Temporary ambiguity in whispered word recognition: a semantic priming study

Sophie Dufour, Yohann Meynadier

This is an Accepted Manuscript of an article published by Taylor & Francis in JOURNAL OF COGNITIVE PSYCHOLOGY on the 24 jan 2019, available online: http://www.tandfonline.com/10.1080/20445911.2019.1573243.

Temporary ambiguity in whispered word recognition: A semantic priming study

Sophie Dufour[1,2] Yohann Meynadier[1,2]

1 : Aix-Marseille Université, CNRS, LPL, UMR 7309, 13100 Aix-en-Provence, France

2 : Brain and Language Research Institute, Aix-Marseille Université, Aix-en-Provence, France

Sophie Dufour
Laboratoire Parole et Langage
Aix-Marseille Université, CNRS
5, avenue Pasteur
13604 Aix-en-Provence
France
Email: sophie.dufour@lpl-aix.fr

Abstract:

We examined whether listeners use acoustic correlates of voicing to resolve lexical ambiguities created by whispered speech in which a key feature, the voicing, is missing. Three associative priming experiments were conducted. The results showed a priming effect with whispered primes that included an intervocalic voiceless consonant (/petal/ “petal”) when the visual targets (FLEUR “flower”) were presented at the offset of the primes. A priming effect emerged with whispered primes that included a voiced intervocalic consonant (/pedal/ “pedal”) when the delay between the offset of the primes and the visual targets (VELO “bike”) was increased by 50 ms. In none of the experiments, the voiced primes (/pedal/) facilitated the processing of the targets (FLEUR) associated with the voiceless primes (/petal/). Our results suggest that the acoustic correlates of voicing are used by listeners to recover the intended words. Nonetheless, the retrieval of the voiced feature is not immediate during whispered word recognition.

Clearly, the way in which listeners deal with spoken language is remarkable since word recognition occurs effortlessly and with few errors under a wide variety of conditions that could be disruptive. Indeed, a given word is never spoken twice in exactly the same way and presents acoustic and phonetic differences according to the age, gender, regional accent, emotional state or speech rate of the talker. Each word can thus be materialised by an infinite number of different sound patterns that listeners have to associate with a unique lexical entry. Most models of spoken word recognition (Marslen-Wilson & Warren, 1994; McClelland & Elman, 1986; Norris, 1994; Norris & McQueen, 2008) posit that each word in the mental lexicon is associated with an abstract representation consisting of an idealised string of discrete symbols, for instance distinctive features or phonemes, that does not include fine-grained acoustic details about how these words are pronounced. In this view, the listener first converts the speech signal into a sequence of discrete segments, removing all acoustic details deemed irrelevant to identification, and then projects the output of this first stage onto the phonological representations of words stored in the mental lexicon. However, in contrast to these idealized phonological representations, the speech signal is often altered and words deviate from their so-called canonical forms. Within this abstractionist view of spoken word recognition, a challenge for researchers is to understand how listeners succeed in recognising the intended words from non-canonical forms. Opposed to this traditional class of models, other models (Godinger, 1998; Johnson, 1997) assume that the lexicon consists of multiple episodic traces including perceptual and contextual details associated to each individual occurrence of a word, thus encoding words in their non-canonical forms.

Up to now, a great deal of interest has been devoted to altered speech sounds resulting from phonological processes such as reduction, liaison or assimilation (e.g. Gow, 2001, 2002, 2003; Mitterer & McQueen, 2009; Snoeren, Segui & Halle, 2008a; 2008b; Spinelli, McQueen & Cutler, 2003). Many of these studies have in fact revealed that altered speech sounds do not disrupt word recognition, and that listeners are fully capable of recovering the intended words from their altered pronunciations. As an illustration, let us consider changes in word realisation that result from assimilation which refers to the process by which a segment becomes more like an adjacent segment in terms of shared phonetic features. In French, for example, the word-final /p/ in *jupe* /ʒyp/ "skirt" tends to become voiced ([b]) when followed by a voiced consonant as in *jupe grise* "grey dress". Using a cross-modal associative priming paradigm, Snoeren et al. (2008a) examined whether the assimilated forms of words presented in isolation activate their underlying lexical representations. Assimilated primes were either ambiguous, in that the altered form create another word (e.g. the assimilated form [sud] of the intended word *soute /sut/* "hold" could also be the word *soude /sud/* "soda"), or unambiguous, in that they had no voiced counterpart (e.g. *jupe* /ʒyp/ being the sole possible underlying word of the assimilated form [ʒyb]). Priming effects were found regardless of potential ambiguity. In particular, Snoeren et al. (2008a) observed that both the ambiguous [sud] and non-ambiguous [ʒyb] assimilated primes facilitated responses to their respective associated targets (BAGAGE "luggage" and ROBE "dress") as much as canonical primes [sut] and [ʒyp] did. Importantly, the priming effect observed with the ambiguous assimilated primes [sud] was not due to the overall phonological similarity of /sud/ with the underlying lexical representation of the word /sut/ (see Connine, Blasko & Titone, 1993), since no priming effect was reported in a subsequent experiment with the *soude* prime canonically pronounced [sud]. Hence, it appeared that listeners are able to recover the intended words from their assimilated forms, likely through the exploitation of subtle traces of the underlying voicing present in the speech signal. Indeed, assimilation is often incomplete (see Gow, 2002, 2003, for English), and regarding voice assimilation in French, an acoustic-phonetic description of assimilated segments in Snoeren et al. (2008a)'s study showed that acoustic properties that contribute to the voicing distinction, in particular the durations of stops and the preceding vowels, resist to assimilation, and thus may be perceptually relevant.

In this study, we focused on altered words that come from whispered speech. Whisper is a common mode of phonation used in natural communication, for example to reduce the level of perceptibility of speech in public situations for discretion or confidentiality purposes (Cirillo, 2004). Nevertheless, whispered speech is very little studied in the field of spoken word recognition, and yet it has two major characteristics that make it interesting for our understanding of how listeners cope with variability in how words are pronounced. The first characteristic is that when speakers whisper they produce speech through specific laryngeal configurations and mechanisms preventing vocal fold vibration (Fleischer, Kothe & Hess, 2007; Tsunoda, Niimi & Hirose, 1994), thus causing stronger changes in the speech input for voiced segments than for voiceless segments (Ito, Kateda & Itakura, 2005; Jovičić & Šarić, 2008). Since in French as in most languages (Lisker & Abramson, 1964), the presence of vocal fold vibration signal the presence of a voiced feature [+voice], and the absence of vocal fold vibration signal the presence of a voiceless feature [-voice], one of the main consequences of whisper is that [+voice] consonants tend to be pronounced more as [-voice] consonants. Thus in French, a word such as *pédale* /pedal/ “pedal” becomes potentially ambiguous, because its pronunciation approaches that of one of its lexical phonological neighbours, namely *pétale* /petal/ “petal”. The second characteristic is that, as it is the case for voice assimilation in French (Snoeren et al., 2008a), the voiced consonants in whispered speech maintain phonetic traces of their underlying identity. Whispered speech production studies have indeed revealed that vowels are longer when followed by a voiced obstruent than when followed by a voiceless obstruent, and voiced obstruents are shorter than voiceless obstruents (Jovičić & Šarić, 2008; Kohlberger & Strycharczuk, 2015; Meynadier & Gaydina, 2013; Parnell, Amerman & Wells, 1977; Sharf, 1964; Schwartz, 1972; Tartter, 1989; van de Velde & van Heuven, 2011), as typically observed in modal speech across languages (Chen, 1970; Flege, Munro & Skelton, 1992; Solé & Ohala, 2010) including French (Abdeli-Beruh, 2004; Mack, 1982; O’Shaughnessy, 1981).

The few studies that have examined how listeners perceive voicing either in whispered words, non-words or syllables have reported that the perception of voicing is rather well maintained - although less accurate than in modal words, non-words or syllables - with better than chance identification performance (Dannenbring, 1980; Fux, 2012; Gilichinskaya & Strange, 2011; Munro, 1990; Tartter, 1989). However, a closer look at the results reported in some of these studies reveals a perceptive bias in favour of [-voice] consonants. For example, in studies conducted in English by Tartter (1989) and Gilinchskaya and Strange (2011), the percentage of correct identification of voiceless obstruents reached 81% and 87%, respectively, and that of voiced obstruents reached 68% and 80%. In unpublished French study, Fux (2012) reported a percentage of correct identification comprised between 80 and 87% for voiceless obstruents, and between 16% and 47% for voiced consonants. This perceptual bias found in favour of [-voice] consonants in Fux (2012)’s study, with performance that did not reach the chance level for the [+voice] consonants, could suggest that French listeners fail to recover whispered words such as *pédale* /pedal/ “pedal” that include a [+voice] intervocalic consonant. As a result, these words could be confused with a phonological neighbour such as *pétale* /petal/ “petal” that include a [-voice] intervocalic consonant. However, all the above-mentioned studies used offline identification tasks, without time pressure, the more often forced-choice tasks between two or more alternatives, which make them rather limited in their ability to reflect automatic real-time processes, not contaminated by conscious decisional processes. This study was thus designed to examine whether listeners used the residual correlates of voicing to recover the intended words using a more online measure of spoken word recognition.

