dotted line) for 25,122 words in common. Lowess smoothing was obtained using the lowess function of R which computes locally-weighted polynomial regression between dependent variables and independent variables (see Cleveland, 1981).

**Table 5**

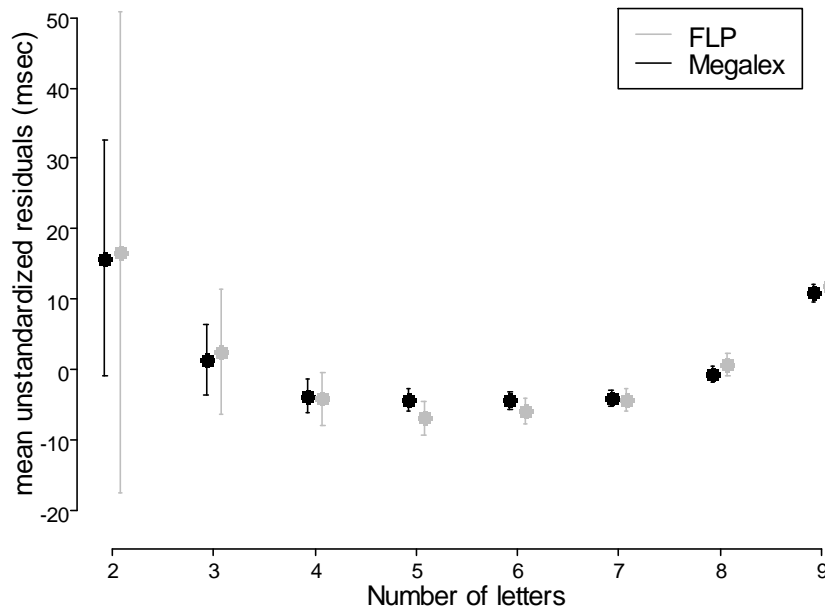
*Percentages of Variance Explained in MEGALEX and The French Lexicon Project (FLP) for 25,122 Common Words and for Frequency, Word Length (in Letters) and Orthographic Similarity (OLD20)*

	MEGALEX $R^2$	FLP $R^2$
	$zRT$	$zRT$
Frequency	35.65	26.77
Frequency + Frequency <sup>2</sup>	40.05	31.42
Frequency + Frequency <sup>2</sup> + $N_{Letters}$	40.59	31.81
Frequency + Frequency <sup>2</sup> + $N_{Letters}$ + OLD20	42.40	32.70

Note.  $zRT$ , standardized reaction times.

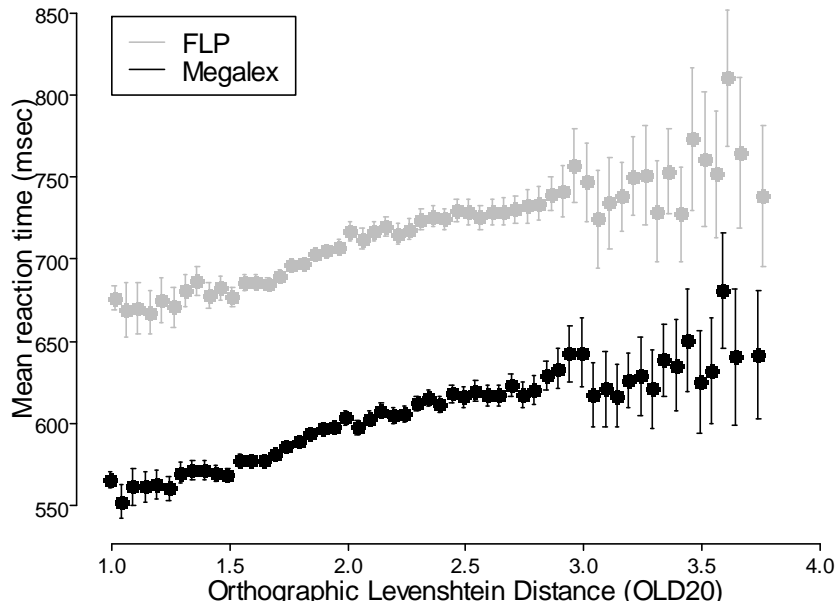
For word length (in letters), Ferrand et al. (2010; see also New et al., 2006, for English) found that this variable had a quadratic effect on visual lexical decision times: the effect of number of letters was facilitatory for words of 3–5 letters, null for words of 5–8 letters, and inhibitory for words of 8–13 letters. We replicated this quadratic length effect in MEGALEX (see

Figure 3), even if the range was more limited in the present study (3 to 9 letters vs. 3 to 13 letters in FLP).



**Figure 3.** The shape of the word length effect in MEGALEX (black dots) and The French Lexicon Project (gray dots) for 25,122 words in common. As in Ferrand et al. (2010), error bars are given for residuals of quadratic regression with RT as the dependent variable and the sum of book and subtitle frequencies (in Log) as an independent variable.

For orthographic similarity, Ferrand et al. (2010) found that similar words (low OLD 20) were responded to more quickly than dissimilar words (high OLD 20). We replicated this orthographic similarity effect in MEGALEX (see Figure 4).



**Figure 4.** The shape of the orthographic similarity effect in MEGALEX (black dots) and The French Lexicon Project (gray dots) for 25,122 words in common.

A comparison of MEGALEX and FLP further confirms the superiority of a design in which the participants process half of the stimuli (MEGALEX) relative to a design in which participants only see a subset of the stimuli (FLP) (see also Ferrand et al., 2015, for the same results based on preliminary data collected in MEGALEX). Indeed, there was more variance accounted for (nearly 10% more) in the lexical decision task of MEGALEX than in FLP (see Table 5). This is in line with findings from the Dutch and the British Lexicon Projects, indicating that there is less noise in megastudies with a complete cross-over of participants and stimuli (Keuleers et al., 2010, 2012).

### 2.2.6. Correlations Between the Variables

Tables 6 and 7 show the intercorrelations between the various predictor variables and the RTs from MEGALEX (for both the visual and auditory modalities). The analysis was limited to the words that were recognized by at least 67% of the participants ( $N = 25,776$  in the visual modality and  $N = 15,842$  in the auditory modality).

In the visual modality (see Table 6), word frequency had the highest correlation with RT.

In the auditory modality (see Table 7), stimulus duration had the highest correlation with RT.

**Table 6**

*Correlations Between Variables in the MEGALEX Data Set for the Visual Modality*

	F <sub>subtitles</sub>	F <sub>books</sub>	F <sub>total</sub>	N <sub>lett</sub>	OLD <sub>20</sub>	OUP	RT	zRT	Acc
Frequency									
Subtitles	1	.81	.93	-.36	-.24	-.33	-.55	-.57	.34
Frequency									
Books		1	.96	-.34	-.24	-.31	-.52	-.54	.32
Frequency									
Total			1	-.37	-.26	-.33	-.58	-.60	.37
Length									
In letters				1	.61	.93	.27	.27	.05
Orthographic									
Levenshtein Distance					1	.57	.29	.30	-.08
Orthographic									
Uniqueness Point						1	.23	.22	.07
Reaction									
Time							1	.97	-.69
Z Reaction									
Time								1	-.71

Note. The contribution of the various word predictors to the visual lexical decision times in the MEGALEX data set. Data are limited to the words that were recognized by more than 67% of the participants (N = 25,776). OLD20, orthographic Levenshtein distance; OUP, orthographic uniqueness point; RT, reaction times; zRT, standardized reaction times; Acc, accuracy.

**Table 7**

*Correlations Between Variables in the MEGALEX Data Set for the Auditory Modality*

	F <sub>subtitles</sub>	F <sub>books</sub>	F <sub>total</sub>	N <sub>phon</sub>	PLD <sub>20</sub>	PUP	Dur	RT	zRT	Acc
Frequency										
Subtitles	1	.82	.93	-.30	-.23	-.27	-.20	-.31	-.31	.18
Frequency										
Books		1	.96	-.32	-.27	-.28	-.22	-.33	-.33	.19
Frequency										
Total			1	-.33	-.26	-.29	-.22	-.35	-.35	.21
Length										
In phonemes				1	.67	.85	.74	.46	.48	.17
Phonological										
Levenshtein Distance					1	.62	.46	.32	.33	.10
Phonological										
Uniqueness Point						1	.59	.37	.39	.17
Stimulus										
Duration							1	.65	.67	.14
Reaction										
Time								1	.99	-.36
Z Reaction										
Time									1	-.35

Note. The contribution of the various word predictors to the visual lexical decision times in the MEGALEX data set. Data are limited to the words that were recognized by more than 67% of the participants ( $N = 15,842$ ). PLD20, phonological Levenshtein distance; PUP, phonological uniqueness point; Dur, acoustic duration; RT, reaction times; zRT, standardized reaction times; Acc, accuracy.

