Comparing word processing times in naming, lexical decision, and progressive demasking: evidence from Chronolex

Ludovic Ferrand[1]*, Marc Brysbaert[2], Emmanuel Keuleers[2], Boris New[3], Patrick Bonin[4], Alain Méot[1], Maria Augustinova[1] and Christophe Pallier[5]

[1] Laboratoire de Psychologie Sociale et Cognitive, Centre National de la Recherche Scientifique, University Blaise Pascal, Clermont-Ferrand, France
[2] Department of Experimental Psychology, Ghent University, Ghent, Belgium
[3] Laboratoire de Psychologie et Neuropsychologie Cognitive, Centre National de la Recherche Scientifique, University Paris Descartes, Paris, France
[4] Laboratoire d'Etude de l'Apprentissage et du Développement, Centre National de la Recherche Scientifique, University of Bourgogne and Institut Universitaire de France, Dijon, France
[5] Cognitive Neuroimaging Unit, Institut National de la Santé et de la Recherche Médicale, Commissariat à l'Energie Atomique, Saclay, France

We report performance measures for lexical decision (LD), word naming (NMG), and progressive demasking (PDM) for a large sample of monosyllabic monomorphemic French words ($N = 1,482$). We compare the tasks and also examine the impact of word length, word frequency, initial phoneme, orthographic and phonological distance to neighbors, age-of-acquisition, and subjective frequency. Our results show that objective word frequency is by far the most important variable to predict reaction times in LD. For word naming, it is the first phoneme. PDM was more influenced by a semantic variable (word imageability) than LD, but was also affected to a much greater extent by perceptual variables (word length, first phoneme/letters). This may reduce its usefulness as a psycholinguistic word recognition task.

Keywords: megastudy approach, cross-task comparisons, visual word recognition, word naming, lexical decision, progressive demasking, word processing times

INTRODUCTION

Visual word recognition in adults is fast, efficient, and relatively effortless, suggesting that word processing times will be the same for all words and invariant across tasks. This view underestimates the complexity of the processes involved. In recent years, authors have focused on the variables affecting the speed and accuracy with which words are processed (e.g., Balota et al., 2004; Baayen et al., 2006; Lemhöfer et al., 2008; Yap and Balota, 2009). In the present study, we focus on ways in which word processing times converge or diverge across three different tasks: lexical decision (LD), word naming (NMG), and progressive demasking (PDM).

Lexical decision and word naming are by far the most frequently used word processing tasks. They have been analyzed extensively for the English language on the basis of megastudies run by Balota, Spieler, and colleagues (Spieler and Balota, 1997, 2000; Balota et al., 2004, 2007). Megastudies are studies in which word processing times are gathered for a large, representative sample of the language (see, e.g., Balota et al., in press, for a review). The findings are typically explored with regression analyses. From these analyses it has emerged that word frequency is the most important predictor of visual LD times, accounting for up to 40% of the variance (of which 25% cannot be accounted for by other variables; Baayen et al., 2006; Cortese and Khanna, 2007; Brysbaert and Cortese,

2011)[1]. In contrast, for naming times of monosyllabic words the articulatory features of the initial phoneme are the most important, explaining up to 40% of the variance (Balota et al., 2004; Bonin et al., 2004; Cortese and Khanna, 2007). In naming, word frequency explains less that 10% of the variance (of which 6% is

---

[1]As pointed out by one reviewer, other studies have investigated the same issues with factorial designs (e.g., Gernsbacher, 1984; Gardner et al., 1987; Monsell et al., 1989; Schreuder and Baayen, 1997; Andrews and Heathcote, 2001). However, we focus our introduction on megastudies using regression designs because these designs have a number of important advantages over the more commonly used factorial designs. For instance, as pointed out by Balota et al. (2004, in press; see also Baayen et al., 2006), there are a number of problems that arise from the *a priori* selection of restricted sets of stimuli in factorial designs, such as difficulties with matching the stimuli on all relevant dimensions, the possible occurrence of experimenter biases during stimulus selection, and the sometimes disproportionate use of words that take extreme values on the target dimensions. In contrast, in regression designs, the variables of interest are not used as selection criteria of the stimulus set, thereby avoiding these difficulties. Furthermore, given that most variables under investigation are continuous (e.g., word frequency), regression analyses avoid the loss of information that is associated with the categorization of continuous variables. Finally, regression designs allow for establishing not only whether a particular factor has an effect on the dependent variable but also how large the contribution of this factor is in accounting for the measured variance; they also allow comparisons across experimental tasks (e.g., lexical decision and naming). Because of these advantages, an increasing number of large-scale psycholinguistic studies have recently used regression designs and we focus mainly on these studies.

pure), implying that, for word naming, it is more critical to match conditions on the first phoneme than on frequency. In addition, LD times are influenced to a greater extent by the morphological and semantic characteristics of the words than word naming times (Balota et al., 2004).

Progressive demasking has been used much less. It is a variant of the once popular perceptual identification task in which participants try to recognize tachistoscopically presented words and which fell from grace because it did not provide RT data (Monsell, 1991). The PDM task was developed by Feustel et al. (1983), who called it the continuous threshold latency identification task. It was given its current name by Grainger and Segui (1990; see also Dufau et al., 2008). The task was designed to have an online perceptual identification variant with an RT measure on each trial. A word is alternated with a pattern mask and progressively presented for a longer duration (see below for further details). Participants must indicate as quickly as possible when they think they have identified the word. This response is used as an index of word processing time. Participants next type in the word, to make sure they identified it correctly.