We used the cross-modal (auditory-visual) semantic/associative priming paradigm which is a well-established method for probing the amount of lexical activation caused by potentially ambiguous sequences (Connine et al., 1993; Gow and Gordon, 1995; Marslen-Wilson, Moss & van Halen, 1996; Snoeren et al., 2008a; Tabossi, 1996). The logic behind semantic priming is that hearing a prime leads to activation in the mental lexicon of its underlying phonological representation, which spreads on words that share semantic properties. If one of these words is then presented as target word, its processing is facilitated due to preactivation. Whispered words that include a [+voice] intervocalic consonant (henceforth, "voiced words"; e.g. /pedal/ "pedal") and whispered words that include a [-voice] intervocalic consonant (henceforth, "voiceless words"; e.g. /petal/ "petal") were used as primes. Our hypotheses were as follows. Because whispered voiceless words (/petal/) fully match their underlying representations for both the primary cue of voicing (i.e., the absence of vocal fold vibration) and the secondary durational cues of voicing, we expected a priming effect on their associated target words (FLEUR "flower"), but not on the target words (VELO "bike") related to the voiced lexical neighbours (/pedal/). Two scenarios have been envisaged for whispered voiced words (/pedal/). First, if the durational cues of voicing in whispered speech are used by listeners during lexical access, and if there are sufficient to compensate the loss of information caused by the absence of vocal fold vibration, we can expect a priming effect on their associated target words (VELO), but not on the target words (FLEUR) related to the voiceless lexical neighbours (/petal/). Indeed, due to the use of the remaining traces of voicing in the speech signal, the whispered prime /pedal/ should lead to a strong activation of the word form /pedal/ in the mental lexicon, thus facilitating the subsequent recognition of the associated target word VELO. Second and by contrast, if the durational cues signaling the presence of voicing in whispered speech are not sufficient for the retrieval of the voicing feature, then the voiced word prime /pedal/ should be treated as the voiceless word prime /petal/. In this case, we can expect a priming effect for the incongruent prime-target pairs (/pedal/-FLEUR), but not for the congruent prime-target pairs (/pedal/-VELO).

Experiment 1

In Experiment 1, the visual targets were presented at the offset of the primes (Inter-Stimulus Interval (ISI) = 0 ms), as it is often the case in studies using the cross-modal semantic priming paradigm (Connine et al., 1993; Marslen-Wilson et al., 1996; Snoeren et al., 2008a). Before to conduct our main experiment in whispered mode, we ran a control experiment with the primes presented in modal phonation to ensure that we have selected the right associated prime-target pairs, and thus that both the voiceless words (e.g. /petal/) and the voiced words (e.g. /pedal/) prime their respective associated target words namely FLEUR and VELO respectively. As a consequence, an eventual failure to observe semantic priming effects in the main experiment with whispered primes cannot be attributed to a problem in the selection of our materials. This control experiment thus consisted in a validation of our prime-target pairs in which each type of words (voiced, voiceless) was tested with their respective associated target words, namely FLEUR for the voiceless prime words (/petal/) and VELO for the voiced prime words (/pedal/).

In the main experiment each type of prime (voiceless, voiced) presented in whispered mode was tested in both a congruent and an incongruent condition. Hence, the voiceless words (/petal/) were tested with their associated targets words (FLEUR), and with the associated targets words (VELO) which correspond to the non-intended lexical neighbours (/pedal/). Also, the voiced words (/pedal/) were tested with their associated targets words (VELO), and with the associated targets words (FLEUR) which correspond to the non-intended lexical neighbours (/petal/).

Method

Materials

Eighty-six minimal pairs of words opposing in an intervocalic position a [+voice] consonant and its [-voice] counterpart were first selected (e.g. *pédale* /pedal/ – *pétale* /petal/). In order to select associated target words, the 172 words were then presented to 38 participants who were instructed to write down the first word that came to their mind upon written presentation of each of the words. Only the minimal pairs for which the two members were associated to a target word with a percentage of agreement equal to or greater than 20% were kept for the experiments. Adopting this criterion, our final set of words contained (i) 20 minimal pairs of words (e.g. /pedal/-/petal/), whose each member served as related primes, (ii) 40 target words, half of them being associated with the voiced member (/pedal/-VELO), and the other half being associated with the voiceless member (/petal/-FLEUR), (iii) 20 other words having no relation with the targets were selected and were used as control primes (e.g. *quittance* /kitɑ̃s/ "receipt"-VELO for the targets used in the voiced condition; and /kitɑ̃s/-FLEUR for the targets used in the voiceless condition). Related and control primes were matched for frequency, number of syllables, number of phonemes and uniqueness point. The characteristics of the prime and target words are summarised in Table 1, and the prime-target pairs are given in the Appendix. As indicated in Table 1, the mean frequency of target words (VELO) associated with the voiced primes (/pedal/) was 44 occurrences per million and the mean frequency of target words (FLEUR) associated with the voiceless primes (/petal/) was 106 occurrences per million. This difference was not significant ($t(38) = 1.18$, $p>.20$). More precisely, the frequency ranges of the target words (VELO) associated with the voiced primes (/pedal/) were 1-290; and those of the target words (FLEUR) associated with the voiceless primes (/petal/) were 1-283 with an "outlier" that has a frequency of 992 occurrences per million, and which is responsible for the descriptive difference in frequency between the two sets of target words.

< Insert Table 1 about here>

Acoustic recordings

All the primes were recorded in both modal (control experiment) and whispered (main experiment) phonation by a male native speaker of French in a soundproof room using a Senheiser HD 415 microphone, and digitised to 25-kHz sampling rate and 32-bit resolution. The 30-cm distance between the mouth of the speaker and the microphone was kept constant during the recording. The speaker was instructed to speak and to whisper in a natural manner. Glottal pulses and F0 were first checked with Praat (Boersma and Weenink, 2017) for their presence for the [+voice] segments and for their absence for the [-voice] segments in the modal mode, and importantly for their absence for both the [+voice] and [-voice] segments in the whispered mode. Figure 1 illustrates the presence of voicing for the prime /pedal/ in the modal mode, and the absence of voicing for the prime /pedal/ in the whispered mode as well as for the prime /petal/ in both the modal and the whispered modes. We also measured the durations of both the intervocalic obstruent (e.g. /d/ in /pedal/ and /t/ in /petal/) and the pre-consonantal vowel (e.g. /e/ in /pedal/ and /petal/) known to contribute to the voicing distinction. In accordance with previous studies (Gilichinskaya and Strange, 2011; Jovičić and Šarić, 2008; Tartter, 1989), we expected shorter durations for voiced obstruents than for voiceless obstruents, and longer pre-consonantal vowel durations for voiced obstruents than for voiceless obstruents. The durational values are provided in Table 2.

<Insert Figure 1 and Table 2 about here>

Modal phonation: The voiced obstruents were in average 62 ms shorter than the voiceless obstruents ($t(38) = 7.63$, $p<.0001$), and vowels were in average 36 ms longer before voiced obstruents than before voiceless obstruents ($t(38) = 3.18$, $p<.01$).

Whispered phonation: Crucially for our purpose, the voiced obstruents were in average 60 ms shorter than the voiceless obstruents ($t(38) = 7.06$, $p<.0001$), and vowels were in average 38 ms longer before voiced obstruents than before voiceless obstruents ($t(38) = 4.20$, $p<.001$). Hence, durational correlates of voicing are preserved in our whispered words, that listeners could exploit to recover the intended words.

Design

For each type of words (voiceless, voiced) two initial lists were constructed. As illustrated in the Appendix, in one list, the primes were combined with their associated target words (e.g. /pedal/-VELO for the voiced words and /petal/-FLEUR for the voiceless words). In the other list, the primes were combined with the associated target words of the other member of the minimal pairs (e.g. /pedal/-FLEUR for the voiced words and /petal/-VELO for the voiceless words). Only the congruent prime-target pairs (e.g. /pedal/-VELO for the voiced words, and /petal/-FLEUR for the voiceless words) were tested in the control experiment (in modal phonation). Both the congruent (e.g. /pedal/-VELO for the voiced words and /petal/-FLEUR for the voiceless words) and incongruent (e.g. /pedal/-FLEUR for the voiced words and /petal/-VELO for the voiceless words) prime-target pairs were tested in the main experiment (in whispered phonation). So that each target word was preceded by both the related and control primes and no participant received the same target twice, each list were then divided in two counterbalanced sub-lists. Each sub-list was thus composed of 20 target words. Half of the 20 target words were preceded by a related prime (e.g. /pedal/-VELO or /petal/-FLEUR in the congruent conditions; /pedal/-FLEUR or /petal/-VELO in the incongruent conditions), and the other half by a control prime (/kitɑ̃s/-VELO or /kitɑ̃s/-FLEUR in both the congruent and incongruent conditions). To achieve a proportion of related prime-target pairs of 20%, 30 prime-target word pairs without any relation (e.g. *nuage* /nyaʒ/ “cloud” - PION “pawn”) were added to each sub-list. For the purpose of the lexical decision task, 50 target non-words were also added to each sub-list. The target non-words were created by changing one letter of words (VIEL from CIEL “sky”) not previously used. Among the 50 target non-words, 10 were preceded by a “related” prime that was associated with the word from which the non-word was derived (e.g. the prime *tomate* /tomat/ “tomato” for the non-word target ROUZE derived from *rouge* “red”). The associated target words of these “pseudo” related non-words were taken from Ferrand and Alario’s (1998) association tables. As for the test items, the association strength was greater than 20% with an average of 31.32%. The 40 other target non-words were preceded by an unrelated prime that had no relation with the word from which the non-word was derived (e.g. the prime *visage* /vizaʒ/ “face” for the non-word PLAME derived from *plume* “feather”).

The participants were tested on one experimental sub-list only. Note that because initially we did not have a clear idea of how whispered voiced words (/pedal/) will be perceived, we preferred to test different participants for each type of words (voiced, voiceless) rather than to use a within-participants design for the factor word type. This was done to avoid repetitions in the primes in the case the voiced words (/pedal/) are perceived as the voiceless words (/petal/). Also for each type of words (voiced, voiceless), different participants were tested in the congruent and incongruent conditions, again to avoid repetitions in the prime words (e.g. /pedal/-VELO and /pedal/-FLEUR for the voiced word conditions). Of course, we are aware that a between-participants design is statistically less powerful than a within-participants design. Note also that we were especially interested in the priming effect (i.e., the difference between control and related prime) triggered by voiceless words in the one hand and by voiced words in the other hand in both the congruent and the incongruent conditions. In each of these conditions, the factor prime type (related, control) was a within-participants factor with participants tested on both control and related primes.

Procedure

The participants were tested individually in a quiet room. The modal (control experiment) and the whispered (main experiment) primes were presented over headphones at a comfortable and audible sound level, and remained constant across participants. At the offset of the primes, the targets were displayed in lower case letters (white 80-point Arial font) in the middle of a black screen until the participant's response. Participants were asked to decide as quickly as possible whether the target was a word or a non-word. "Word" responses were made with the dominant hand. The participant's response and the onset of the prime of the following trial were separated by a two second silence. Each participant began the experiment with a block of 10 practice trials.