### 2.2.8. Regression Analyses in the Visual and Auditory Modalities

To further examine the impact of the predictor variables for the two modalities, we conducted multiple regression analyses at the item-level for each modality separately (25,776 words in the visual modality and 15,842 words in the auditory modality). Regressions have the advantage that non-linear relations can be examined (by making use of cubic splines). A forward approach was used in order to choose the nonlinear terms: i) at each step, such terms were included for the predictor variable for which the inclusion of splines lead to the greatest variance increase in comparison with a model in which no such terms were included, and ii) the stopping rule used to choose the number of knots was the value for which adding more knots did not lead to an increase of  $R^2$  above .001. Maximum number of knots were limited to 4 for length (N-letters and N-phonemes) and 7 for frequency, OLD20 (PLD20) and acoustic duration<sup>10</sup>. Table 8 shows the results.

**Table 8**  
*Results of the Forward Procedure Used to Choose Knots for the Predictor Variables*

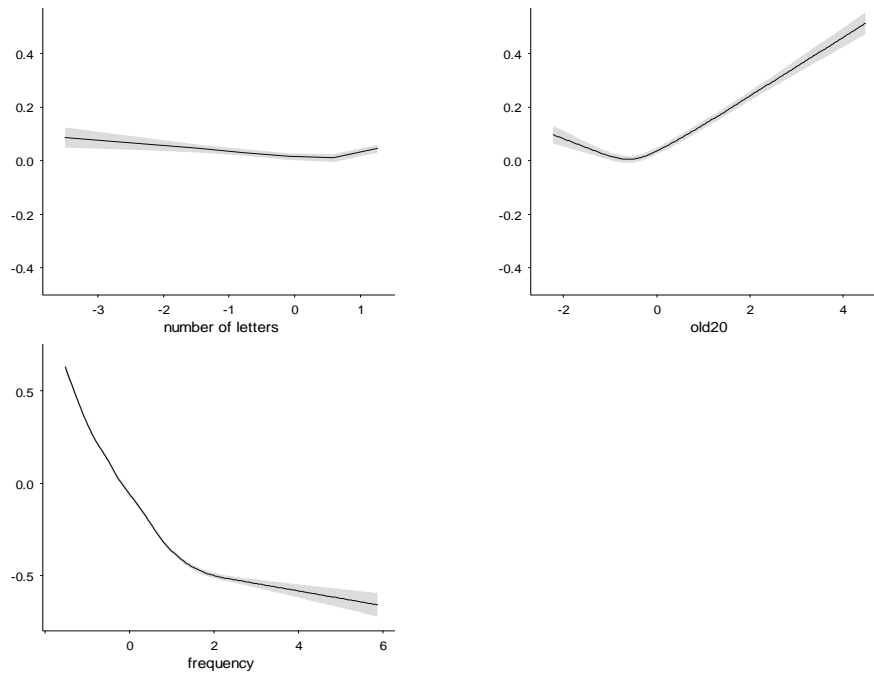
		init $R^2$	Frequency	OLD20	N-letters	Final $R^2$	
Visual (25,776 words)	RTs	.3829	.0437 (7)	.0074 (4)	.0011 (4)	.4351	
	Accuracy	.1897	.0496 (5)	.0039 (4)		.2432	
Auditory (15,842 words)	RTs	.5008	Frequency .0105 (5)	N-phonemes .0077 (3)	PLD20 .0016 (3)	AccD .0006 (4)	.5212
	Accuracy	.1061	Frequency .0282 (5)	Duration .0062 (4)	PLD20 .0021 (3)		.1426

*Note.* Predictor variables are shown by the order of entry of nonlinear terms in the equation. init (final)  $R^2 = R$  square with only linear terms (all nonlinear terms) included. For each line, first term =  $R^2$  increase when including the nonlinear terms; in brackets = number of knots for the variable.

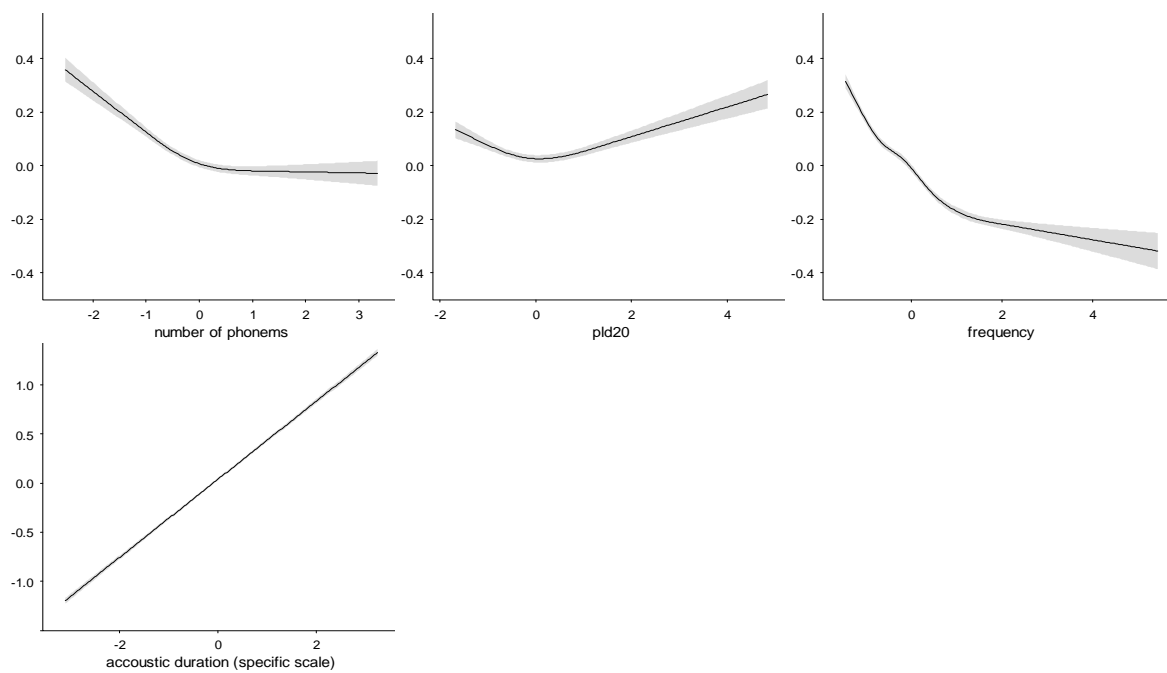
<sup>10</sup> Because uniqueness point values were both highly correlated with lengths, we postponed the analysis including this variable to Section 2.2.9. comparing effects across modalities and their interactions.

For both modalities and both dependent variables (see Tables 8 and 9), all included variables were significant at  $p < .001$ . For RTs, the effects of frequency and word similarity (OLD20 and PLD20) were nonlinear in both modalities, with relatively comparable patterns (see Figure 5).

**Visual**

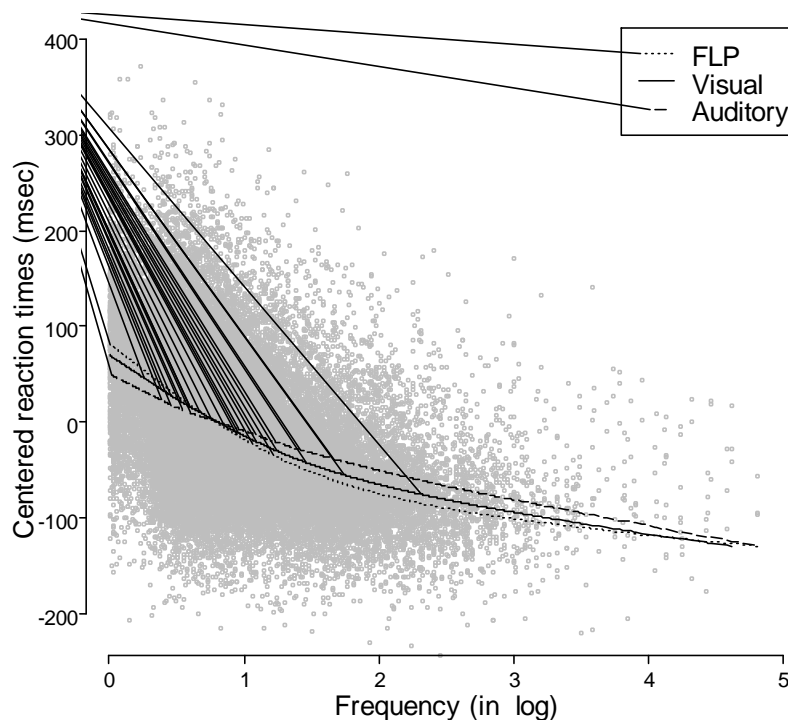


**Auditory**



**Figure 5.** The shape of the frequency effects, word length effects (number of letters and number of phonemes) and of word similarity effects (orthographic and phonological) in each modality (visual and auditory) for zRT. X-axis: standardized variable. Y-axis: predicted within task standardized zRT. 95% confidence intervals are indicated by gray shade.