The PDM presentation conditions are thought to slow down word recognition, thereby making the task sensitive to factors affecting the early stages of visual word recognition. Carreiras et al. (1997) argued that the PDM task represents a purer measure of orthographic word processing than tasks like LD or naming, because (unlike LD) PDM requires the unambiguous identification of the word and it is not influenced by external factors such as the nature of non-word foils or articulatory factors. However, in a number of small-scale, factorial designs the PDM task has been shown to produce patterns similar to those of the more popular LD task (e.g., Grainger and Segui, 1990; Grainger and Jacobs, 1996; Carreiras et al., 1997; Schreuder and Baayen, 1997; Ferrand and Grainger, 2003). PDM is also used in memory research, where it is called the continuous identification (CID) task (e.g., Berry et al., 2008).

To our knowledge, there is only one word recognition megastudy using PDM. Lemhöfer et al. (2008) compared English processing times in native speakers and three groups of bilinguals with different mother tongues. The authors concluded that overall (and against their expectations) the similarities between word processing in first and second language were greater than the differences. As a matter of fact, the latter only emerged from detailed analyses of subsets of data.

The current study presents a database of word processing times for French monosyllabic, monomorphemic words in NMG, LD, and PDM, providing researchers with a new database in a new language. In this paper, we describe the collection of the data and the outcome of some basic analyses. By making the data already available, other researchers can start to run their own analyses rather than having to wait until we finalize our detailed ones.

MATERIALS AND METHODS
PARTICIPANTS
One-hundred and five university students (mean age = 20.6 years, SD = 1.94, min = 18 max = 27) from Blaise Pascal University, Clermont-Ferrand, France, participated in the experiments (35 in the LD task, 37 in the NMG task, and 33[2] in the PDM task). All were native speakers of French, and had normal or corrected-to-normal vision. Participants were paid 25€ for the two sessions in the naming task, or 50€ for the four sessions in the LD task and the PDM task. The difference of payment was due to the fact that the LD task and the PDM task were twice as long as the naming task. Each session lasted 1 h. Each participant participated either in the NMG task, the LD task, or the PDM task.

STIMULI
All French monosyllabic monomorphemic word forms were extracted from Lexique (New et al., 2004b, 2007). These are based on very large corpora of contemporary French texts and television subtitles. From this sample, we excluded the words we would never use as targets in word recognition experiments (such as loan words from English, names, sexually charged words, abbreviations), all polymorphemic words (in particular plurals and verb inflections other than the infinitive form), and all single letter words (as it is counterintuitive to define letters as non-words). This left us with a total sample of 1,826 words ranging from two to eight letters. For each word, printed frequency, subtitle frequency, number of letters, number of phonemes, the orthographic distance to the 20 nearest neighbors (OLD20)[3], and the phonological distance to the 20 nearest neighbors (PLD20) were taken from Lexique 3 (New et al., 2007). Length, frequency of occurrence, and similarity to other words are three fundamental word characteristics. We additionally had information about three other variables: age-of-acquisition (AoA), subjective frequency (taken from Ferrand et al., 2008), and imageability (published by Bonin et al., 2011).

For the purpose of the LD task, 1,826 length-matched pronounceable non-words were created by recombining the onsets and rimes of the words. We additionally made sure that the transition frequency distributions were very similar and that the recombination did not lead to less frequent bigrams or trigrams. The non-words were further matched on number of letters and number of orthographic neighbors (by deletion, substitution, addition, or transposition) using a nearest neighbor algorithm (implemented by the knn1 function provided by the class package of the R software)[4]. We did not follow Balota et al.'s (2004) procedure of changing a single letter in each word because this makes the non-words too similar to the words (Keuleers and Brysbaert, 2011).

---

[2]Six additional participants did not return for the fourth and last session; these are not included in our analyses.

[3]OLD20 (proposed by Yarkoni et al., 2008) is obtained by calculating the average number of operations (letter deletion, insertion, or substitution) needed to change a word into another word. For instance, the OLD from *smile* to *similes* is two (two insertions: I and S).

[4]The quality of the non-words was assessed *post hoc* with the newly developed LDNN1 algorithm developed by Keuleers (Keuleers and Brysbaert, 2011). This algorithm decides whether a new stimulus is a word or a non-word on the basis of the orthographic similarity with the previously presented words and non-words (i.e., without online lexical mediation). Whereas for all previously tested megastudies the algorithm was significantly better than chance, this was not true for the present study. Odds of correct selection on the basis of the algorithm were 1.1:1 (compared to 0.34:1 for the English Lexicon Project, 1.3:1 for the British Lexicon Project, and 1.6:1 for the French Lexicon Project). This means that the quality of our non-words was good.

All 1,826 words were presented in the LD and NMG tasks. However, because it became clear in the LD study that many words were not known to the participants, only the 1,482 words recognized by at least two thirds of the participants in the LD study were presented in the PDM task. The analyses are limited to this subsample. The main characteristics of the words are shown in **Table 1**.

APPARATUS
The LD and NMG experiments were run using DMDX (Forster and Forster, 2003), the PDM experiment using Dufau et al. (2008) software. All experiments were run on the same platform and had the same visual characteristics. Stimuli were presented on a 17″ Dell LCD monitor with a refresh rate of 70 Hz and a resolution of 1280 by 1024 pixels, placed at a distance of about 60 cm from the participants. The monitor was controlled by a PC Core Duo (Dell Precision 390). Stimuli were presented in lowercase Courier New font 12, and appeared on the screen as white characters on a dark background. For the LD task, participants responded on a Logitech Dual Action Gamepad. For the NMG task, we used a Plantronics microphone (Audio 350). Data were stored digitally and analyzed off-line with the CheckVocal software (Protopapas, 2007). In the PDM task, participants pressed the space bar whenever they recognized a word and typed it on the keyboard.