Participants

A total of two hundred and four participants took part in Experiment 1 and were paid 10 euros for their participation. All were native speakers of French self-reported no hearing or speech disorders, and had normal or corrected to normal vision. Sixty-eight of them were assigned to the control experiment (modal primes) and were spread across the two types of prime words (voiced, voiceless). 34 were thus tested on voiceless words, and the other 34 were tested on voiced words. In each condition the 34 participants were spread across the two sub-lists that counterbalance the related and control primes.

The remaining 136 participants were assigned to the main experiment (whispered primes), and were spread across the two types of prime words (voiced, voiceless) and across the two types of semantic matching (congruent, incongruent). The repartition of the participants was thus as follow. 68 participants were tested on voiceless words, 34 with the congruent target words and the remaining 34 with the incongruent target words. The 68 other participants were tested on voiced words, 34 with the congruent target words and 34 with the incongruent target words. Again, in each condition the 34 participants were spread across the two sub-lists that counterbalance the related and control primes.[1]

Results and discussion

The data of the two experiments were analysed by means of mixed-effects modelling (Baayen, 2008; Baayen, Davidson and Bates, 2008) taking into account in the same model both participant and item variability, using the statistical software R (R Development Core Team, 2016) and the package *lme4* (Bates and Sarkar 2007). In both experiments, reaction times (RTs) analyses were performed on correct responses. RTs longer than 1300 ms were considered as outliers and removed from the RTs analyses. So that the model met the assumptions of normally-distributed residuals and homogeneity of variance, a log transformation was then applied on RTs (Baayen and Milin, 2010), and residuals that were 2.5 standard deviations away from the mean for each participant were also removed. Adopting these criteria, less than 5% of the data were rejected in each of the following models. Percentage of correct responses was also analysed using mixed-effects logit model (Jaeger, 2008). In both the control and the main experiment no significant accuracy effect was found. So we do not discuss them further.

Control experiment (modal primes): To assess the semantic priming effect caused by the voiceless words on the one hand, and by the voiced words on the other hand, we fitted linear mixed-effects model for each word type (voiceless, voiced) with the variable prime type (related, control) entered as fixed effect. The model also included participants and items as random intercepts, plus random participant and item slopes for prime type (see Barr, Levy, Scheepers, & Tily, 2013). Figure 2 shows the mean RTs as a function of prime type (related, control) for the voiceless and voiced words.

<Insert Figure 2 about here>

The results for the *voiceless words* revealed a significant effect of prime type (t= 2.34, p<.05) with target words responded to 26 ms more quickly when preceded by related primes in comparison to control primes.

The results for the *voiced words* also revealed a significant effect of prime type (t=2.64, p<.05) with target words responded to 26 ms more quickly when preceded by related primes in comparison to control primes.

Together the results of this control experiment revealed that both the voiced words (/pedal/) and the voiceless (/petal/) words prime well their respective associated target words since a significant priming effect was observed with the two types of words. Let us now examine how the two types of words (voiceless, voiced) behave when presented in whispered mode.

Main experiment (whispered primes): For each type of words (voiced, voiceless), we fitted linear mixed-effects model with the variables prime type (related, control), matching (congruent, incongruent) and their interaction entered as fixed effects[2]. The model also included participants and items as random intercepts, plus random participant and item slopes for prime type (see Barr et al., 2013). The p values when reported were obtained using the package *lmertest* (Kuznetsova et al., 2016), and pairwise comparisons assessing the priming effects were done using the *glht* function from the *multcomp* package (Bretz, Hothorn and Westfall, 2011) with a Bonferonni correction. Figure 3 shows the mean RTs as a function of prime type (related, control) and matching (congruent, incongruent) for the voiceless and voiced words separately.

<Insert Figure 3 about here>

The results for the *voiceless words* revealed a significant effect of prime type (t=2.32, p<.05) with target words responded to more quickly when preceded by related primes in comparison to control primes. The effect of matching was not significant (t=0.72, p>.20). Crucially, there was a significant prime type × matching interaction (t=2.04, p<.05) due to the fact that the priming effect was in the opposite direction in the congruent and incongruent condition. Subsequent pairwise comparisons revealed a significant priming effect for the congruent prime-target pairs (z=3.09, p<.05) only, with target words responded to 30 ms more quickly when preceded by related primes in comparison to control primes. No priming effect was observed for the incongruent prime target pairs (z= -0.68, p>.20). Hence as expected, whispered words that include a [-voice] intervocalic consonant (/petal/) facilitated the subsequent recognition of their associated target words (FLEUR), but not that of the associated target words (VELO) corresponding to the lexical neighbours (/pedal/).

The results for the *voiced words* revealed no significant effect (prime type, t=0.86, p>20; matching, t=1.45, p=.15; prime type × matching, t=1.12, p>.20), and subsequent pairwise comparisons assessing the priming effect within each matching condition revealed that both the congruent (z=1.16, p>.20) and the incongruent prime-target pairs (z= -0.91, p>.20) failed to show a significant priming effect. Hence, whispered words that include a [+voice] intervocalic consonant (/pedal/) caused no significant priming effect either on their associated target words (VELO) or on the associated target words (FLEUR) corresponding to the lexical neighbours (/petal/).

To summarise, the results observed in the main experiment with whispered prime words revealed, as expected, that the voiceless words (/petal/) led to a significant priming effect only when they were paired with their associated target words (FLEUR), but not when they were paired with incongruent target words (VELO). This priming effect was of similar magnitude (around 30 ms) to that observed with modal primes in the control experiment. Note that the differential priming effect observed between the whispered voiceless (/petal/) and voiced (*/pedal/*) prime words in the congruent conditions cannot be attributed to the fact that the two set of target words differ descriptively (not statically) in frequency, since the same amount of priming was observed in the control experiment when the two types of primes were presented in modal phonation[3]. Our results thus revealed no ambiguity in the processing of whispered words containing a [-voice] intervocalic consonant. However, no priming effect was found for whispered words that include a [+voice] intervocalic consonant (*/pedal/*), either on their associated target words (VELO), or on the incongruent target words (FLEUR). Such an observation thus suggests that the durational cues of voicing are used by listeners in that they prevent the activation of the inappropriate word (/petal/). However, the lack of a priming effect for the voiced words /pedal/ on their associated target words (VELO) also suggests that at 0 ms of ISI, the durational cues of voicing present in whispered speech are not fully exploited by listeners to compensate the loss of information due the absence of vocal fold vibration, and to guarantee an immediate recognition of these words. In Experiment 2, a delay of 50 ms was introduced between the offset of the primes and the visual targets. We expect that this delay would let the time for the intended voiced words to be fully activated when the target words are presented.

Experiment 2

In this experiment, we re-ran Experiment 1 with an ISI of 50 ms rather than 0 ms. Note that the semantic priming effect is known to be short-lived (Andruski et al., 1994; Marslen-Wilson et al., 1996, for the intra-modal semantic priming), and so with a longer ISI, we take the risk of seeing the effect decreases when the primes are unambiguous, namely with whispered voiceless primes (/petal/) and with the two types of primes (voiced, voiceless) of the control experiment in modal phonation. For example, although it is difficult to compare the timings of the events in intra-modal (auditory primes and targets) and cross-modal (auditory primes and visual targets) situations (see Marslen-Wilson et al., 1996 for a discussion), Andruski et al.

(1994) in an intra-modal situation reported a clear semantic priming effect at 50 ms ISI that vanished at 250 ms, thus indicating that the size of the semantic priming effect decreases as the ISI increases. Crucially however, 50 ms delay between the offset of the auditory primes and the onset of the visual targets could also allow the whispered voiced words - in which a key feature, the voicing, is missing - to fully activate their underlying lexical representations, so that a semantic priming effect could emerge. This was examined in this second experiment. Note that the 50-ms ISI is the interval between the primes and the targets which is the most frequently used in studies using the phonological priming paradigm (i.e. primes and targets overlapping in some of their phonemes; Dufour (2008) for a review), and thus it was chosen in reference to these studies.

Method

Materials and procedure

They were the same as in Experiment 1, except that a 50-ms interval separated the offset of the primes and the display of the targets.

Participants

A total of two-hundred and four participants took part in the experiments and were paid 10 euros for their participation. All were native speakers of French, self-reported no hearing or speech disorders, and had normal or corrected to normal vision. Sixty-eight of them were assigned to the control experiment (modal primes), and 136 to the main experiment (whispered primes). None had taken part in Experiment 1. The repartition of the participants was the same as in Experiment 1.

Results and discussion

The data were analysed according to the same criteria as in Experiment 1. Less than 5% of the data were rejected in each of the following models. Percentage of correct responses was also analysed using mixed-effects logit model (Jaeger, 2008). In both the control and the main experiment no significant accuracy effect was found. So we do not discuss them further.

Control experiment (modal primes): To assess the semantic priming effect caused by the voiceless words on the one hand, and by the voiced words on the other hand, we fitted linear mixed-effects model for each word type (voiceless, voiced) with the variable prime type (related, control) entered as fixed effect. The model also included participants and items as random intercept, plus random participant and item slopes for prime type (see Barr et al., 2013). Figure 4 shows the mean RTs as a function of prime type (related, control) for the voiceless and voiced words. Contrary to what was observed in Experiment 1, no significant priming effect was observed either for the *voiceless words* (t=1.71, p=.10) or for the *voiced words* (t=1.78; p=.10). Hence, increasing the ISI from 0 to 50 ms caused the dissipation of the semantic priming effects when the words were presented in modal phonation. Let us now examine how the two types of words (voiceless, voiced) behave when presented in whispered mode with 50-ms ISI.

<Insert Figure 4 about here>

Main experiment (whispered primes): For each type of word (voiced vs. voiceless), we fitted linear mixed-effects model with the variables prime type (related, control), matching (congruent, incongruent) and their interaction entered as fixed effects[2]. The model also included participants and items as random intercepts, plus random participant and item slopes for prime type (see Barr et al., 2013). The p values when reported were obtained using the package *lmertest* (Kuznetsova et al., 2016), and pairwise comparisons assessing the priming effects were done using the *glht* function from the *multcomp* package (Bretz et al., 2011) with a Bonferonni correction. Figure 5 shows the mean RTs as a function of prime type (related, control) and matching (congruent, incongruent) for the voiceless and voiced words separately.