As shown in Figure 5, for both modalities, word frequency effects (frequent words were responded to faster) were manifested in the significant negative relationship between RTs and frequency. Figure 6 also shows the shape of word frequency effects for both modalities in MEGALEX and for the visual modality in The French Lexicon Project (Ferrand et al., 2010) for 14,868 words in common. A weak negative effect of word length appeared in the visual modality (number of letters), whereas this effect (number of phonemes) was more pronounced in the auditory modality. For both modalities, the word similarity effect had a negative effect at the left of the scale and then inverted near the mean: similar words (low OLD20 or PLD20) were responded to more quickly than dissimilar words (high OLD20 or PLD20). Finally, acoustic duration showed the most important effect, linear and positive, in the auditory modality: words whose tokens had longer durations took longer to recognize.



**Figure 6.** The shape of word the frequency effects in the visual (black line) and auditory modality (dashed line) in MEGALEX, as well as in the French Lexicon Project (dotted line) for 14,868 words in common. The RT scales being different in MEGALEX (both modalities) and the French Lexicon Project), we centered RTs (RTs - MeanRTs).

Table 9 shows the results of regression analyses including all words in each modality.

**Table 9**  
*Regression Analyses With All Words in Each Modality (25,776 in the Visual Modality and 15,842 in the Auditory Modality)*

Visual $z$ RTs ( $R^2=.4351$ )					Auditory $z$ RTs ( $R^2=.5212$ )				
	SS	spR <sup>2</sup>	df	F		SS	spR <sup>2</sup>	df	F
Frequency	1852.477	0.330	6	2510.03	Frequency	294.563	0.062	4	512.45
Nonlinear	210.190	0.037	5	341.76	Nonlinear	37.541	0.008	3	87.08
OLD20	127.170	0.023	3	344.62	PLD20	14.678	0.003	2	51.07
Nonlinear	26.965	0.005	2	109.61	Nonlinear	8.091	0.002	1	56.30
Length (Nlett)	6.369	0.001	3	17.26	Length (Nphon)	29.996	0.006	2	104.37
Nonlinear	6.123	0.001	2	24.89	Nonlinear	8.938	0.002	1	62.20
					Duration	1144.683	0.241	3	2655.20
					Nonlinear	2.852	0.001	2	9.92
REGRESSION	2440.687		12	1653.51	REGRESSION	2475.937		11	1566.32
ERROR	3168.981		25763		ERROR	2274.824		15830	
Visual ACCURACY ( $R^2=.2432$ )					Auditory ACCURACY ( $R^2=.1426$ )				
	SS	spR <sup>2</sup>	df	F		SS	spR <sup>2</sup>	df	F
Frequency	32,962	0.219	4	1865.48	Frequency	9,990	0,108	4	497.18
Nonlinear	7,004	0.047	3	528.55	Nonlinear	2,285	0,025	3	151.6
OLD20	2,623	0.017	3	197.95	PLD20	0,195	0,002	2	19.38
Nonlinear	0,599	0.004	2	67.83	Nonlinear	0,194	0,002	1	38.71
Length (Nlett)	5,742	0.038	1	1299.93	Length (Nphon)	1,241	0,013	1	247.04
					Duration	0,429	0,005	3	28.5
					Nonlinear	0,428	0,005	2	42.62
REGRESSION	36.584		8	1035.26	REGRESSION	13,224	0,143	10	263.27
ERROR	113,821		25767		ERROR	79,522	0,857	15831	

Note. SS = sum of squares; spR<sup>2</sup> = semi-partial R square ( $R^2$  increase when the iv was the last entered in the equation)

For the visual modality, frequency, orthographic similarity and word length (in letters) explained 43.51% of variance in RTs and 24.32% in accuracy. For the auditory modality,

frequency, phonological similarity, word length (in phonemes) and acoustic duration explained 52.12% of variance in RTs and 14.26% in accuracy.

### 2.2.9. Comparing Effects Across Modalities and Their Interactions

Table 10 shows the intercorrelations between the various predictor variables and the RTs from MEGALEX for both the visual and auditory modalities. The analysis was limited to the 14,868 words common in both modalities. Table 10 shows in particular that the correlation between visual RTs and auditory RTs is limited ( $r = .37$ ), suggesting that there are differences between visual and auditory lexical decision times. These differences between modalities might suggest that people always wait until a perceptual decision can be reached, but are constrained differently across modalities<sup>11</sup>.

**Table 10**

*Bivariate and Multiple Correlations Between Variables in the MEGALEX Data Set for the Visual and the Auditory Modalities (14,868 Words in Common)*

	zRTv	Acc-v	RTa	zRTa	Acc-a	Dur	N <sub>lett</sub>	N <sub>phon</sub>	OLD	PLD	F <sub>total</sub>
RTv	.97	-.69	.37	.37	-.26	.21	.31	.29	.29	.25	-.62
zRTv		-.71	.37	.37	-.27	.20	.30	.28	.30	.24	-.64
Acc-v			-.22	-.22	.37	.01	.04	.02	-.08	.02	.40
RTa				.99	-.33	.66	.39	.47	.34	.33	-.33
zRTa					-.32	.68	.40	.48	.36	.33	-.33
Acc-a						.16	.12	.19	.07	.11	.15
Dur							.58	.74	.54	.46	-.23
N <sub>lett</sub>								.72	.59	.59	-.38
N <sub>phon</sub>									.62	.67	-.34
OLD										.54	-.25
PLD											-.27

Note. The contribution of the various word predictors to the visual and auditory lexical decision times in the MEGALEX data set. Data are limited to the words common to both modalities (14,868). RTv, visual reaction times; zRTv, standardized visual reaction times; Acc-v, visual accuracy; RTa, auditory reaction times; zRTa, standardized auditory reaction times; Acc-a, auditory accuracy; Dur, acoustic duration; OLD, orthographic Levenshtein distance; PLD, phonological Levenshtein distance; Ftotal, combination of subtitle and book frequencies.

A comparison of the visual and auditory lexical decision times of MEGALEX (14,868 words in common) reveals that whereas the effect of word frequency is the most important

<sup>11</sup> We would like to thank an anonymous reviewer for this interesting suggestion.

factor in the visual modality (accounting for at least 45% unique variance; see Table 11), stimulus duration is the main variable in the auditory modality (accounting for at least 46% of the variance; see Table 12).

**Table 11**

*Percentages of Variance Explained for the Visual Lexical Decision Task (14,868 Words in Common with the Auditory Lexical Decision Task) and for Frequency, Word Length (in Letters) and Orthographic Similarity (OLD20)*

	MEGALEX $R^2$ zRT
Frequency	41.21
Frequency + Frequency <sup>2</sup>	45.05
Frequency + Frequency <sup>2</sup> + N <sub>Letters</sub>	45.60
Frequency + Frequency <sup>2</sup> + N <sub>Letters</sub> + N <sub>Letters</sub> <sup>2</sup>	45.83
Frequency + Frequency <sup>2</sup> + N <sub>Letters</sub> + N <sub>Letters</sub> <sup>2</sup> + OLD20	47.34

Note. zRT, standardized reaction times.