PROCEDURE
Lexical decision
Each person took part in four sessions of an hour on separate days within a period of 7 days. Participants were tested individually in a quiet room. They were seated in front of a computer and were told to indicate as rapidly and accurately as possible if the presented letter string was a correctly spelled French word or not. The participants responded using response buttons on a Logitech Dual Action Gamepad. They answered "yes" by pressing the button corresponding to the index of the preferred hand and "no" by pressing the button corresponding to the index of the non-preferred hand. The intertrial interval was 1.5 s.

Each trial consisted of the following sequence of events. First, two vertical fixation lines appeared slightly above and below the center of the screen, with a gap between them wide enough to clearly present a horizontal string of letters (word or non-word).

Table 1 | Characteristics of the words used in the experiments.

	Min	Max	Mean
N-letters	2	8	4.73
N-phonemes	1	6	3.28
Frequency books	0	38,929	313
Frequency subtitles	0.14	25,988	325
OLD20	1.00	2.85	1.50
PLD20	1.00	2.45	1.22
Age-of-acquisition (years)	2.82	15.45	7.78
Subjective frequency (1–7 scale)	2.32	7	4.25
Imageability (1–7 scale)	1.07	6.93	4.55

OLD20, orthographic Levenshtein distance; PLD20, phonological Levenshtein distance.

Participants were asked to fixate the gap as soon as the lines appeared. Five hundred milliseconds later the stimulus was presented in the gap with the center between the vertical lines, while the vertical lines remained on the stimulus. The stimulus and the vertical lines stayed on the screen until the participant made a response.

Overall, participants received 1,826 words and 1,826 non-words. Stimuli (words and non-words) were organized in four blocks of trials, 912 for the first and third session (456 words and 456 non-words), and 914 for the second and fourth session (457 words and 457 non-words). Blocks were counterbalanced across participants in a Latin square design. Trials within each block were randomly presented for each participant. Within a block, breaks occurred every 228 trials. Twenty practice trials preceded the experiment.

Naming
The naming task was in all aspects the same as the LD task, except that participants had to read aloud the words as fast and accurately as possible. After the computer detected the response, the stimulus word was erased from the screen. The next trial started after the participant pressed on the spacebar. If there was a pronunciation error or if an extraneous sound triggered the voice key, participants were asked to press the letter "b" on the keyboard instead of the spacebar (and the trial was coded as an error). The intertrial interval was 1,500 ms.

Overall, participants received 1,826 words. Stimuli were organized in two blocks of trials, 913 for the first session and 913 for the second session. Blocks were counterbalanced across participants. Trials within each block were randomly presented for each participant. Within a block, breaks occurred every 228 trials. Twenty practice trials preceded the experiment.

Progressive demasking
In the PDM task the display durations of the masking stimulus (a row of 11 hash marks ###########) and the target stimulus (a word) were manipulated. On each trial, a target–mask pair was presented consecutively several times until a response was made by the participant. During the trial, the total duration of the target and mask pair was held constant at 210 ms, but the ratio of the target display time to the mask display time increased progressively. At the beginning of a trial, the target was shown for one screen display cycle only (about 14 ms) and the mask for the rest of the time. As the trial continued, the target presentation time was increased by one cycle at a time (28, 42, 56,. . .ms) and that of the mask decreased (182, 168, 154,. . .ms). To the participant it looked as if the target gradually emerged from the mask. The target–mask cycles continued until the target was identified by the participant. A response was then made by pressing the spacebar of the keyboard. To check the correctness of the identification, participants also had to type in the word. Only the words known in the LD task were presented in the PDM task.

RESULTS
Because there is no point in analyzing words unknown to the participants, we followed Balota et al. (2004) and included only those

words that achieved at least 66% accuracy in the LD task[5]. This left us with 1,482 items (out of 1,826).

PERFORMANCE IN THE TASKS

Any response that was coded as an error in the LD task, the NMG task, or the PDM task was excluded from the response latency analyses. The percentage of removed errors was 9.2% for LD, 2.0% for NMG and 0.2% for PDM. In addition, response latencies were discarded when they were lower than three times the Q3–Q1 distance below Q1 (Q1 and Q3 being the first and the third quartile of the distribution of a participant) or higher than three times the Q3–Q1 distance above Q3 (see, e.g., Tukey, 1977, for such a procedure applied to outliers). The percentage of outliers was 0.2% in the LD task, 5.5% in the NMG task, and 1.8% in the PDM task.

Table 2 presents the mean RTs (and their SDs), together with the reliabilities of these RTs for the three tasks[6]. The latter are based on the split-half correlation corrected for length (e.g., Ferrand et al., 2010; Keuleers et al., 2010). They indicate how much variance in principle can be accounted for in the data. All three tasks were

---

[5]Two other stimuli (false and true) were lost because they had been translated automatically into Booleans during the stimulus handling.
[6]Unfortunately, because the data were collected a few years ago as a pilot study of the French Lexicon Project, before mixed effects analyses became common, the raw data were not archived and have been partly lost since.

comparable in reliability, certainly when one takes into account the differences in sample size (e.g., Rey and Courrieu, 2010). For comparison purposes, we also include the data from the French Lexicon Project (FLP; Ferrand et al., 2010) for the words included in the present study (there were 66 missing observations for this variable).

CORRELATIONS BETWEEN THE VARIABLES

Analyses indicated that word frequencies based on film subtitles slightly outperformed word frequencies based on books or the average frequency based on both sources (see also New et al., 2007; Brysbaert and New, 2009). Therefore, we limit our analysis to the film-based frequency measure. For word length, number of letters correlated more with task performance than number of phonemes. So, we took this variable. Table 3 shows the correlations between the three tasks and the various predictor variables.