<Insert Figure 5 about here>

The results for the *voiceless words* revealed no significant effect (prime type, t=1.65, p=.11; matching, t=0.40, p>.20; prime type × matching, t=0.66, p>.20), and subsequent pairwise comparisons assessing the priming effects within each matching condition revealed that both the congruent (z=2.06, p>.20) and the incongruent (z=0.92, P>.20) prime-target pairs failed to show a significant priming effect. Hence, the priming effect found in Experiment 1 with whispered words that include a [-voice] intervocalic consonant (/petal/) was no longer observed with 50-ms ISI.

The results for the *voiced words* revealed a significant effect of prime type (t=2.73, p<.01) with target words responded to more quickly when preceded by related primes in comparison to control primes. The effect of matching was not significant (t=0.72, p>.20). Crucially, there was a significant prime type × matching interaction (t=2.56, p<.05) due to the fact that the priming effect was in the opposite direction in the congruent and incongruent condition. Subsequent pairwise comparisons revealed a significant priming effect for the congruent prime-target pairs (z=3.03, p<.05) only, with target words responded to 29 ms more quickly when preceded by related primes in comparison to control primes. No priming effect was observed for the incongruent prime target pairs (z= -1.01, p>.20). Hence, with 50-ms ISI, whispered words that include a [+voice] intervocalic consonant (/pedal/) succeeded in priming their associated target words (VELO), so that a facilitation effect was now observed.

To summarise, the semantic priming effect found in Experiment 1 with voiceless words in the congruent condition (/petal/-FLEUR) was no longer observed when the ISI was increased from 0 to 50 ms, and critically a semantic priming effect with voiced words in the congruent condition (/pedal/-VELO) was now found.

Experiment 3

Together the results of Experiments 1 and 2 clearly show that voiced words are in no case confounded with their voiceless counterparts. This observation thus suggests that listeners used residual correlates of voicing to activate the intended words, and more crucially to prevent the activation of inappropriate candidates. Nonetheless, our results also showed that the retrieval of the voiced feature is not immediate during whispered word recognition and requires a certain amount of time to be done. Voiceless words showed a semantic priming effect with target words presented at the offset of the primes. However, a longer delay of 50 ms between the offset of the whispered primes and the visual target words was necessary such that a semantic priming effect occurs with voiced words. From a methodological point of view, our results confirm that the semantic priming effect, when it occurs, is rather short-lived (see Marslen-Wilson et al., 1996). The lack of priming effects at 50-ms ISI with voiceless words likely results from the fact that the residual activation caused by the prime in the lexical/semantic space of the target had had the time to dissipate before target presentation (Marslen-Wilson et al., 1996; see also Goldinger, Luce & Pisoni, 1989, for the same reasoning with phonetic priming effet). In Experiment 3, we again increased the ISI by 50 ms to examine whether the semantic priming effect observed with voiced words also vanishes when we increase the ISI from 50 to 100 ms. Note that because in the two previous experiments, no priming effect was reported in the incongruent conditions, only the congruent conditions (/pedal/-VELO; /petal/-FLEUR) were tested in Experiment 3.

Method

Materials and procedure

Only the whispered primes in the congruent conditions were tested. A 100 ms interval separated the offset of the primes and the display of the targets.

Participants

Sixty-eight participants took part in the experiments and were paid 10 euros for their participation. All were native speakers of French, self-reported no hearing or speech disorders, and had normal or corrected to normal vision. 34 were tested on voiceless words, and the other 34 were tested on voiced words. In each condition, the 34 participants were spread across the two sub-lists that counterbalance the related and control primes.

Results and discussion

The data were analysed according to the same criteria as in the previous experiments. Less than 5% of the data were rejected in each of the following models. Percentage of correct responses was also analysed using mixed-effects logit model (Jaeger, 2008). No significant accuracy effect was found. So we do not discuss them further.

To assess the semantic priming effect caused by the whispered voiceless words on the one hand, and by the whispered voiced words on the other hand, we fitted linear mixed-effects model for each word type (voiceless, voiced) with the variable prime type (related, control) entered as fixed effect. The model also included participants and items as random intercept, plus random participant and item slopes for prime type (see Barr et al., 2013). Figure 6 shows the mean RTs as a function of prime type (related, control) for the voiceless and voiced words. No significant priming effect was observed either for the *voiceless words* ($t=0.38$, $p>.20$) or for the *voiced words* ($t=0.60$, $p>.20$). Hence, increasing the ISI by 50 ms also caused the dissipation of the semantic priming effect for voiced words. We illustrated the time-course of the semantic priming effect for both the voiced and voiceless words in Figure 7.

<Insert Figures 6 and 7 about here>

General discussion

Although very little studied, whisper is one mode of natural communication that could bring information on how listeners cope with variability present in the speech signal. When speakers whisper, they send to their interlocutor a message which is altered, mainly because of the absence of vocal fold vibration, which removes the most obvious signature of the [+voice] feature from the speech signal. Nonetheless, just as in the case of voice assimilation (Gow, 2001, 2002, 2003; Snoeren et al., 2008a), in whispered speech, residual correlates of voicing, and in particular durational correlates, are present in the speech signal that could be exploited by listeners to recover the intended words in case of lexical ambiguity. In Experiment 1 with visual target words presented at the offset of the auditory primes, a priming effect was found only with whispered primes that included an intervocalic voiceless consonant (*/petal/*-FLEUR). Whispered words with an underlying voiced obstruent, such as /pedal/, did not prime their associated target words (VELO). Importantly, the results of Experiment 1 also showed that [+voice] words was not perceived as [-voice] words, since no priming effect was observed in the incongruent priming condition (/pedal/-FLEUR). This observation is particularly interesting since it suggests that the durational cues of voicing are used by listeners in that they prevent the activation of the inappropriate word (/petal/). However, the lack of a priming effect for the [+voice] words /pedal/ on their associated target words (VELO) suggests that at 0 ms of ISI the durational cues of voicing present in whispered speech are not fully exploited by listeners to compensate the loss of information due the absence of vocal fold vibration in [+voice] words, and to guarantee an immediate recognition of these words. Crucially, Experiment 2 revealed that whispered [+voice] words such as /pedal/ are effective primes provided that a 50-ms delay is introduced between the offset of the prime and the display of the target word. Hence, our results show that the retrieval of the [+voice] feature is not immediate during the recognition of whispered words and requires a certain amount of time to be done. Listeners appears to use the residual cues of voicing present in the speech signal to recover whispered words, but a certain amount of time is necessary to compensate the loss of information due to the absence of vocal fold vibration in [+voice] words.

Whispered voiceless primes showed a significant priming effect when the targets were presented at the offset of the primes. Although the magnitude of the priming effect is relatively small, around 30 ms, it corresponds to what is usually observed in studies using cross-modal associative priming (e.g. Marslen-Wilson and Warren, 1994; Snoeren et al., 2008a). The priming effects observed with whispered voiceless primes as well as with modal voiced and voiceless primes disappeared when the ISI between the primes and the targets increased from 0 to 50 ms. At this longer delay, the residual activation caused by the primes in the lexical/semantic space of the targets likely had the time to dissipate by the time the targets were presented (Marslen-Wilson et al., 1996). Interestingly, the pattern of priming effect observed with whispered voiced primes followed a different time course. Indeed, while no priming effect was observed at 0 ms, at 50 ms of ISI a priming effect occurred, and its magnitude was also around 30 ms. Such an observation thus clearly suggests that the 50-ms delay introduced between the offset of the primes and the display of the associated target words has left the time for the target words to be fully activated in the lexical/semantic space, so that an associative priming effect is observed. In accordance with what was observed with voiceless words, the priming effect observed with voiced words at 50 ms ISI also vanished when the ISI was increased by 50 ms.

At a more theoretical level, our study shows that the residual correlates of voicing in whispered speech are exploited by listeners to recover the intended words. Consequently, voiced words were not confounded with their voiceless counterparts. Such a finding does not replicate the observation made by Fux (2012) that [+voice] consonants are poorly identified and confounded with [-voice] consonants in an off-line identification task. However, as in the study of Fux (2012), participants were engaged in an explicit categorisation of the voiced/unvoiced consonants, his findings are likely the result of a late post-lexical decisional process rather than the result of processes involved in lexical access. Our results are in line with previous studies showing that listeners are sensitive to fine-grained differences and use them in online word recognition (e.g. Mitterer and McQueen, 2009; Snoeren et al., 2008a; Spinelli et al., 2003). Nonetheless, they are not totally in line with studies examining the precise moment at which fine-grained acoustic information contributes to the disambiguation of words, by means of response measures such as eye fixations that are closely time-locked to the input as it unfolds over time. Using eye-tracking, Shatzman and McQueen (2006) presented Dutch participants with ambiguous sentences in which stop-initial words (*pot* “jar”) were preceded by *eens* (*eens pot* “once jar”), and which thus could also refer to cluster-initial words (*een spot* “a spotlight”). The pictures of the target objects (*pot*) were presented among three other pictures, one of them corresponding to the competing interpretation (*spot*). The authors showed that the participants were slower to fixate the target pictures (*pot*) when the targets and the preceding /s/ were spliced from a recording of the cluster-initial word (*een spot*) than when the targets and the preceding /s/ were spliced from a recording of the stop-initial words (*eens pot*). In a subsequent experiment, the sentences containing the stop-initial words were manipulated so that the duration of the /s/ preceding the target word was either lengthened or shortened. More fixations to the stop-initial target words were observed for shortened /s/ than for lengthened /s/, thus suggesting that /s/ duration guides listeners in the correct segmentation in case of ambiguous sequences. More importantly, the time course of the effect in the second experiment showed that segment duration biased lexical interpretation almost immediately, as soon as the disambiguating information was heard. This was not the case in our study since disambiguation of the [+voice] words (/pedal/) from its lexical neighbours (/petal/) was not observed when the target words were presented at prime offset. Also, our finding that disambiguation requires a certain amount of time is potentially in contradiction with studies examining the impact of voice assimilation by means of the cross-modal associative priming paradigm. For example, Snoeren et al. (2008a) reported clear priming effect with ambiguous assimilated primes with 0-ms ISI between the prime and the associated target word. Because in Snoeren et al. (2008a)’s study, the ambiguity was on the word final consonant (e.g. the assimilated pronunciation [sud] of the word /sut/), the observation of an associative priming effect with target words presented at prime offset also argues for an immediate effect of the presence of subtle acoustic cues in lexical access. The discrepancy between our study and the preceding ones regarding the rapidity with which subtle acoustic cues bias listeners’ interpretation is likely due to the loss of information in [+voice] words that delay the retrieval of the voiced feature.