**Table 12**

*Percentages of Variance Explained for the Auditory Lexical Decision Task (14,868 Words in Common with the Auditory Lexical Decision Task) and for Stimulus Duration, Frequency, Word Length (in Phonemes) and Phonological Similarity (PLD20)*

	MEGALEX $R^2$ zRT
Stimulus duration	46.02
Stimulus duration + Frequency	49.23
Stimulus duration + Frequency + Frequency <sup>2</sup>	50.20
Stimulus duration + Frequency + Frequency <sup>2</sup> + N <sub>Phonemes</sub>	50.87
Stimulus duration + Frequency + Frequency <sup>2</sup> + N <sub>Phonemes</sub> + N <sub>Phonemes</sub> <sup>2</sup>	51.68
Stimulus duration + Frequency + Frequency <sup>2</sup> + N <sub>Phonemes</sub> + N <sub>Phonemes</sub> <sup>2</sup> + PLD20	51.82

Note. zRT, standardized reaction times.

To further examine the impact of the predictor variables across the two modalities, we conducted multiple regression analyses at the item-level for the 14,868 words common in both modalities. In a first step, using the same forward procedure as above, two preliminary analyses were done within tasks in order to choose the numbers of splines to use in each. Both analyses included all variables under study, that is to say frequency subtitles, the numbers of letters and phonemes, the orthographic and phonologic Levenshtein distances and

acoustic duration. In a second step, a global analysis was conducted on scores from both tasks and the inclusion of nonlinear terms retained in the first step (when the numbers of knots differed between the tasks in the first step, the highest one was chosen). In addition, interactions between the two modalities and the predictor variables were entered in the regression model in order to compare the effects of the predictor variables between the two modalities. Tables 13 and 14 show the results of interaction effects and simple effects for both  $z$ RTs and accuracy.

**Table 13**  
*Across Modalities Comparisons for 14,868 Common Words in the Visual and Auditory Lexical Decision ( $z$ RTs)*

Interaction effects - $z$ RTs				
	SS	df1	F	p
Modality	3.4	1	2.8	<.05
Modality*Frequency.7	1068.1	6	450.2	<.0001
Modality*OLD20.6	4.5	5	11.5	<.001
Modality*PLD20.3	18.4	2	23.3	<.0001
Modality*Nlett	0.2	1	0.5	>.1
Modality*Nphon.3	33.5	2	42.3	<.0001
Modality*Duration	1787.3	1	4520.2	<.0001

Simple effects – $z$ RTs										
	Visual ( $R^2 = .478$ )					Auditory ( $R^2 = .52$ )				
	SS	spR <sup>2</sup>	df	F	p	SS	spR <sup>2</sup>	Df	F	p
Frequency	5311,34	0,3573	6	1692.74	<.0001	723,47	0,0487	6	251.11	<.0001
Nonlinear	472,86	0,0318	5	180.84	<.0001	105,21	0,0071	5	43.82	<.0001
OLD20	248,81	0,0167	5	95.16	<.0001	9,00	0,0006	5	3.75	<.01
Nonlinear	53,20	0,0036	4	25.43	<.0001	2,56	0,0002	4	1.33	>.1
PLD20	0,78	0,0001	2	0.75	>.1	48,41	0,0033	2	50.40	<.0001
Nonlinear	0,39	0,0000	1	0.75	>.1	22,48	0,0015	1	46.82	<.0001
Nlett	4,16	0,0003	1	7.96	<.01	2,00	0,0001	1	4.16	<.05
Nphon	8,01	0,0005	2	7.66	<0.001	102,38	0,0069	2	106.61	<.0001
Nonlinear	7,98	0,0005	1	15.26	<0.0001	52,64	0,0035	1	109.62	<.0001
Duration	0,02	0,0000	1	0.04	>.1	3557,29	0,2393	1	7408.31	<.0001
REG	7101,14		17	798.76	<.0001	7736,39		17	947.74	<.0001
ERROR	7765,86		14850			7130,61		14850		

Note. For interaction terms, p-values were computed using the satterthwaite approximation for degrees of freedom furnished by the lmerTest package of R (df2=14,868). Number of knots are added as “.n” expression next to the predictor variable abbreviation. Simple effects were tested using within tasks by item analyses.

For RTs, only the interaction between modalities and number of letters was not significant (all the other interactions between modalities and predictor variables were significant; see Table 13 and Figure 7). For the visual modality, frequency, orthographic similarity, phonological similarity, word length (in letters and in phonemes) and acoustic duration explained 47.80% of variance in RTs and 26.52% in accuracy. For the auditory modality, frequency, orthographic similarity, phonological similarity, word length (in letters and in phonemes) and acoustic duration explained 52% of variance in RTs and 12.30% in accuracy.

**Table 14**  
*Across Modalities Comparisons for 14,868 Common Words in the Visual and Auditory Lexical Decision (Accuracy)*

Interaction effects - Accuracy										
	SS	df1	F	p						
Modality	.0096	1	3.04	<.1						
Modality*Frequency.5	3.1056	4	246.93	<.0001						
Modality*OLD20.7	0.4329	6	22.95	<.0001						
Modality*PLD20.3	0.2014	2	32.03	<.0001						
Modality*Nlett.3	0.7818	2	124.33	<.0001						
Modality*Nphon	0.1387	1	44.10	<.0001						
Modality*Duration.4	0.1543	3	16.36	<.0001						