Table 3 first shows that the correlation between the LD experiment of Chronolex and the data from the FLP ($r = 0.62$) is lower than expected. If a test is expected to correlate about 0.80 with itself (which is the definition of reliability), we expect it to correlate nearly as much with a very similar test (see Keuleers et al., 2011, for such a case). This suggests that differences between the tasks have induced slightly different response strategies. In FLP the monosyllabic, monomorphemic stimuli were only a small part of the words (less than 5%). Word length is a likely candidate to be

Table 2 | Mean reaction times and SD across stimuli for the three Chronolex tasks and the words present in FLP ($N = 1,482$ for Chronolex and 1,416 for FLP).

	LDT	Naming	PDM	LDT$_{FLP}$
M$_{RT}$	665	485	1194	658
SD$_{RTstim}$	64	38	82	61
Reliability	0.85	0.87	0.82	0.77*

Reliability of the FLP-data, as given by Ferrand et al. (2010). LDT, lexical decision task; PDM, progressive demasking; LDT$_{FLP}$, lexical decision from the French Lexicon Project. Last row gives the reliability of the data, based on the split-half correlation.

Table 3 | Correlations between the various measures (correlations lower than $r = -0.051$ or higher than $r = +0.051$ are significant at $p < 0.05$, $N = 1482$).

	NMG	PDM	LDT$_{FLP}$	Nlett	OLD20	PLD20	FREQ	AoA	SUBJ	IMA
LDT	0.343	0.560	0.622	0.338	0.159	0.204	−0.633	0.568	−0.566	−0.119
NMG		0.216	0.254	0.234	0.186	0.136	−0.232	0.230	−0.223	−0.004
PDM			0.434	0.375	0.122	0.163	−0.355	0.310	−0.299	−0.088
LDT$_{FLP}$				0.101	0.059	0.061	−0.384	0.412	−0.348	−0.188
Nlett					0.491	0.438	−0.349	0.199	−0.284	0.149
OLD20						0.615	−0.288	0.239	−0.217	0.087
PLD20							−0.284	0.278	−0.223	0.039
FREQ								−0.601	0.795	−0.255
AoA									−0.567	−0.324
SUBJ										−0.277

NMG, naming; PDM, progressive demasking; LDT, lexical decision task; LDT$_{FLP}$, lexical decision from the French Lexicon Project; Nlett, number of letters; OLD20, orthographic Levenshtein distance; PLD20, phonological Levenshtein distance; FREQ, frequency; AoA, age-of-acquisition; SUBJ, subjective frequency; IMA, imageability.

affected by this context and indeed the length effect is significantly smaller for FLP than for Chronolex (see also below for non-linear analyses).

Table 3 further shows that the PDM task correlates more with LD ($r = 0.56$) than with NMG ($r = 0.22$). OLD20 and PLD20 have positive correlations with word processing times, meaning that a word was processed more efficiently the more it resembled other words. This was also (surprisingly) the case for the PDM task.

IMPACT OF WORD LENGTH, SIMILARITY TO OTHER WORDS, AND WORD FREQUENCY

To further examine the impact of the predictor variables for the various tasks, we ran stepwise regression analyses. Regressions have the advantage that non-linear relations can be examined (by making use of cubic splines in the R statistical package). We used splines with three knots. First, we limited us to the three traditional variables in word recognition research: word length, word frequency, and similarity to other words[7]. For each variable we calculated the percentage of variance explained when it was entered first in the analysis and the extra variance it explained when it was entered last. The former value gives the upper limit of the impact; the later value gives the lower limit (Baayen et al., 2006). Table 4 shows the results. For the similarity measure we took the measure that explained most variance (OLD20 for naming, and PLD20 for LD and PDM).

Table 4 clearly shows the importance of word frequency for LD, PDM, and to a lesser extent NMG. Word length had a strong influence on PDM and contributed to the other tasks. The unique contribution of PLD20/OLD20 for the short words of Chronolex was limited in all tasks. Figure 1 illustrates the impact of each variable when the two other predictors are taken into account[8]. Because the differences with OLD20 only gave slightly better estimates for NMG, we limit the analyses to PLD20. Attentive readers may notice that the effect of PLD20 tended to be negative when the other two variables were taken into account (i.e., slower responses to words similar to many other words). This is different from what seemed to be the case in the raw correlations (Table 3).

IMPACT OF THE FIRST PHONEME

It is well established that the first phoneme is the most important variable to predict naming times for short words (Balota et al., 2004; Bonin et al., 2004; Cortese and Khanna, 2007). Usually, the impact of this variable is assessed by using a set of binary features to represent the first phoneme. An alternative is to enter the different phonemes as a non-continuous variable in the regression analyses. An advantage of this procedure is that you get estimates of the delays associated with the various phonemes; a disadvantage is that there may be some overfitting of the data (i.e., that the regression weights capture some noise in the data). Table 5 shows the effect of the introduction of the first phoneme into the regression for the different tasks. In line with the English findings, the first phoneme accounts for much variance in word naming, whereas the contribution to LD is limited and rather inconsistent between Chronolex and FLP. Surprisingly, the first phoneme (or letter) seems to have a stronger effect on PDM than on LD. The Section "Appendix" contains the estimates of the various phonemes in French for the different tasks.

AGE-OF-ACQUISITION, SUBJECTIVE FREQUENCY, AND IMAGEABILITY

To assess the impact of AoA, subjective frequency, and imageability, we entered these variables as well. Because of the limited contribution of PLD20/OLD20, we dropped it from Table 6. As can be seen in Table 6, the addition of each variable explained extra variance, except for imageability in the word naming task. For LD and PDM, the contribution of AoA and imageability was larger than that of subjective frequency. To make sure that the effect of subjective frequency was not due to overfitting the data, we entered log10 (frequency books) as an additional predictor to the regression. Subjective frequency remained significant after this variable had been added (which itself was not significant in the overall analysis)[9].