Given that listeners appear to use the durational cues of voicing present in whispered speech, how to account that activation of voiced but not unvoiced words takes time? Consider the two main cues of the voiced/voiceless distinction, namely the presence/absence of vocal fold vibration (primary cue) and phoneme durations (secondary cue). The most obvious explanation is that since whispered [+voice] words preserve the secondary but not the primary cue to voicedness, they only partly match their underlying representations, and consequently require time to be fully activated. In contrast, for whispered [-voice] words, the two main cues to voicelessness being preserved, they fully match their underlying representations, and thus are more quickly activated during whispered word recognition. Models assuming abstract representations (TRACE: McClelland and Elman, 1986; Shortlist: Norris, 1994; Norris and McQueen, 2008; Cohort: Marslen-Wilson and Warren, 1994) could account for our results provided that they revise their assumption in such a way that they allow fine-grained acoustic information such as phonemic duration to influence the word recognition process, and in particular the pre-lexical stage of word recognition. For example, TRACE (McClelland and Elman, 1986) assumes three levels of representation consisting in distinctive features, phonemes and words, with excitatory links between them. Thus, when the [+voice] feature unit becomes activated, it increases the activation level of [+voice] phonemes, which in turn pass on their activation to the words that contain them. Consider the whispered voiced words /pedal/ in which a key feature, namely the voicing, is missing. The intervocalic phoneme /d/ should receive less activation from the feature level than when the word /pedal/ is produced in modal voice, because all the features that compose it have not been extracted from the speech signal. In order that the /d/ phoneme gains in activation and surpasses the activation of the [-

voice] /t/ phoneme, acoustic cues such as phoneme duration would play a crucial role at this level of processing, so that for example shorter phoneme duration reinforces activation of [+voice] phonemes and longer phoneme duration reinforces activation of [-voice] phonemes. The reinforcement of the correct phoneme by acoustic duration would allow the intended word to be fully activated at the lexical level and recognised. The assumption that both abstract features and fine-grained information concur in phoneme activation and, at a higher level, in word activation could account for the correct recognition of whispered voiced words, but also, for our observation that these words take time to be recognised. There is less bottom-up evidence in its favour at a featural level of processing, which has thus for consequence to slow down the moment at which the recognition threshold is reached, and thus the moment at which the whispered voiced words can be reliably identified. Of course, the implementation of fine-grained acoustic details in abstractionist models inevitably make them "less" abstract, thus arguing in favour of hybrid models that allow for abstraction, but in the same time reserves an important place for acoustic/phonetic details (see Ernestus, 2014).

Although, in accordance with other studies (Gow, 2002, 2003; Shatzman and McQueen, 2006; Snoeren et al., 2008a; Spinelli et al., 2003), we argue for an important role of durational cues to resolve lexical ambiguity, an alternative explanation in model like TRACE would be that words containing [+voice] intervocalic consonant suffer more from lexical competition than [-voice] consonant. This model assumes a lexical competition process by the means of intra-level inhibition between phonologically similar words. The more a word is activated, the more it sends inhibition to its lexical neighbours. Due to the loss of information in whispered words containing [+voice] intervocalic consonant, a word like /pedal/ should receive strong inhibition from its lexical neighbours /petal/ which is more compatible with the information present in the speech signal, thus receiving strong activation from the input. As a result, for words like /pedal/ the resolution of the lexical competition in its favour requires a certain amount of time, thus delaying its recognition. In accordance with our claim, the resolution of the lexical competition process for [+voice] words could be done by the use of the durational cues present in the speech signal.

Models of spoken-word recognition in which fine-grained acoustic information such as durational information is viewed as inherent properties of lexical representations could account for our finding that listeners use subtle acoustic cues during the recognition of whispered words. In exemplar-based models of spoken word recognition (Godinger, 1998; Johnson, 1997), instances or exemplars of words are stored in memory with their acoustic details, and word recognition consists in finding the nearest match in a vast collection of exemplars. Assuming that our prime words are frequently used when speakers whisper, and thus are frequently encountered by listeners, both whispered voiced words and whispered voiceless words would be stored with their own durational characteristics. As results, due to the presence in memory of whispered voiced and voiceless exemplars, both voiced and voiceless words should find a match among the whispered exemplars and should be immediately recognized. Consequently, it is unclear how exemplar-based models could account for our observation that whispered voiced words require more time to be recognised than whispered voiceless words.

Interestingly, the main finding of this study, namely that the retrieval of the voiced feature during whispered word recognition takes time, suggests in a more general way that the perception and then the successful recognition of non-prototypical stimuli (here, one token of the word /pedal/ without voicing) is not so easy. The difficulty which causes additional processing time has been often observed in speech or in written word perception whenever the input deviates from the listener's abstract internalised representations, and thus requires repair. An example is the no-permissible sequences of phonemes /tl/ at the beginning of French words which is perceived and repaired as /kl/ by French listeners (Hallé, Segui, Frauenfelder & Meunier, 1998). Using a phoneme monitoring task, Hallé et al. (1998) have reported longer monitoring times for /k/ in the auditory sequence /tlabo/ - repaired into /klabo/ - than in the "true" auditory sequence /klabo/. In a more recent study, Hallé and collaborators (Hallé, Dominguez, Cueto & Segui, 2008) tested in visual masked priming experiments, Spanish participants who pronounce *esport* for *sport* or *especial* for *special*, and thus hear a prothetic /e/ in foreign words. They observed that a prime such as *special* facilitated the processing of Spanish target word "especial" as the prime *especial* do, at 88 ms and 132 ms SOAs (Stimulus Onset Asynchrony; the delay between the beginning of the prime and the beginning of the target) but not at 44 ms SOA. Although the authors used masked priming and consequently manipulated the SOA, which renders difficult to establish a clear parallelism between the delays used in their study and in this study (i.e. ISI in our experiments), the two studies concur in showing that altered speech sounds require additional processing time, so that the corresponding underlying representation be fully activated in the mental lexicon.

To conclude, our study adds to the growing body of research showing that listeners are sensitive to fine phonetic details, and use them in word recognition. However, what our study also revealed is that in spite of the acoustic correlates of voicing in whispered speech, the retrieval of the voiced feature takes time. At a more theoretical level, our study argues for the existence of abstract representations whose activation can be modulated by fine-grained acoustic information.

FOOTNOTES

1: In the main experiment, the number of participants was twice as high as in the control experiment, because as explained above, in the control experiment, each type of prime words (voiced/voiceless) was tested only with their respective associated (congruent) target words.

2: We first conducted a mixed-effects model on all data (collapsed across voiced and voiced word types) to check that the predictor word type interacted significantly with the predictors prime type (related, control) and semantic matching (congruent, incongruent). Note that in this model, the three-way interaction between word type, prime type and semantic matching failed to reach significance. This is not surprising due to the fact that for methodological reasons two of our variables were between-participants factors, namely semantic matching (congruent, incongruent) and type of words (voiced, voiceless), and thus we likely lack statistical power. Note also that as explained at the end of the introduction, we are specifically concerned by the interaction between prime type (related, control) and semantic matching (congruent, incongruent) within each type of words (voiced, voiceless).

3: A possibility however is that the differential priming observed with the whispered voiceless (/petal/) and voiced (*/pedal/*) prime words in the congruent conditions merely results from a difference in the speed of responses of the two groups of participants. Indeed, as indicated in Figure 3, in the congruent conditions, the participants that heard whispered voiceless primes /petal/ responded in average faster that the participants that heard whispered voiced primes /pedal/, and thus it could be argued that long RTs cause no semantic priming effect. Additional analyses performed in the voiceless priming condition (/petal/-FLEUR) in which a clear priming effect was observed did not confirm this claim. The 12 slowest participants (RTs on the control primes greater than 700 ms) showed a strong priming effect around 75 ms (mean RTs 786 ms and 711 ms for the control and related primes respectively) whereas the 12 fastest participants (RTs on the control primes smaller than 640 ms) showed a priming effect around 10 ms (not significant; mean RTs 584 ms and 574 ms for the control and related primes respectively).

Acknowledgements

We are grateful to anonymous reviewers for their helpful comments on earlier versions of this manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Labex Brain and Language Research Institute [grant number ANR-11-LABX-0036] and benefited support from the French National Agency of Research (ANR), under the project entitled "Investments of the Future" A*MIDEX [grant number ANR-11-IDEX-0001-02.

.

REFERENCES

Abdelli-Beruh, N. B. (2004). The stop voicing contrast in French sentences: Contextual sensitivity of vowel duration, closure duration, voice onset time, stop release and closure voicing. *Phonetica*, 61(4), 201–219.

Andruski, J. E., Blumstein, S. E., and Burton, M. (1994). The effects of subphonetic differences on lexical access. *Cognition, 52*, 163–187.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge: Cambridge University Press.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59,* 390-412.

Baayen, R. H., and Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, 3(2), 12–28.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal *Journal of Memory and Language, 68*, 255–278.

Bates, D. M. and Sarkar, D. (2007). lme4: Linear mixed-effects models using S4 classes. R package version 2.6, retrieved 20 November 2017 from http://lme4.r-forge.r-project.org/.

Boersma, P. and Weenink, D. (2017). Praat: doing phonetics by computer [Computer program]. Version 6.0.26. Retrieved 2 March 2017 from http://www.praat.org/.

Bretz, F., Hothorn, T., and Westfall, P. H. (2011). *Multiple comparisons using R*. Boca Raton: CRC Press.

Chen, M. (1970). Vowel length ariation as a function of the voicing of the consonant environment. *Phonetica*, 22(3), 129–159.

Cirillo, J. (2004). Communication by unvoiced speech: The role of whispering. *Anais Da Academia Brasileira de Ciências*, 76(2), 413–423.