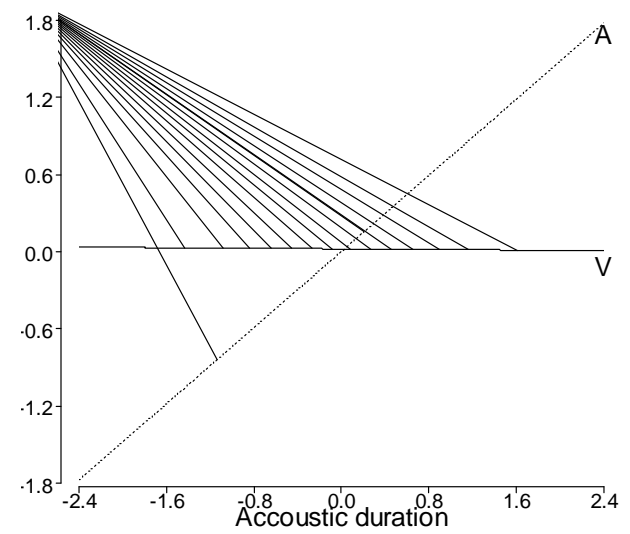
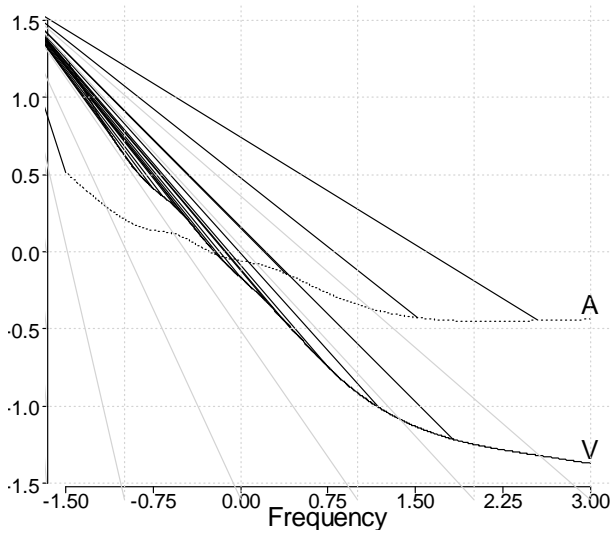
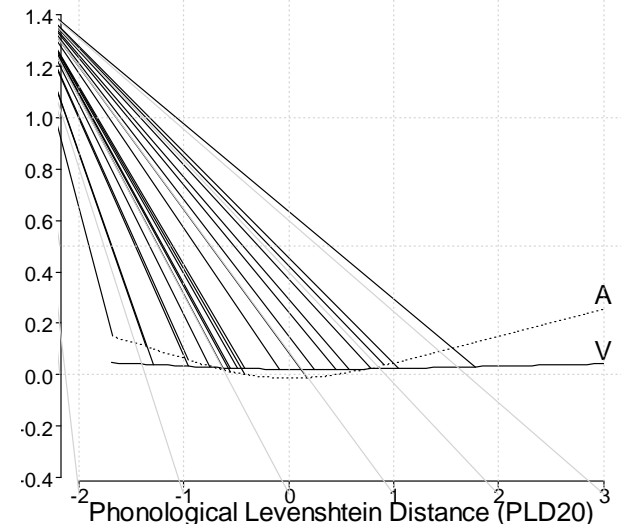
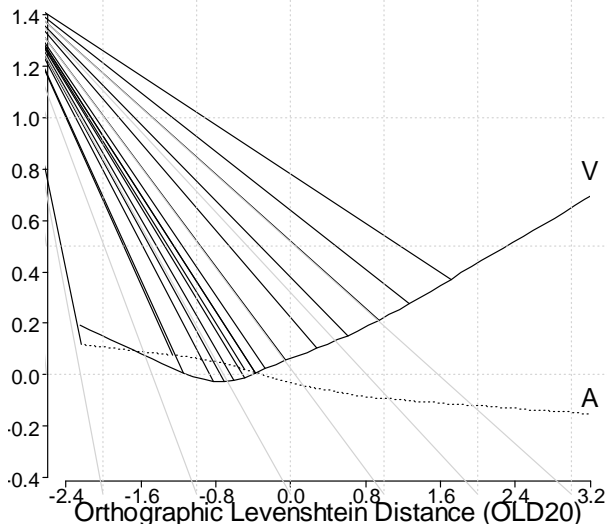
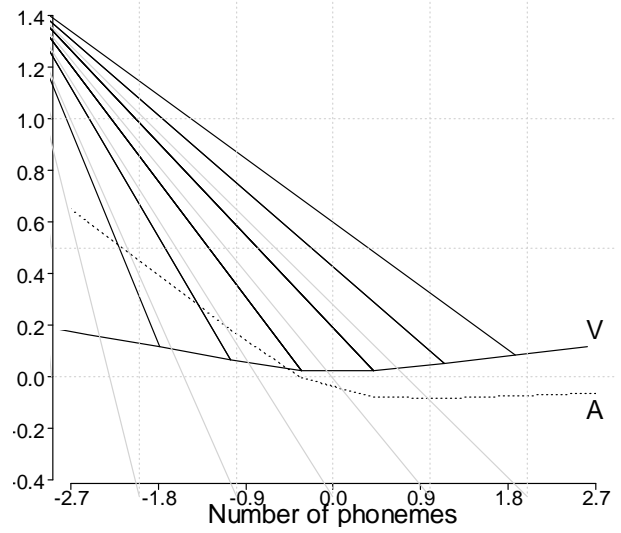
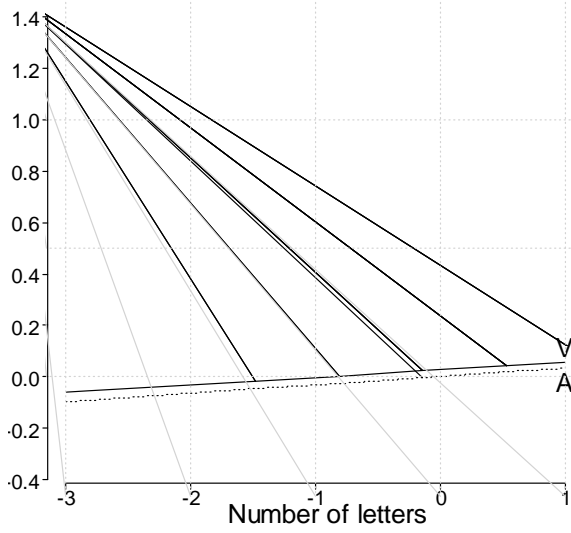
  

Simple effects – Accuracy										
	Visual ( $R^2 = .2652$ )					Auditory ( $R^2 = .123$ )				
	SS	spR <sup>2</sup>	df	F	p	SS	spR <sup>2</sup>	Df	F	p
Frequency	22.4136	0.2430	4	1227.50	<.0001	5.06	0.0706	4	298.66	<.0001
Nonlinear	3.6556	0.0396	3	266.94	<.0001	1.01	0.0141	3	79.41	<.0001
OLD20	1.8523	0.0201	6	67.63	<.0001	0.31	0.0043	6	12.13	<.0001
Nonlinear	0.4299	0.0047	5	18.83	<.0001	0.13	0.0018	5	6.18	<.0001
PLD20	0.1072	0.0012	2	11.74	<.0001	0.25	0.0036	2	30.06	<.0001
Nonlinear	0.0023	0.0000	1	0.50	>.1	0.25	0.0035	1	60.09	<.0001
Nlett	1.6493	0.0179	2	180.65	<.0001	0.00	0.0001	2	0.47	>.1
Nonlinear	0.0586	0.0006	1	12.84	<.001	0.00	0.0000	1	0.81	>.1
Nphon	0.1955	0.0021	1	42.84	<.0001	0.94	0.0131	1	221.51	<.0001
Duration	0.0580	0.0006	3	4.23	<.01	0.31	0.0043	3	24.04	<.0001
Nonlinear	0.0432	0.0005	2	4.73	<.01	0.28	0.0039	2	33.04	<.0001
REG	24.4651		18	297.74	<.0001	8.82		18	115.65	<.0001
ERROR	67.7840		14849			62.92		14849		

Note. For interaction terms, p-values were computed using the Satterthwaite approximation for degrees of freedom furnished by the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2013) of R (R Core Team, 2016) (df2=14,868). Numbers

of knots are added as “.n” expression next to the predictor variable abbreviation. Simple effects were tested using within tasks by item analyses.

Figure 7 show the interaction effects for RTs between modalities and predictor variables. The effect of number of letters did not interact with modalities: it was significant and positive in both modalities. Given the high number of observations, it however appeared quite weak. All the other effects interacted with modalities. The effect of number of phonemes was also significant in both modalities, with a negative sign in the left part of the scale, becoming then slightly positive. This pattern was more pronounced in the auditory modality. Phonological similarity and acoustic duration effects were only observed in the auditory modality. Orthographic similarity effect had a comparable pattern to the phonological one, but it was more pronounced in the visual modality; a weak negative effect - nearly linear - was however also observed in the auditory modality. Finally, classical frequency effects were observed in both modalities, with more important decreases at the left of the scale. The effect was however more pronounced in the visual modality. These results were generally in agreement with those obtained previously using all the words available in both modalities. The only exception concerned the effect of number of letters in the visual modality which was negative in the preceding analysis and positive in this one. This effect was however rather weak in both analyses.

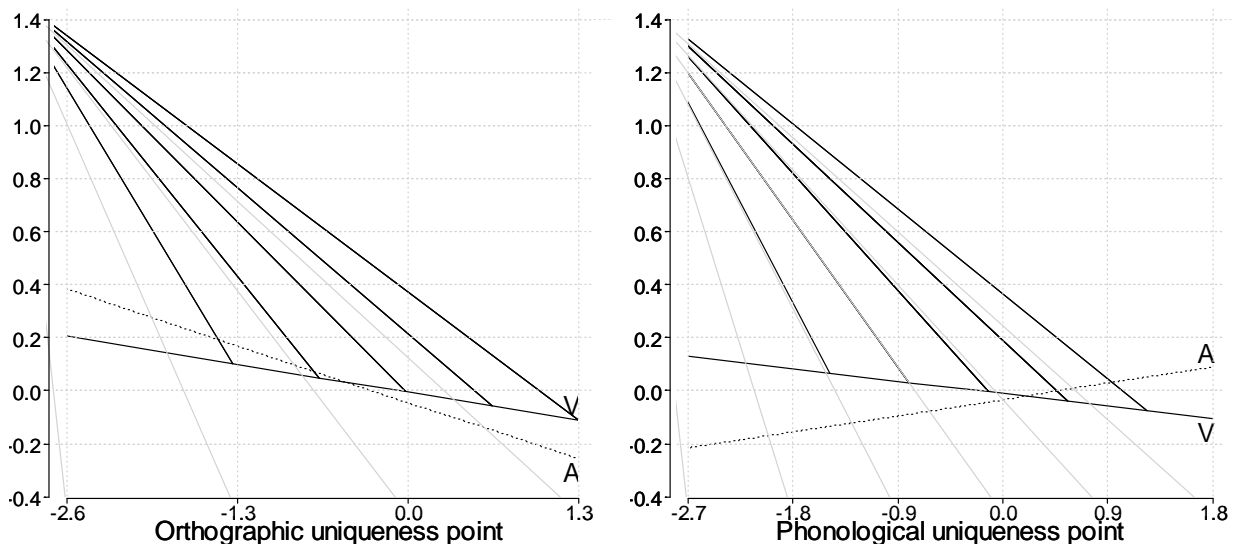


**Figure 7.** The shape of the word frequency effects, word length effects (number of letters and number of phonemes), word similarity effects (orthographic and phonological) and acoustic duration for both modalities (V: visual and A: auditory) for  $z$ RT. X-axis: standardized variable. Y-axis: predicted within task standardized  $z$ RT.