PROGRESSIVE DEMASKING VS. LEXICAL DECISION

Finally, to examine whether PDM captured some unique aspect of word processing that was not present in LD, we ran regressions with LD RT and the other variables as predictors. The data of these analyses can be found in Table 7. As can be seen, only length and, to a lesser degree, imageability had different effects in PDM and in LD. Both variables had a significantly stronger impact on PDM.

---

[7]Collinearity of the three predictors is 14.2, well within the limits of acceptability (Baayen, 2008, p. 182).

[8]For the sake of simplicity, coefficients are not presented in the text, since readers can run the analysis and check the coefficients themselves.

[9]This is another example where access to the raw data would have been interesting, because then it would have been possible to develop the model on part of the data and test it on the remainder. Such a cross-validation usually results in a smaller number of significant predictors. In particular the contribution of subjective frequency must be treated with caution as its inclusion leads to a collinearity index of 37 (against 26 without this variable).

Table 4 | Percentages variance explained (differences of 0.01 are significant at the $p < 0.05$ level).

	LDT	Naming	PDM	LDT$_{FLP}$
Nlett	0.11/0.03	0.05/0.03	0.17/0.12	0.02/0.02
PLD20/OLD20	0.04/0.01	0.03/0.00	0.03/0.01	0.01/0.01
FREQ	0.44/0.36	0.06/0.04	0.16/0.13	0.24/0.24

Lower and upper limits of percentages variance are shown. LDT, lexical decision task; PDM, progressive demasking; LDT$_{FLP}$, lexical decision from the French Lexicon Project; Nlett, number of letters; PLD20/OLD20, phonological/orthographic Levenshtein distance; FREQ, frequency.
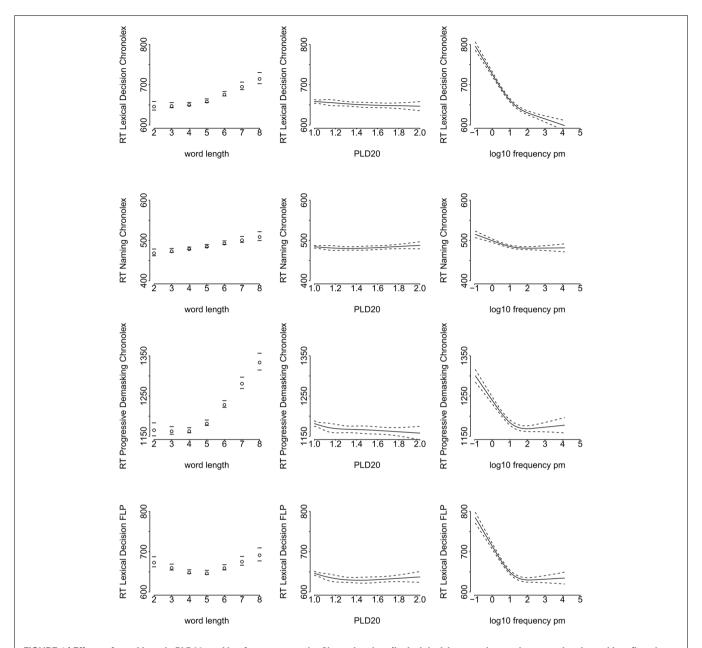
FIGURE 1 | Effects of word length, PLD20, and log frequency on the Chronolex data (lexical decision, naming, and progressive demasking; first three lines) and the lexical decision data of FLP (last line). (95% confidence intervals indicated by dashed lines).

Table 5 | Improvement in percentage of variance explained due to the inclusion of the first phoneme in a multiple regression model with word length (Nlett), PLD20, and word frequency (FREQ).

	LDT	Naming	PDM	LDT$_{FLP}$
Nlett + PLD20 + FREQ	0.48	0.10	0.29	0.27
Nlett + PLD20 + FREQ + first_phon	0.50*	0.67**	0.42**	0.29

*$p < 0.05$, **$p < 0.0001$ (determined with an ANOVA between the two models). Cubic splines used for Nlett, PLD20, and FREQ.

DISCUSSION

The aim of the present study was to investigate the influence of some psycholinguistic variables in different word recognition tasks (with monosyllabic monomorphemic words): NMG, LD, and PDM. This study extends the seminal study of Balota et al. (2004) by testing a new language (French) and adding a new task (PDM).

Table 6 | Percentage of variance explained by AoA, subjective frequency (SUBJ), and imageability (IMA) in addition to the variance explained by word length and word frequency.

	LDT	Naming	PDM	LDT$_{FLP}$
Nlett + FREQ	0.474	0.095	0.282	0.261
Nlett + FREQ + AoA	0.520**	0.106**	0.291**	0.298**
Nlett + FREQ + SUBJ	0.496**	0.102**	0.287**	0.277**
Nlett + FREQ + IMA	0.528**	0.096	0.296**	0.302**
Nlett + FREQ + AoA + SUBJ	0.531**	0.110*	0.294*	0.306**
Nlett + FREQ + AoA + IMA	0.536**	0.106	0.297**	0.309**
Nlett + FREQ + AoA + IMA + SUBJ	0.557**	0.110*	0.302**	0.323**

*p < 0.05, **p < 0.01 (determined with an ANOVA between the two models) in which model1 is Nlett + FREQ for AoA and SUBJ, and model1 = Nlett + FREQ + AoA for AoA + SUBJ and AoA + IMA, and model1 = Nlett + FREQ + AoA + IMA for the full model in the last row). Nlett, number of letters; FREQ, frequency; AoA, age-of-acquisition; SUBJ, subjective frequency; IMA, imageability. Cubic splines used for Nlett and FREQ.*

Table 7 | Variables other than lexical decision RT that predict PDM performance (linear regression without splines).