Connine, C. M., Blasko, D. G., and Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language*, *32*, 193–210.

Dannenbring, G. L. (1980). Perceptual discrimination of whispered phoneme pairs. *Perceptual and Motor Skills*, 51(3), 979–985.

Dufour, S. (2008). Phonological priming in auditory word recognition: When both controlled and automatic processes are responsible for the effects. *Canadian Journal of Experimental Psychology*, 62, 33-41.

Ernestus, M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua. International Review of General Linguistics. Revue internationale de Linguistique Générale*, 142, 27–41.

Ferrand, L., and Alario, F.-X., (1998). Normes d'association verbales pour 366 d'objets concrets. *L'Année Psychologique, 98*, 659-709.

Flege, J. E., Munro, M. J., and Skelton, L. (1992). Production of the word-final English /t/–/d/ contrast by native speakers of English, Mandarin, and Spanish. *The Journal of the Acoustical Society of America*, 92(1), 128–143.

Fleischer, S., Kothe, C., and Hess, M. (2007). Glottal and supraglottal configuration during whispering. *Laryngo-Rhino-Otology*, 86(4), 271–275.

Fux, T. (2012). *Vers un système indiquant la distance d'un locuteur par transformation de sa voix*. Ph Dissertion, Université de Grenoble : Grenoble. Available from https://tel.archives-ouvertes.fr/tel-01557936.

Gilichinskaya, Y., Strange, W. (2011). Perception of final-consonant “voicing” in whispered speech. *The Journal of the Acoustical Society of America*, 129, 2420.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251–279.

Goldinger, S. D., Luce, P. A., and Pisoni, D. B. (1989). Priming lexical neighbours of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, *28,* 501–518.

Gow, D. W. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language, 24*, 133–159.

Gow, D. W. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance, 28*, 163–179.

Gow, D. W. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception and Psychophysics, 65*, 575–590.

Gow, D. J., Jr., and Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 344–359.

Hallé, P., Dominguez, A., Cuetos, F., and Segui, J. (2008). Phonological mediation in visual masked priming: Evidence from phonotactic repair. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 177-192.

Hallé, P., Segui, J., Frauenfelder, U., and Meunier, C. (1998). The processing of illegal consonant clusters: A case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 592-608.

Ito, T., Takeda, K., and Itakura, F. (2005). Analysis and recognition of whispered speech. *Speech Communication*, 45(2), 139–152.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434-446.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson and J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145–165). San Diego, CA: Academic Press.

Jovičić, S. T., and Šarić, Z. (2008). Acoustic analysis of consonants in whispered speech. *Journal of Voice, 22*, 263-274.

Kohlberger, M., and Strycharczuk, P. (2015). Voicing assimilation in whispered speech. *Proceedings of the 18th International Conference on Phonetic Sciences.* Glasgow.

Kuznetsova, A., Bruun Brockhoff, P., Haubo Bojesen Christensen, R. (2016): Tests in linear mixed effect models. Available from https://cran.r-project.org/package=lmerTest.

Lisker, L., and Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustic measurements. *Word, 20*, 324–422.

Mack, M. (1982). Voicing-dependent vowel duration in English and French: monolingual and bilingual production. *Journal of the Acoustical Society of America*, *71*, 173–178.

Marslen-Wilson, W.D., Moss, H.E., and van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 1376-1392.

Marslen-Wilson, W. D., and Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, 101, 653–675.

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

Meynadier, Y., and Gaydina, Y. (2013). Aerodynamic and durational cues of phonological voicing in whisper. *Proceedings of the 14th Annual Conference of the International Speech Communication Association* (pp. 335–339). Lyon.

Mitterer, H., and McQueen, J.M. (2009). Processing reduced word-forms in speech perception using probabilistic knowledge about speech production. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 244-263.

Munro, M. J. (1990). Perception of "voicing" in whispered stops. *Phonetica*, 47(3-4), 173–181.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52, 189–234.

Norris, D., and McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115, 357–395.

O'Shaughnessy, D. (1981). A study of French vowel and consonant durations. *Journal of Phonetics*, 9, 385–406.

Parnell, M., Amerman, J. D., and Wells, G. B. (1977). Closure and constriction duration for alveolar consonants during voiced and whispered speaking conditions. *The Journal of the Acoustical Society of America*, 61(2), 612–613.

R Development Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available from *http://www.R-project.org*.

Schwartz, M. F. (1972). Bilabial closure durations for /p/, /b/, and /m/ in voiced and whispered vowel environments. *The Journal of the Acoustical Society of America*, 51(6B), 2025–2029.

Sharf, D. J. (1964). Vowel duration in whispered and in normal *speech. Language and Speech*, 7(2), 89–97.

Shatzman, K. B., and McQueen, J. M. (2006). Segment duration as a cue to word boundaries in spoken-word recognition. *Perception and Psychophysics*, 68, 1–16.

Snoeren, N., Segui, J., and Hallé, P. (2008a). On the role of regular phonological variation in lexical access: Evidence from voice assimilation in French. Cognition, 108, B512-B521.

Snoeren, N., Segui, J., and Hallé, P. (2008b). Perceptual processing of partially and fully assimilated words in French. *Journal of Experimental Psychology: Human Perception and Performance*, 34, 193-204.

Solé, M.-J., and Ohala, J. J. (2010). What is and what is not under the control of the speaker: Intrinsic vowel duration. In C. Fougeron, B. Kühnert, M. D'Imperio, and N. Vallée (Eds.), *Phonology and Phonetics*, vol. 10, pp.607-656. Berlin: De Gruyter Mouton.

Spinelli, E., McQueen, J. and Cutler, A. (2003). Processing resyllabified words in French. *Journal of Memory and Language*, 48, 233-254.

Tabossi, P. (1996). Cross-modal semantic priming. *Language and Cognitive Processes*, 11, 569–576.

Tartter, V. C. (1989). What's in a whisper? *The Journal of the Acoustical Society of America*, 86, 1678–1683.

Tsunoda, K., Niimi, S., and Hirose, H. (1994). The roles of the posterior cricoarytenoid and thyropharyngeus Muscles in whispered speech. *Folia Phoniatrica et Logopaedica*, 46(3), 139–151.

van de Velde, D. J., and van Heuven, V. (2011). Compensatory strategies for voicing of initial and medial plosives and fricatives in whispered speech in Dutch. *Proceedings of the 17th International Conference on Phonetic Sciences* (pp. 2058–2061). Hong-Kong.

Table 1: Characteristics of the prime and target words

	Frequency[1]	Number of syllables	Number of phonemes/ graphemes[2]	Uniqueness point[3]	Percentage of association
Voiced consonants					
Target words (VELO)	44	1.75	5.55	-	41.22
Related primes (/pedal/)	21	2.20	4.35	4.95	41.22
Control primes (/kitɑ̃s/)	16	2.20	4.35	4.85	-
Voiceless consonants					
Target words (FLEUR)	106	1.80	5.90	-	40.30
Related primes (/petal/)	14	2.20	4.35	4.80	40.30
Control primes (/kitɑ̃s/)	16	2.20	4.35	4.85	-

Note: [1] In number of occurrences per million. [2] Number of phonemes for the auditory primes/Number of graphemes for the visual targets. [3] The phonemic position at which the auditory word primes can be reliably identified.

Table 2: Intervocalic consonant durations (in ms) and pre-consonantal vowel durations (in ms) for each minimal pair.

	Modal phonation				Whispered phonation			
	Consonnant duration		**Vowel duration**		**Consonnant duration**		**Vowel duration**	
Minimal pairs	**Voiceless**	**Voiced**	**Voiceless**	**Voiced**	**Voiceless**	**Voiced**	**Voiceless**	**Voiced**
appât/abat	192	114	96	109	175	118	126	134
affaler/avaler	132	63	90	106	139	76	109	178
agrafer/aggraver	136	58	84	147	161	94	100	137
amphi/envie	158	71	150	228	117	67	195	206
bateau/badaud	133	101	117	131	144	77	144	171
briquer/briguer	138	100	125	140	131	95	99	175
capot/cabot	167	119	92	112	181	130	118	126
cachou/cajou	168	111	109	175	200	135	129	170
casser/caser	201	111	116	190	215	123	134	191
compas/combat	86	53	180	210	113	75	148	198
comte/condé	122	47	195	242	105	88	186	209
dessert/désert	167	100	95	151	184	107	97	197
écho/égaux	183	121	84	101	159	80	139	174
empocher/embaucher	106	72	128	148	104	91	110	135
embrasser/embraser	172	86	98	148	175	124	107	132
gaffer/gaver	161	74	130	187	168	92	146	208
motel/modèle	151	71	128	133	169	86	145	161
pétale/pédale	145	85	91	116	142	90	142	143
râteau/radeau	148	107	125	135	194	95	125	177
refus/revue	162	112	123	167	189	114	177	211
mean	**151**	**89**	**118**	**154**	**158**	**98**	**134**	**172**

Figure 1: Waveform and spectrogram of the /pedal/ (left) and /petal/ (right) words in modal (top) and whispered (bottom) phonations. The critical consonants /d/ and /t/ are bounded by the solid vertical lines.