A similar analysis was conducted with the inclusion of orthographic uniqueness point and phonological uniqueness point values. For the visual modality, frequency, orthographic similarity, phonological similarity, word length (in letters and in phonemes), and uniqueness point (orthographic and phonological) explained 47.95% of variance in RTs and 26.84% in accuracy. For the auditory modality, frequency, orthographic similarity, phonological similarity, word length (in letters and in phonemes), acoustic duration and orthographic and phonological uniqueness point explained 52.50% of variance in RTs and 13.26% in accuracy.

Some studies have reported that words with an early OUP were processed more quickly than words with a late OUP in the visual modality (Kwantes & Mewhort, 1999; Lindell, Nicholls, & Castles, 2003), an effect similar to the effect obtained in auditory tasks (e.g., Radeau, Mousty, & Bertelson, 1989; Radeau & Morais, 1990). This has been taken to suggest that readers process the letters of words sequentially in a left-to-right order (but see Lamberts, 2005, for another interpretation). However, other studies found the opposite pattern, words with an early OUP were processed more slowly than words with a late OUP (Izura, Wright, & Fouquet, 2014; Miller, Juhasz, & Rayner, 2006; Radeau, Morais, Mousty, Saerens, & Bertelson, 1992), suggesting that word processing is likely to occur in a parallel manner. Here, we found that words with an early OUP were processed more slowly than words with a late OUP in the visual modality (see Figure 8). These results are consistent with those of Izura et al. (2014), Miller et al. (2006) and Radeau et al. (1992). The same OUP effect was obtained in the auditory modality. To our knowledge, this is the first time that an OUP effect is reported in this modality.

In the auditory modality, it has also been reported that words with an early PUP were processed more quickly than words with a late PUP (Radeau, Morais, Mousty, & Bertelson, 2000; Radeau, Mousty, & Bertelson, 1989; Radeau & Morais, 1990). In particular, Radeau et al. (2000) found a PUP effect only in the slow rate condition, not in the fast one (this last condition being closer to natural speech). They also reported that the PUP effects were smaller with synthetic than with natural speech. Here, we found that words with an early PUP were processed more quickly than words with a late PUP (see Figure 8), replicating Radeau et al. (2000). However, concerning the visual modality, words with an early PUP were processed more slowly than words with a late PUP, an effect similar to the OUP.



**Figure 8.** The shape of the orthographic uniqueness point and phonological uniqueness point effects for both modalities (V: visual and A: auditory) for zRT. Analysis conducted on 14,868 words present in both modalities. X-axis: standardized variable. Y-axis: predicted within task standardized zRT.

### 3. General Discussion

#### 3.1. Summary of Findings

In general, the findings of the visual lexical decision times of MEGALEX are consistent with the results of previous megastudies conducted in English (Balota et al., 2007; Brysbaert

et al., 2016; Cortese & Khanna, 2007; Keuleers et al., 2012; Yap & Balota, 2009), Dutch (Brysbaert et al., 2016; Keuleers et al., 2010) and French (Ferrand et al., 2010). Without question, word frequency is the most important predictor in visual lexical decision (accounting for at least 45% unique variance in our study; 40% in ELP; Balota et al., 2007; Brysbaert et al., 2016; 34% in DLP1; Keuleers et al., 2010; 33% in DLP2; Brysbaert et al., 2016; and 38% in FLP; Ferrand et al., 2010). In line with what Ferrand et al. (2010) reported for French, the extra contribution of word length was significant but rather limited (0.8%). The extra contribution of orthographic similarity was larger (around 1.5%). Furthermore, the present megastudy confirms the superiority of a design in which the participants process half of the stimuli (MEGALEX) relative to a design in which participants only see a subset of the stimuli (FLP). This is in line with findings from the Dutch and the British Lexicon Projects, indicating that there is less noise in megastudies with a complete cross-over of participants and stimuli (Keuleers et al., 2010, 2012).

Concerning the auditory modality, stimulus duration is the most important variable (accounting for 46% of the variance in our study), which is consistent with the results of a previous relatively large-scale study conducted in Dutch (Ernestus & Cutler, 2015; see also Goh et al., 2016). The extra contribution of frequency is around 4% and the extra contribution of number of phonemes is 1.5%. Phonological similarity and phonological uniqueness point also influenced auditory lexical decision times (less than 1%). However, we showed that words that are phonologically similar to many other words (i.e., with low PLD20 values) were responded to faster than phonologically dissimilar words (i.e., with high PLD20 values). This latter result departs from the general finding that, in the auditory modality, words with more similar sounding (or closer phonological neighbours) are usually recognized more *slowly* than more distinct word-forms (e.g., Goh et al., 2009, 2016; Suarez, Tan, Yap, & Goh, 2011; Ziegler et al., 2003), but it converges with the general finding that, in the visual

modality, words that are orthographically similar to many other words are responded to faster than orthographically dissimilar words (e.g., Brysbaert et al., 2016; Ferrand et al., 2010; Keuleers et al., 2010, 2012). Concerning phonological uniqueness point, our results are consistent with the general finding that words with an early phonological uniqueness point are responded to faster than words with a late phonological uniqueness point (e.g., Radeau et al., 1989, 1990, 2000).

### **3.1.1. Similarities and Differences Across Modalities**

A comparison of the effects of lexical variables across modalities revealed interesting differences between auditory and visual lexical decision times, as was predicted by Goldinger (1996; see also Segui, 1994, for a discussion). A first difference across modalities is that word frequency is the most important factor in the visual modality, whereas it is word duration in the auditory modality. A second difference is that the proportion of variance explained by frequency is relatively low in the auditory modality (11.3%) compared to the visual modality (45.1%). A third difference is that listeners tended to wait until they had heard the last phoneme in the word before making their decision (as shown by a strong effect of stimulus duration), whereas readers can make their decision before the word has been identified (as shown by a strong effect of OLD20). Indeed, in the visual modality, as suggested by Ernestus and Cutler (2015, p. 1471), “Participants can see at a glance whether a presented form is long or short, but in the auditory modality, listeners must wait for silence to tell them that the presented form has ended”. In the visual modality, as suggested by Brysbaert et al. (2016, p. 443), “Lexical decisions are partly based on the overall activation in the orthographic lexicon, such that when this overall activation exceeds a certain level, a word response is initiated before the target word itself has been identified. This overall activation is higher when a target word activates many similar word forms than when it has a unique letter sequence” (see

Grainger & Jacobs, 1996; see also Dufau, Grainger, & Ziegler, 2012; Grainger, 2017). A fourth difference is that readers process words with an early orthographic uniqueness point more slowly than words with a late orthographic uniqueness point (see also Izura et al. 2014; Miller et al. 2006; Radeau et al., 1992), whereas listeners process words with an early phonological uniqueness point more quickly than words with a late phonological uniqueness point (see also Radeau et al., 1989, 2000). These results suggest that processing of visually presented words is likely to occur in a parallel manner, whereas processing of auditory presented words is likely to occur in a sequential manner. A fifth difference is that more variance is explained in the auditory modality (max=53%) than in the visual modality (max=48%).

Turning to those effects that revealed a similar pattern in the two modalities, a first converging pattern concerns word similarity: words orthographically similar to many other words (i.e., with low OLD20 values) were responded to more quickly than more orthographically distinct words (i.e., with high OLD20 values) in the visual modality, and words phonologically similar to many other words (i.e., with low PLD20 values) were responded to faster than more phonologically distinct words (i.e., with high PLD20 values) in the auditory modality (see Figures 5 and 7). A second converging pattern concerns the shape of the word frequency effect in both modalities. Even if less proportion of variance is explained by frequency for the auditory modality compared to the visual modality, the shapes of the word frequency effects are very similar (see Figure 6).