PDM = 715.5 + 0.72 LDT**	$R^2 = 0.314$
PDM = 703.7 + 0.63 LDT** + 15.3 Nlett**	$R^2 = 0.353$
PDM = 672.9 + 0.66 LDT** + 15.8 Nlett** + 4.0 FREQ	$R^2 = 0.354$
PDM = 708.8 + 0.63 LDT** + 16.3 Nlett** − 9.9 PLD20	$R^2 = 0.354$
PDM = 700.4 + 0.64 LDT** + 15.3 Nlett** − 0.5 AoA	$R^2 = 0.353$
PDM = 657.4 + 0.67 LDT** + 15.7 Nlett** + 4.4 SUBJ*	$R^2 = 0.355$
PDM = 724.8 + 0.61 LDT** + 16.3 Nlett** − 3.4 IMA**	$R^2 = 0.357$
PDM = 696.5 + 0.63 LDT** + 16.3 Nlett** − 2.8 IMA* + 2.4 SUBJ	$R^2 = 0.357$

**p < 0.001, *p < 0.05. PDM, progressive demasking; LDT, lexical decision task; Nlett, number of letters; FREQ, frequency; PLD20, phonological Levenshtein distance; AoA, age-of-acquisition; SUBJ, subjective frequency; IMA, imageability.

COMPARISON OF CHRONOLEX WITH PREVIOUS MEGASTUDIES CONDUCTED IN ENGLISH

In general, the findings of LD and NMG agree very well with those in English (Balota et al., 2004; Cortese and Khanna, 2007; Yap and Balota, 2009). Without question word frequency is the most important factor in LD (accounting for at least 36% unique variance in our study) and the first phoneme is the main variable in word naming (accounting for more than 50% of the variance in our study; see also Bonin et al., 2004, who reported that over 40% of variance was explained by the first phoneme in their regression analyses of French word naming latencies). In line with what Brysbaert and Cortese (2011) reported for English, the extra contribution of subjective frequency seems to be limited when a good objective frequency measure (based on subtitles) is used (**Table 6**). This is the more remarkable because in the present study the ratings (Ferrand et al., 2008) were obtained at the same university as the word processing times. One would have expected this to favor the subjective frequency measure. In contrast, the contribution of AoA ratings is larger (**Table 6**), again in line with the English findings (Brysbaert and Cortese, 2011). Imageability contributed to LD times and PDM times, but not to word naming times. This also is in line with the pattern reported for English (Balota et al., 2004).

COMPARISON OF CHRONOLEX WITH THE FRENCH LEXICON PROJECT

A comparison of the LD times of Chronolex and the FLP (first and last row of **Figure 1**) reveals that the effect of word length seems to depend on the context in which the words are presented. When all words and non-words are short (monosyllabic), as in Chronolex, the length function is J-shaped. In contrast, when the short words are presented among many long words, as in FLP, the length function becomes U-shaped. The latter was reported by New et al. (2006) and Ferrand et al. (2010) as well. The difference between Chronolex and FLP suggests that the effect of word length in the LD task is not entirely due to word processing itself, but partly depends on expectancies of what words are likely to be presented in an experiment. Given that optimal performance is for word lengths below the middle of the range, the geometric mean[10] of the word lengths rather than the arithmetic mean may better capture the expected length, in line with the observation that magnitudes in the human brain are stored logarithmically (Brysbaert, 1995; Dehaene et al., 1998). Participants may also take into account the probabilities of the various lengths.

A comparison of Chronolex and FLP further confirms the superiority of a design in which the participants process all stimuli (Chronolex) relative to a design in which participants only see a subset of the stimuli (FLP). Indeed, there was more variance accounted for in the LD task of Chronolex than in 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, 2011; see also Courrieu and Rey, 2011).

Similarity to other words, as measured by OLD20 and PLD20 had little influence on the processing times of the monosyllabic, monomorphemic words tested in Chronolex (**Table 4**; **Figure 1**). This deviates from the large impact OLD20 had on the RTs from the full FLP (Ferrand et al., 2010). A possible interpretation is that nearly all words had a high similarity to other words, so that the variability in OLD20 and PLD20 values was strongly reduced.

---

[10]The geometric mean of word length is the antilog of the mean log (word length). So, for words with lengths of 2–8, the geometric mean = antilog[(log(2) + log(3) + log(4) + log(5) + log(6) + log(7) + log(8))/7] = 4.5.

PROGRESSIVE DEMASKING VS. LEXICAL DECISION

Finally, the findings with the PDM task seem to be less exciting than we had hoped for. The results do not add much to those of the LD task and seem to be particularly influenced by perceptual factors (word length, first phoneme/letters). Note that the claim that PDM mainly taps into visual factors is not new (see e.g., Grainger and Segui, 1990; Schreuder and Baayen, 1997). It is interesting to see that our megastudy approach yielded the same results as other approaches. There is no evidence that response times in PDM are more influenced by unique word identification than response times in LD, as OLD20 and PLD20 did not have a stronger effect in PDM than in LD. On the other hand, there was a slightly stronger effect of word imageability in PDM, suggesting that the long reaction times in this task provide more scope for semantic influences. A similar observation was made by Dunabeitia et al. (2008), who reported that RTs in the PDM task were 145 ms shorter for words with many associates (as measured in a word association task) than for words with few associates. All in all, however, because of the substantial influence of perceptual variables, PDM does not look like a strong competitor of LD, except in cases where it is not clear how best to construct non-words (as in research on bilingual language processing, where non-words must be matched not only to the characteristics of target words that are presented but also to the different languages known to the participants; e.g., Lemhöfer et al., 2008).