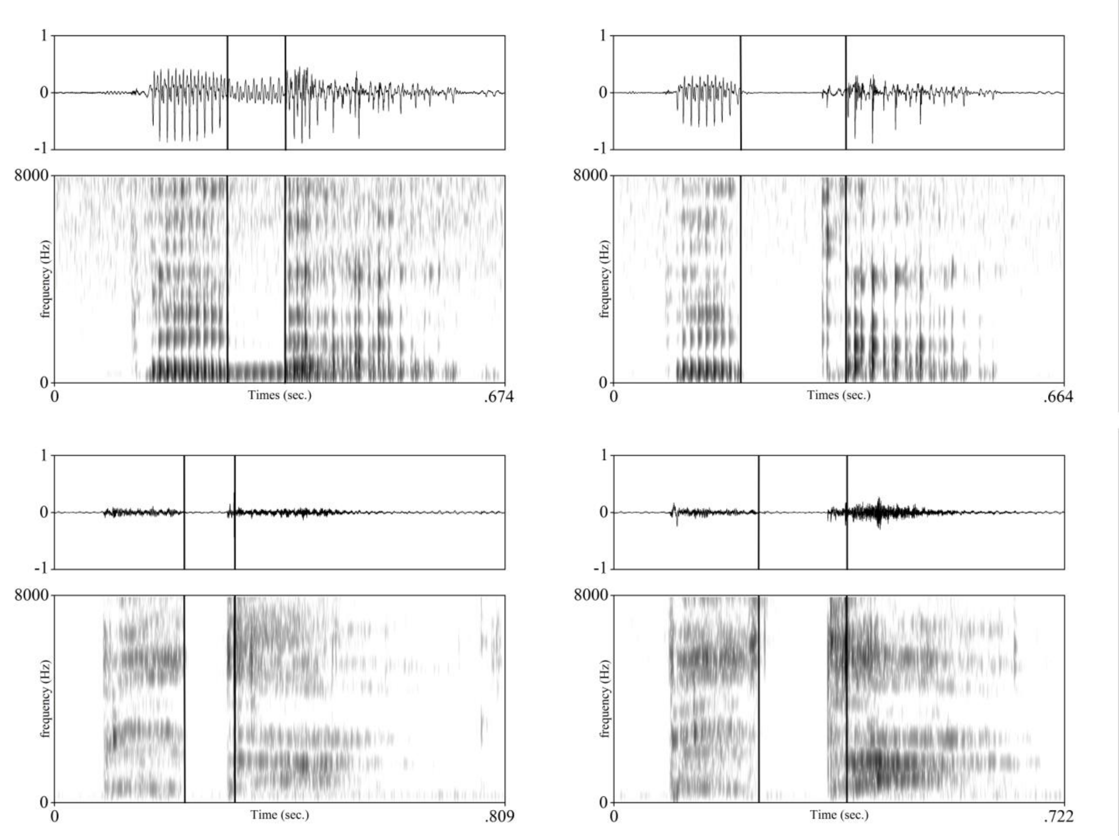


Figure 2: Mean Reaction Times (in ms) and Standard Errors for the control and related primes, and for the voiceless (left) and voiced (right) words in modal phonation in Experiment 1 (control experiment). Percentages of correct responses are shown below the bar for each condition. * indicates significant priming effects.

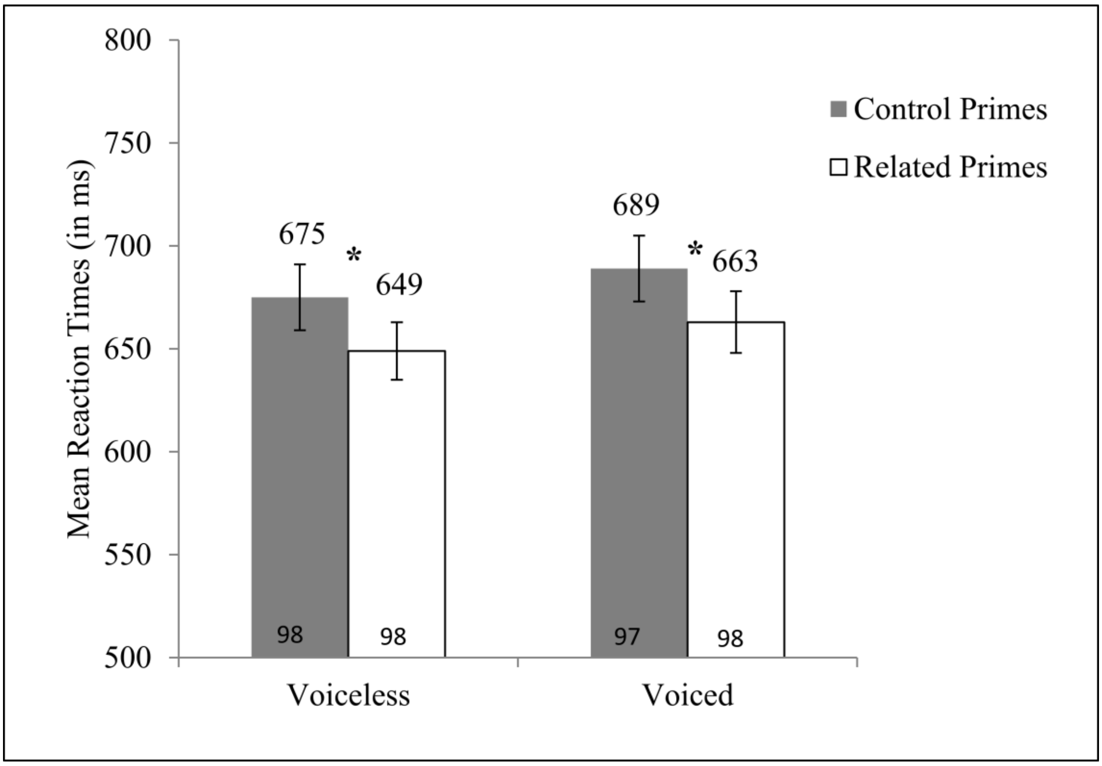


Figure 3: Mean Reaction Times (in ms) and Standard Errors for the control and related primes as a function of the semantic matching (congruent, incongruent) for the whispered voiceless (top) and voiced (bottom) words in Experiment 1 (main experiment). Percentages of correct responses are shown below the bar for each condition. * indicates significant priming effect.

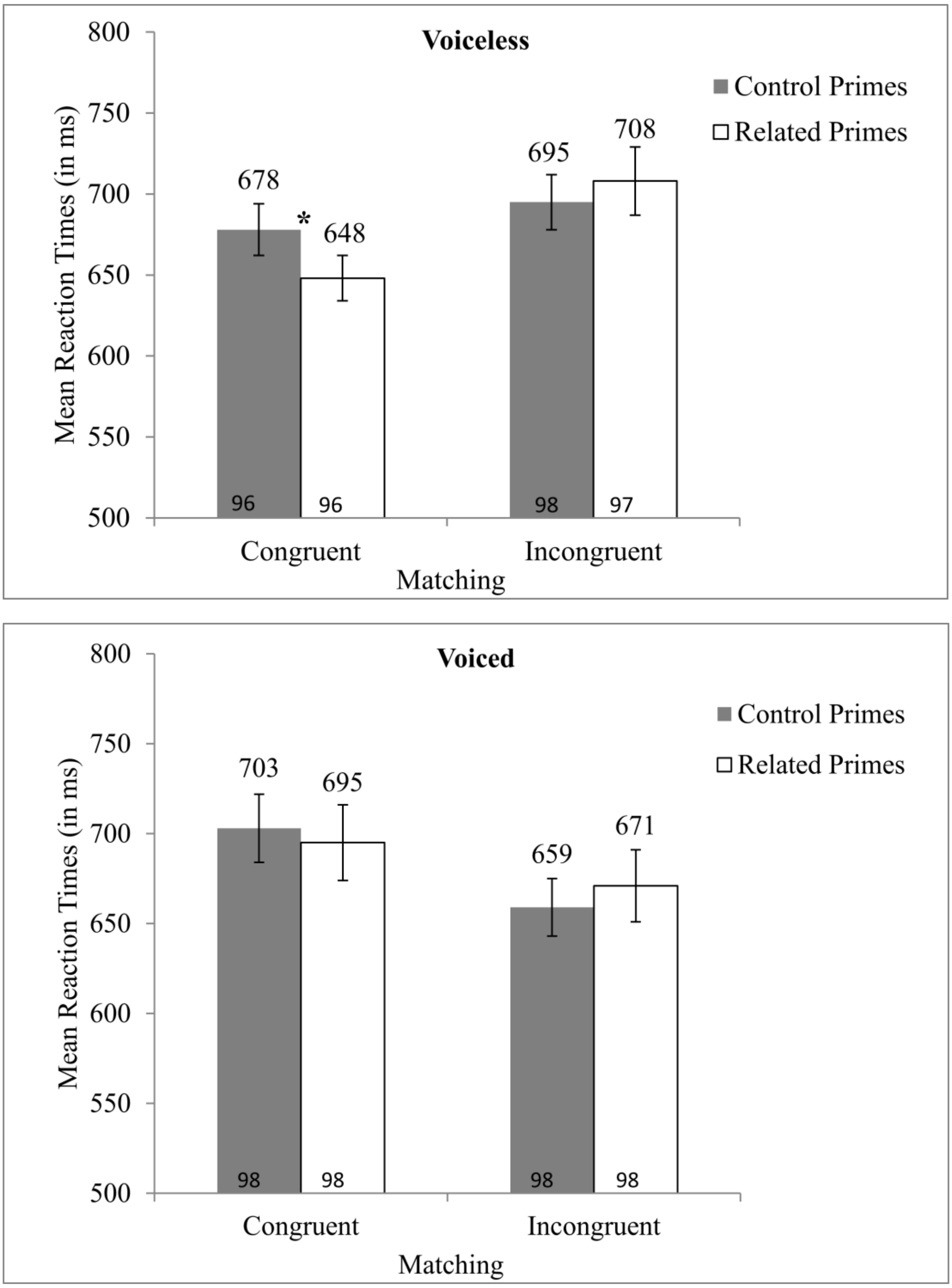


Figure 4: Mean Reaction Times (in ms) and Standard Errors for the control and related primes and for the voiceless (left) and voiced (right) words in modal phonation in Experiment 2 (control experiment). Percentages of correct responses are shown below the bar for each condition.