#### **4. Future Directions**

Although this study has examined the influence of a certain number of measures on visual and auditory word recognition performance, a number of questions remain open. For instance, it will be useful to test other potentially interesting variables (when these variables will be

available for French), such as age of acquisition (AoA; see e.g., Bonin, Barry, Méot, & Chalard, 2004; Cortese & Khanna, 2007; Ferrand, Bonin, Méot, Augustinova, New, Pallier, & Brysbaert, 2008), word prevalence (Wp; see e.g., Keuleers, Stevens, Mandera, & Brysbaert, 2015), sensory experience rating (SER; see e.g., Bonin et al., 2015), semantic variables (such as imageability, concreteness or number of semantic features, to name a few; see e.g., Goh et al., 2016; Yap, Pexman, Wellsby, Hargreaves, & Huff, 2012), morphological variables (see e.g., Baayen et al., 2006; Gimenes, Brysbaert, & New, 2016), etc. Indeed, the maximal percentages of variance explained (48% in the visual modality and 53% in the auditory modality) suggest that there are other variables to discover.

## **5. Conclusion**

In this article, we have described the results of a project in which lexical decision data were collected for tens of thousand French words presented in the visual and auditory modalities. The results of the present megastudy provide an empirical “goldmine” for researchers interested in understanding the core mechanisms of word recognition, independent of modality, as well as for providing a better specification of modality-specific processing. This is clearly a crucial step toward developing a general psycholinguistic theory of word comprehension.

## Acknowledgments

This work was supported by an ANR Grant n° 12-CORP-0001 (Agence Nationale de la Recherche) to Ludovic Ferrand. We would like to thank Mickaël Berthon, Pierre Chausse, and Julien Guégan from the LAPSCO and CATECH (Université Clermont Auvergne, France) for their invaluable help at different stages of the project.

## Appendix A. Supplementary material

The data for the words and the pseudowords are made available as two Excel files and text files. These files can be found at our website (<https://sedufau.shinyapps.io/megalex/>) and are searchable on <http://www.lexique.org> inside the Open Lexique search engine. Our Web site also allows researchers to correlate the MEGALEX data with the word characteristics available for French words at the Web site and to generate lists of words that correspond to certain constraints. Each Excel and text file contains the following columns: Item (word or pseudoword), Ntrials (total number of observation for the item), Err (percentage of errors), RT (mean RT of the correct trials for the item), SD (standard deviation of the RTs for that item) and  $z$ RT (mean RT of the correct trials for the item after the RTs of the individual participants have been transformed into standardized  $z$  scores). Furthermore, the soundfiles (for words and nonwords) are available upon request.

## References

- Adelman, J. S., Marquis, S. J., Sabatos-DeVito, M. G., & Estes, Z. (2013). The unexplained nature of reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1037-1053.
- Baayen, R. H., Feldman, L. F., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language*, *55*, 290-313.
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.
- Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition for single syllable words. *Journal of Experimental Psychology: General*, *133*, 283-316.
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. I., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*, 445-459.
- Balota, D. A., Yap, M. J., Hutchison, K. A., & Cortese, M. J. (2013). Megastudies: What do millions (or so) of trials tell us about lexical processing? In Adelman, J. S. (Ed.), *Visual word recognition: Models and methods, orthography and phonology*. Hove, UK: Psychology Press.
- Bonin, P., Barry, C., Méot, A., & Chalard, M. (2004). The influence of age of acquisition in word reading and other tasks: A never ending story? *Journal of Memory and Language*, *50*, 456-476.
- Bonin, P., Méot, A., Ferrand, L., & Bugajska, A. (2015). Sensory experience ratings (SERs) for 1,659 French words: Relationships with other psycholinguistic variables and visual word recognition. *Behavior Research Methods*, *47*, 813-825.

- Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*, 977-990.
- Brysbaert, M., Stevens, M., Mander, 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*, 441-458.
- Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*, 904-911.
- Cleland, A. A., Gaskell, M. G., Quinlan, P. T., & Tamminen, J. (2006). Frequency effects in spoken and visual word recognition: Evidence from dual-task methodologies. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 104-119.
- Cleveland, W. S. (1981) LOWESS: A program for smoothing scatterplots by robust locally weighted regression. *The American Statistician, 35*, 54.
- Connine, C. M., Mullenix, J., Shernoff, E., & Yellen, J. (1990). Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 1084-1096.
- Cortese, M. J., & Khanna, M. M. (2007). Age of acquisition predicts naming and lexical-decision performance above and beyond 22 other predictor variables: An analysis of 2,342 words. *Quarterly Journal of Experimental Psychology, 60*, 1072-1082.
- Dufau, S., Duñabeita, J. A., Moret-Tatay, C., McGonigal, A., Peeters, D., Alario, F. X., Balota, D. A., Brysbaert, M., Carreiras, M., Ferrand, L., Ktori, M., Perea, M., Rastle, K., Sasburg, O., Yap, M. J., Ziegler, J. C., & Grainger, J. (2011). Smart phone, smart science: How the use of smartphones can revolutionize research in cognitive science. *Plos One, 6(9): e24974*. doi:10.1371/journal.pone.0024974

- Dufau, S., Grainger, J., & Ziegler, J. C. (2012). How to say "no" to a nonword: A leaky competing accumulator model of lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1117-1128.
- Dufau, S., Grainger, J., Midgley, K. J., & Holcomb, P. J. (2015). A thousand words are worth a picture: Snapshots of printed-word processing in an event-related potential megastudy. *Psychological Science*, *26*, 1887-1897.
- Ernestus, M., & Cutler, A. (2015). BALDEY: A database of auditory lexical decisions. *The Quarterly Journal of Experimental Psychology*, *68*, 1469-1488.
- Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Individual differences in information-processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin*, *125*, 777-799.
- Ferrand, L., Bonin, P., Méot, A., Augustinova, M., New, B., Pallier, C., & Brysbaert, M. (2008). Age of acquisition and subjective frequency estimates for all generally known monosyllabic French words and their relation with other psycholinguistic variables. *Behavior Research Methods*, *40*, 1049-1054.
- Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, 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*, 488-496.
- Ferrand, L., Brysbaert, M., Keuleers, E., New, B., Bonin, P., Méot, A., Augustinova, M., & Pallier, C. (2011). Comparing word processing times in naming, lexical decision, and progressive demasking: Evidence from Chronolex. *Frontiers in Psychology*. doi: 10.3389/fpsyg.2011.00306
- Ferrand, L., Méot, A., Spinelli, E., New, B., Pallier, C., Bonin, P., Dufau, S., Mathôt, S., & Grainger, J. (2015). *MEGALEX- A new mega-study of visual word recognition: Some*

*preliminary data*. 19<sup>th</sup> Meeting of the European Society for Cognitive Psychology (ES COP 2011), Paphos, Cyprus.

- Forster, K. I. (2000). The potential for experimenter bias effects in word recognition experiments. *Memory & Cognition*, 28, 1109-1115.
- Gimenes, M., Brysbaert, M., & New, B. (2016). The processing of singular and plural nouns in English, French, and Dutch: New insights from megastudies. *Canadian Journal of Experimental Psychology*, 70, 316-324.
- Gimenes, M., & New, B. (2016). Worldlex: Twitter and blog word frequencies for 66 languages. *Behavior Research Methods*, 48, 963-972.
- Goh, W. D., Suarez, L., Yap, M. J., & Tan, S.H. (2009). Distributional analyses in auditory lexical decision: Neighborhood density and word-frequency effects. *Psychonomic Bulletin & Review*, 16, 882-887.
- Goh, W. D., Yap, M. J., Lau, M. C., Ng, M. M. R., & Tan, L-C. (2016). Semantic richness effects in spoken word recognition: A lexical decision and semantic categorization megastudy. *Frontiers in Psychology*. doi: 10.3389/fpsyg.2016.00976
- Goldinger, S. D. (1996). Auditory lexical decision. *Language & Cognitive Processes*, 11, 559-567.
- Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, 28, 501-518.
- Goodman, J. C., & Huttenlocher, J. (1988). Do we know how people identify spoken words? *Journal of Memory and Language*, 27, 684-698.
- Grainger, J. (2017). Orthographic processing: A “mid-level” vision of reading. *Quarterly Journal of Experimental Psychology*.  
<http://dx.doi.org/10.1080/17470218.2017.1314515>

- Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, *103*, 518-565.
- Izura, C., Wright, V., & Fouque, N. (2014). Hemispheric asymmetries in word recognition as revealed by the orthographic uniqueness point. *Frontiers in Psychology*, *5*, 244. Doi: 10.3389/fpsyg.2014.00244
- Juhasz, B. J., & Yap, M. J. (2013). Sensory experience ratings for over 5,000 mono- and disyllabic words. *Behavior Research Methods*, *45*, 160–168.
- Juhasz, B., Yap, M. J., Dicke, J., Taylor, S. C., & Gullick, M. M. (2011). Tangible words are recognized faster: The grounding of meaning in sensory and perceptual systems. *Quarterly Journal of Experimental Psychology*, *64*, 1683-1691.
- Keuleers, E., & Balota, D. (2015). Megastudies, crowdsourcing, and large datasets in psycholinguistics: An overview of recent developments. *Quarterly Journal of Experimental Psychology*, *68*, 1457-1468.
- Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new frequency measure for Dutch words based on film subtitles. *Behavior Research Methods*, *42*, 643-650.
- 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 Language Sciences, Psychology*, *1*, 174, doi: 10.3389/fpsyg.2010.00174
- Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, *44*, 287-304.