LIMITATIONS OF THE PRESENT STUDY AND FUTURE DIRECTIONS

Although this study has examined the influence of a certain number of measures on visual word recognition performance, a number of questions remain open. A current limitation of the present study is that it did not include morphological variables. Some studies (e.g., Baayen et al., 1997; Schreuder and Baayen, 1997; New et al., 2004a) suggest that the speed with which monomorphemic words are recognized is determined not only by their own frequencies of use, but also by the frequencies of the inflectional variants of these words. In particular, Schreuder and Baayen (1997) showed that the frequency of the (unseen) plural forms affects the recognition latencies of monomorphemic singular words, such that singular nouns with high-frequency plurals are responded to faster than equally frequent singulars with low-frequency plurals (see also Baayen et al., 1997; New et al., 2004a). They also showed that a monomorphemic noun with a large family size (i.e., the number of different words in the family) elicited shorter LD times than a monomorphemic noun with a small family size. Because these morphological variables have been shown to play a prominent role in the visual word recognition of monomorphemic words, it will be important to include these variables in future analyses (see, e.g., Baayen et al., 2006, for such an endeavor).

It is also becoming clear that to take full advantage of megastudies, it is important to have access to the data at trial level rather than the average word data. Although the latter was customary in early megastudies (e.g., Seidenberg and Waters, 1989; Spieler and Balota, 1997), current statistical sophistication allows researchers to go beyond mean reaction times and accuracy levels, a trend which is likely to increase in the coming years. In this respect, it may be interesting to further include a feature of the Dutch and the British Lexicon Projects (Keuleers et al., 2010, 2011). Here the authors not only stored the participants' responses, but also the time and date of stimulus presentation and the computer on which the stimulus was presented. This makes it possible to use these variables in analyses as well.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at http://www.frontiersin.org/language_sciences/10.3389/fpsyg.2011.00306/abstract

The Chronolex data (mean RTs per word, PEs for LDT, and the predictor variables) are available in an Excel file as supplementary materials to this paper. They allow readers to repeat and extend our analyses.

REFERENCES

Andrews, S., and Heathcote, A. (2001). Distinguishing common and task-specific processes in word identification: a matter of some moment? *J. Exp. Psychol. Learn. Mem. Cogn.* 27, 514–544.

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R.* Cambridge: Cambridge University Press.

Baayen, R. H., Dijkstra, T., and Schreuder, R. (1997). Singular and plurals in Dutch: evidence for a parallel dual route model. *J. Mem. Lang.* 37, 94–117.

Baayen, R. H., Feldman, L. B., and Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *J. Mem. Lang.* 55, 290–313.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., and Yap, M. J. (2004). Visual word recognition for single syllable words. *J. Exp. Psychol. Gen.* 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., and Treiman, R. (2007). The English Lexicon Project. *Behav. Res. Methods* 39, 445–459.

Balota, D. A., Yap, M. J., Hutchison, K. A., and Cortese, M. J. (in press). "Megastudies: what do millions (or so) of trials tell us about lexical processing?" in *Visual Word Recognition: Models and Methods, Orthography and Phonology*, ed. J. S. Adelman (Hove: Psychology Press).

Berry, C. J., Shanks, D. R., and Henson, R. N. A. (2008). A single-system account of the relationship between priming, recognition, and fluency. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 97–111.

Bonin, P., Barry, C., Méot, A., and Chalard, M. (2004). The influence of age of acquisition in word reading and other tasks: a never ending story? *J. Mem. Lang.* 50, 456–476.

Bonin, P., Méot, A., Ferrand, L., and Roux, S. (2011). L'imageabilité: normes et relations avec d'autres variables psycholinguistiques [imageability: norms and relationships with other psycholinguistic variables]. *Ann. Psychol.* 111, 327–357.

Brysbaert, M. (1995). Arabic number reading – on the nature of the numerical scale and the origin of phonological recoding. *J. Exp. Psychol. Gen.* 124, 434–452.

Brysbaert, M., and Cortese, M. J. (2011). Do the effects of subjective frequency and age of acquisition survive better word frequency norms? *Q. J. Exp. Psychol.* 64, 545–559.

Brysbaert, M., and 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. *Behav. Res. Methods* 41, 977–990.

Carreiras, M., Perea, M., and Grainger, J. (1997). Orthographic neighborhood effects on visual word recognition in Spanish: cross-task comparisons. *J. Exp. Psychol. Learn. Mem. Cogn.* 23, 857–871.

Cortese, M. J., and 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. *Q. J. Exp. Psychol.* 60, 1072–1082.

Courrieu, P., and Rey, A. (2011). Missing data imputation and corrected statistics for large-scale behavioral databases. *Behav. Res. Methods* 43, 310–330.

Dehaene, S., Dehaene-Lambertz, G., and Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends Neurosci.* 21, 355–361.

Dufau, S., Stevens, M., and Grainger, J. (2008). Windows executable software for the progressive demasking task (PDM). *Behav. Res. Methods* 40, 33–37.

Dunabeitia, J. A., Aviles, A., and Carreiras, M. (2008). NoA's ark: influence of the number of associates in visual word recognition. *Psychon. Bull. Rev.* 15, 1072–1077.

Ferrand, L., Bonin, P., Méot, A., Augustinova, M., New, B., Pallier, C., and Brysbaert, M. (2008). Age-of-acquisition and subjective frequency estimates for all generally known monosyllabic French words and their relation with other psycholinguistic variables. *Behav. Res. Methods* 40, 1049–1054.

Ferrand, L., and Grainger, J. (2003). Homophone interference effects in visual word recognition. *Q. J. Exp. Psychol.* 56A, 403–419.

Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., Augustinova, M., and Pallier, C. (2010). The French Lexicon Project: lexical decision data for 38,840 French words and 38,840 pseudowords. *Behav. Res. Methods* 42, 488–496.

Feustel, T. C., Shiffrin, R. M., and Salasoo, A. (1983). Episodic and lexical contributions to the repetition effect in word identification. *J. Exp. Psychol. Gen.* 112, 309–346.

Forster, K. I., and Forster, J. C. (2003). DMDX: a windows display program with millisecond accuracy. *Behav. Res. Methods Instrum. Comput.* 35, 116–124.

Gardner, M. K., Rothkopf, E. Z., Lapan, R., and Lafferty, T. (1987). The word frequency effect in lexical decision: finding a frequency-based component. *Mem. Cognit.* 15, 24–28.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *J. Exp. Psychol. Gen.* 113, 256–281.

Grainger, J., and Jacobs, A. M. (1996). Orthographic processing in visual word recognition: a multiple read-out model. *Psychol. Rev.* 103, 518–565.

Grainger, J., and Segui, J. (1990). Neighborhood frequency effects in visual word recognition. *Percept. Psychophys.* 47, 191–198.

Keuleers, E., and Brysbaert, M. (2011). Detecting inherent bias in lexical decision experiments with the LD1NN algorithm. *Ment. Lex.* 6, 34–52.

Keuleers, E., Diependaele, K., and 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. *Front. Psychol.* 1:174. doi:10.3389/fpsyg.2010.00174

Keuleers, E., Lacey, P., Rastle, K., and Brysbaert, M. (2011). The British Lexicon Project: lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behav. Res. Methods.* PMID: 21720920. [Epub ahead of print].

Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, H. R., Grainger, J., and Zwitserlood, P. (2008). Native language influences on word recognition in a second language: a mega-study. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 12–31.

Monsell, S. (1991). "The nature and locus of word frequency effects in reading," in *Basic Processes in Reading. Visual Word Recognition*, eds D. Besner and G. W. Humphreys (Hillsdale, NJ: Lawrence Erlbaum), 148–197.

Monsell, S., Doyle, M. C., and Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: where are they? *J. Exp. Psychol. Gen.* 118, 43–71.

New, B., Brysbaert, M., Segui, J., Ferrand, L., and Rastle, K. (2004a). The processing of singular and plural nouns in French and English. *J. Mem. Lang.* 51, 568–585.

New, B., Pallier, C., Brysbaert, M., and Ferrand, L. (2004b). Lexique 2: a new French lexical database. *Behav. Res. Methods Instrum. Comput.* 36, 516–524.

New, B., Brysbaert, M., Veronis, J., and Pallier, C. (2007). The use of film subtitles to estimate words frequencies. *Appl. Psycholinguist.* 28, 661–677.

New, B., Ferrand, L., Pallier, C., and Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: new evidence from the English Lexicon Project. *Psychon. Bull. Rev.* 13, 45–52.

Protopapas, A. (2007). CheckVocal: a program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behav. Res. Methods* 39, 859–862.

Rey, A., and Courrieu, P. (2010). Accounting for item variance in large-scale databases. *Front. Psychol.* 1:200. doi:10.3389/fpsyg.2010.00200

Schreuder, R., and Baayen, R. H. (1997). How complex simplex words can be. *J. Mem. Lang.* 37, 118–139.

Seidenberg, M. S., and Waters, G. S. (1989). Word recognition and naming: a mega study. *Bull. Psychon. Soc.* 27, 489.

Spieler, D. H., and Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychol. Sci.* 8, 411–416.

Spieler, D. H., and Balota, D. A. (2000). Factors influencing word naming in younger and older adults. *Psychol. Aging* 15, 225–231.

Tukey, J. W. (1977). *Exploratory Data Analysis.* Reading, MA: Addison-Wesley Publishing Company.

Yap, M., and Balota, D. A. (2009). Visual word recognition of multisyllabic words. *J. Mem. Lang.* 60, 502–529.

Yarkoni, T., Balota, D., and Yap, M. (2008). Moving beyond Coltheart's N: a new measure of orthographic similarity. *Psychon. Bull. Rev.* 15, 971–979.

APPENDIX

Table A1 | Estimates of the processing times induced by different French phonemes in the different tasks (relative to the phoneme @ as in an).

Phoneme	Example	Chrono_LDT	Chrono_NMG	Chrono_PDM	FLP_LDT
§	on	−9	−13	−32	−43
1	un	15	57**	74	−49
2	eux	−35	−1	−61	−43
8	huile	−41	−17	44	−30
9	heure	−4	12	−9	−29
a	acte	−8	3	−15	−25
b	bac	−8	−47**	−38	−27
d	dans	−20	−52**	−24	−33
e	est	9	19	−55	−34
E	aide	17	5	14	−22
f	fable	−12	−73**	−34	−28
g	gag	−15	−37**	−51*	−21
i	il	80**	17	134**	27
j	hier	−13	23	−4	67
k	cadre	−16	−2	−52	−23
l	lac	−19	−47**	−27	−30
m	main	−14	−55**	22	−19
n	nage	−18	−57**	−16	−43*
o	au	−1	16	27	−9
O	homme	0	4	12	−26
p	page	−15	0	−77**	−30
R	rare	−11	−59**	−17	−42*
s	celle	−19	−69**	−62**	−24
S	chair	−38*	−69**	−63**	−38*
t	table	−13	2	−25	−20
u	ou	−27	−5	−13	−30
v	vache	−23	−53**	−73**	−32
w	oui	11	1	46	10
y	urne	−5	−19	120*	26
z	zone	−17	−43**	−82*	−22
Z	gel	−13	−38**	−76**	−46*

*p < 0.05, **p < 0.01.