- 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*, 1665-1692.
- Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology: General*, *143*, 1065-1081.
- Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30 thousand English words. *Behavior Research Methods*, *44*, 978-990.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2013). lmerTest. Retrieved from <https://cran.r-project.org/web/packages/lmerTest/>
- Kwantes, P. J., & Mewhort, D. J. K. (1999). Evidence for sequential processing in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 376-381.
- Lamberts, K. (2005). Interpretation of orthographic uniqueness point effects in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 14-19.
- Lindell, A. K., Nicholls, M. E. R., & Castles, A. E. (2003). The effect of orthographic uniqueness and deviation points on lexical decisions: Evidence from unilateral and bilateral-redundant presentations. *Quarterly Journal of Experimental Psychology*, *56*, 287-307.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*, 1-36.
- Marlsen-Wilson, W. D. (1990). Activation, competition, and frequency in lexical access. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing*. Cambridge, MA: MIT Press.

- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*, 314-324.
- Miller, B., Juhasz, B. J., & Rayner, K. (2006). The orthographic uniqueness point and eye movements during reading. *British Journal of Psychology*, *97*, 191-216.
- New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics*, *28*, 661-677.
- New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review*, *13*, 45-52.
- New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE. *L'Année Psychologique*, *101*, 447-462.
- New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior, Research, Methods, Instruments, & Computers*, *36*, 516-524.
- Petrova, A., Gaskell, G., & Ferrand, L. (2011). Orthographic consistency and word-frequency effects in auditory word recognition: New evidence from lexical decision and rime detection. *Frontiers in Psychology*, *2*:263. doi:10.3389/fpsyg.2011.00263
- Radeau, M., & Morais, J. (1989). The uniqueness point effect in shadowing of spoken words. *Speech Communication*, *9*, 155-164.
- Radeau, M., Morais, J., Mousty, P., Saerens, M., & Bertelson, P. (1992). A listener's investigation of printed word processing. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 861-871.
- Radeau, M., Morais, J., Mousty, P., & Bertelson, P. (2000). The effect of speaking rate on the role of the uniqueness point in spoken word recognition. *Journal of Memory and Language*, *42*, 406-422.

- Radeau, M., Mousty, P., & Bertelson, P. (1989). The effect of the uniqueness point in spoken-word recognition. *Psychological Research*, *51*, 123-128.
- R Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria. Retrieved from <http://www.r-project.org>
- Schröter, P., & Schroeder, S. (2017). The Developmental Lexicon Project: A behavioral database to investigate visual word recognition across the lifespan. *Behavior Research Methods*. Doi 10.3758/s13428-016-0851-9
- Segui, J. (1994). Language perception in visual and auditory modalities: Similarities and differences. In P. Bertelson & P. Eelen (Eds.), *International Perspectives on Psychological Science, II: The State of the Art*. Hove: Psychology Press.
- Seidenberg, M. S., & Waters, G. S. (1989). Word recognition and naming: A mega study. *Bulletin of the Psychonomic Society*, *27*, 489.
- Shimizu, H. (2002). Measuring keyboard response delays by comparing keyboard and joystick inputs. *Behavior Research Methods, Instruments, & Computers*, *34*, 250-256.
- Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science*, *8*, 411-416.
- Suarez, L., Tan, S. H., Yap, M. J., & Goh, W. D. (2011). Observing neighbourhood effects without neighbours. *Psychonomic Bulletin and Review*, *18*, 605-611.
- Taft, M., & Hambly, G. (1986). Exploring the cohort model of spoken word recognition. *Cognition*, *22*, 259-282.
- Tse, C-H., Yap, M. J., Chan, Y-L., Sze, W-P., Shaoul, C., & Lin, D. (2016). The Chinese Lexicon Project: A megastudy of lexical decision performance for 25,000+ traditional Chinese two-character compound words. *Behavior Research Methods*. Doi 10.3758/s13428-016-0810-5

- Treiman, R., Mullenix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General*, *124*, 107-136.
- Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, *40*, 374-408.
- Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language*, *60*, 502-529.
- Yap, M. J., & Brysbaert, M. (2009). *Auditory lexical decision: On the relative weights of word frequency, neighborhood density, word length, and onset duration*. Unpublished manuscript retrieved at <http://crr.ugent.be/members/marc-brysbaert#pu6>
- Yap, M. J., Rickard Liow, S. J., Jalil, S. B., & Faizal, S. S. B. (2010). The Malay Lexicon Project: A database of lexical statistics for 9,592 words. *Behavior Research Methods*, *42*, 992-1003.
- Yap, M. J., Pexman, P. M., Wellsby, M., Hargreaves, J. S., & Huff, M. (2012). An abundance of riches: Cross-task comparisons of semantic richness effects in visual word recognition. *Frontiers in Human Neuroscience*, *6*:72. doi: 10.3389/fnhum.2012.00072
- Yarkoni, T., Balota, D. A., & Yap, M. (2008). Moving beyond Coltheart's *N*: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*, 971-979.
- Ziegler, J. C., Muneaux, M., & Grainger, J. (2003). Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory and Language*, *48*, 779-793.

**Table 2.** Descriptive statistics of the words and behavioral measures (RT,  $z$ RT, and accuracy) involved in item-level regression analyses used in the visual and auditory lexical decision tasks

Modality	RT	$z$ RT	Accuracy	N-Letters	OLD20	FreqBooks	FreqFilms	OUP		
Visual	Mean	603.31 (596.73)	0.09 (0.03)	0.934 (0.940)	7.21 (7.14)	2.02 (1.96)	31.37 (29.75)	32.21 (30.38)	7.02 (6.98)	
	SD	59.08 (56.86)	0.48 (0.47)	0.079 (0.076)	1.49 (1.47)	0.45 (0.43)	498.75 (448.75)	453.59 (441.74)	1.52 (1.50)	
	Min	469.54 (469.29)	-1.01 (-1.06)	0.667	2	1	0	0	1	
	Max	881.75	2.20	1	9	4.55	38,928.9 (38,928.9)	25,220.9 (25,988.4)	9	
Modality	RT	$z$ RT	Accuracy	N-Phonemes	PLD20	FreqBooks	FreqFilms	PUP	Duration	
Auditory	Mean	1063.44 (1066.91)	0.00 (0.02)	0.947 (0.940)	5.44	2.27 (2.26)	31.37 (29.53)	32.21 (30.27)	5.18 (5.16)	635.99 (635.83)
	SD	81.44 (82.46)	0.54 (0.55)	0.069 (0.077)	1.36	0.75	498.75 (483.23)	453.59 (439.49)	1.47	124.40 (124.49)
	Min	819.78	-1.58	0.667	1	1	0	0	1	167
	Max	1433.86 (1433.86)	2.08 (2.1)	1	11	6.6	38,928.9	25,220.9	11	1149

*Note.* These characteristics are taken from *Lexique* (New et al., 2001, 2004, 2007; [www.lexique.org](http://www.lexique.org)) and from the *French Lexicon Project* (Ferrand et al., 2010; <https://sites.google.com/site/frenchlexicon/>). Statistics are given after elimination rules were applied. For each task, plain text = common words for both tasks (14,868); (brackets)

## A megastudy of visual and auditory word recognition

= for all words (visual: 25,776, auditory: 15,842) - if different from the preceding value.. RT, reaction times; zRT, standardized reaction times; OLD20, orthographic Levenshtein distance; PLD20, phonological Levenshtein distance; OUP, orthographic uniqueness point; PUP, phonological uniqueness point.