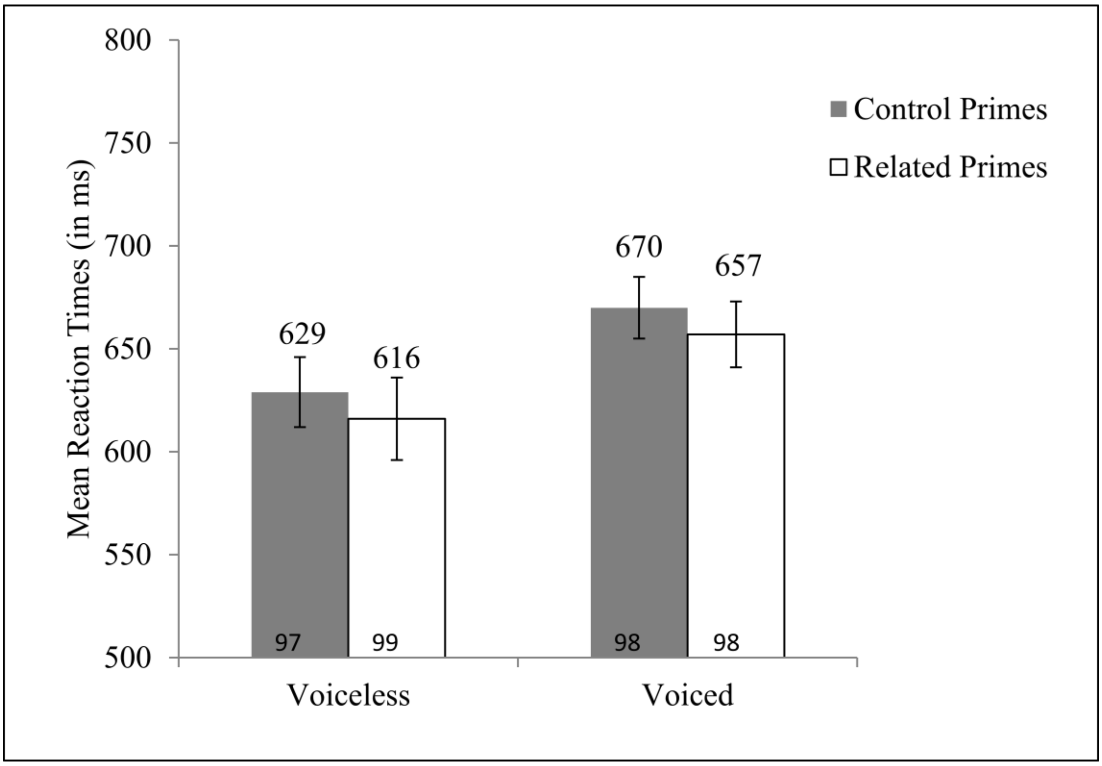
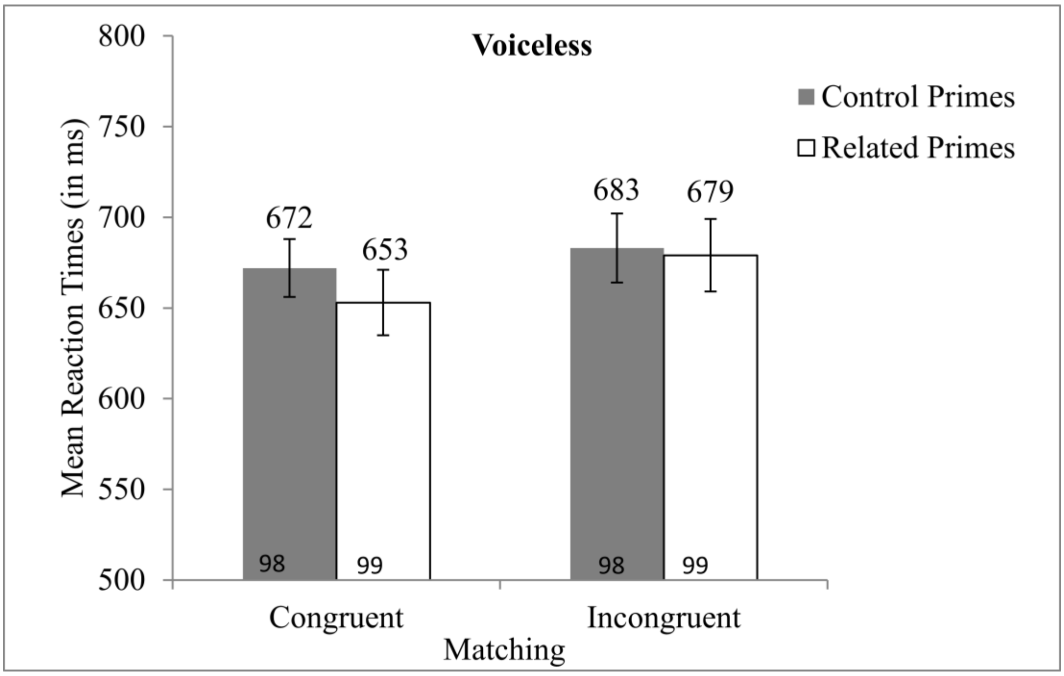


Figure 5: Mean Reaction Times (in ms) and Standard Errors for the control and related primes as a function of the semantic matching (congruent, incongruent) for the whispered voiceless (top) and voiced (bottom) words in Experiment 2 (main experiment). Percentages of correct responses are shown below the bar for each condition. * indicates significant priming effect.



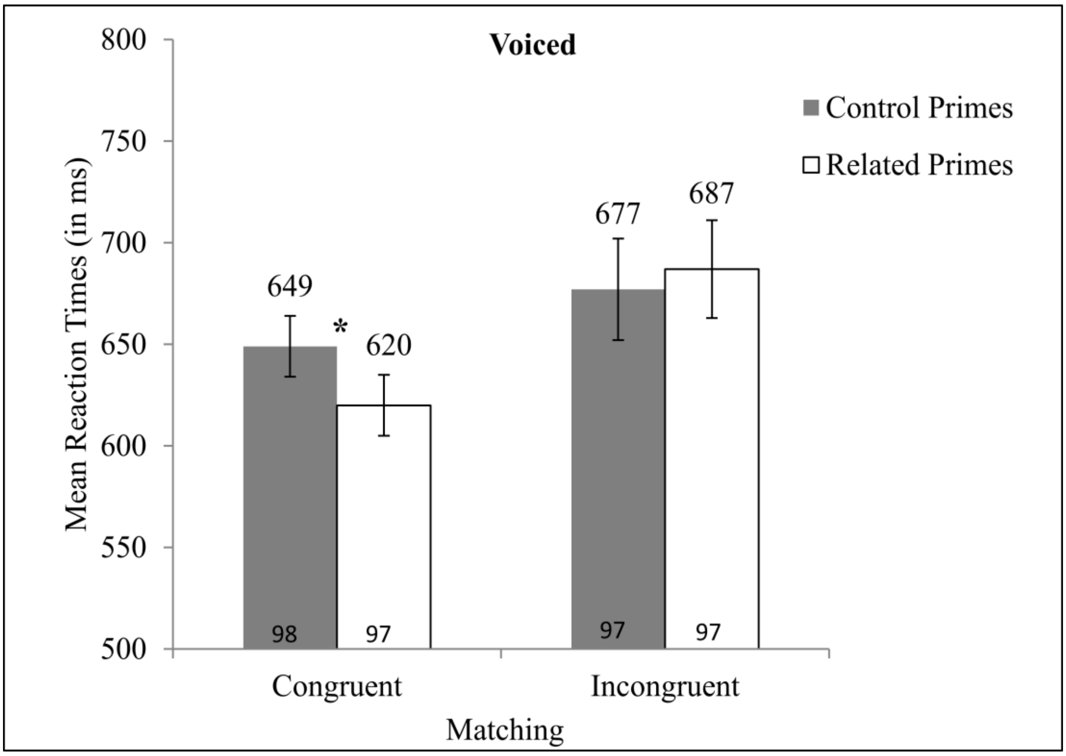


Figure 6: Mean Reaction Times (in ms) and Standard Errors for the control and related primes and for the voiceless (left) and voiced (right) words in the congruent conditions of Experiment 3. Percentages of correct responses are shown below the bar for each condition.

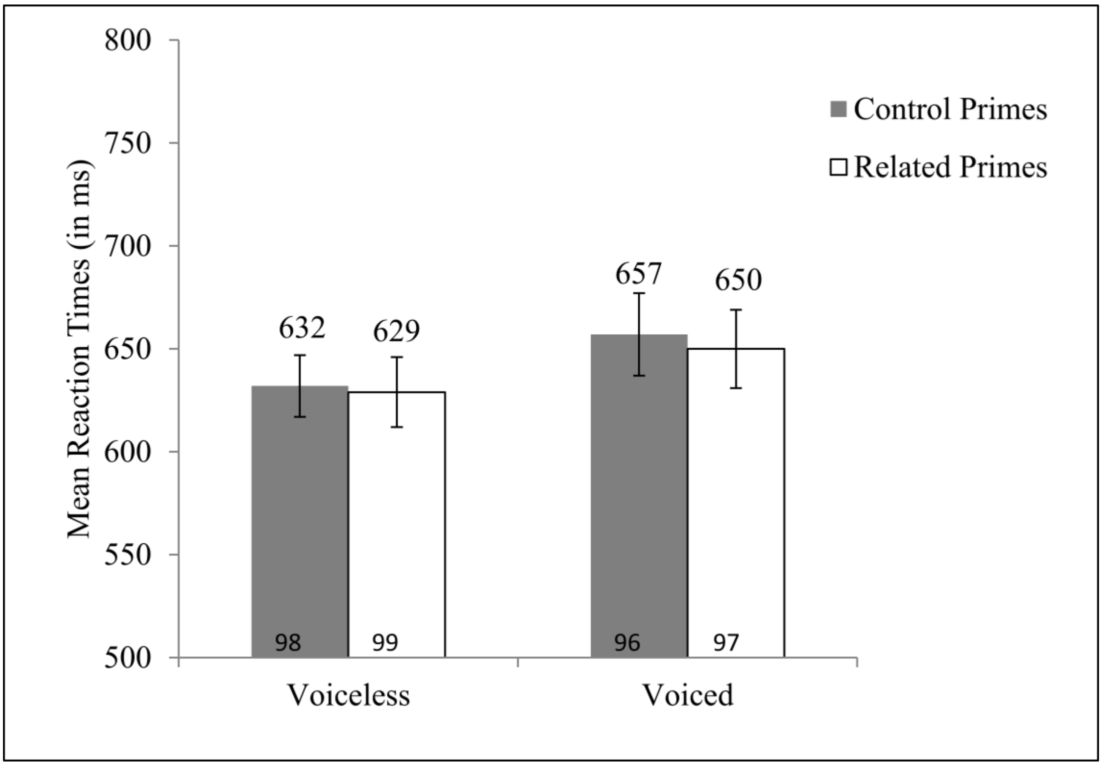
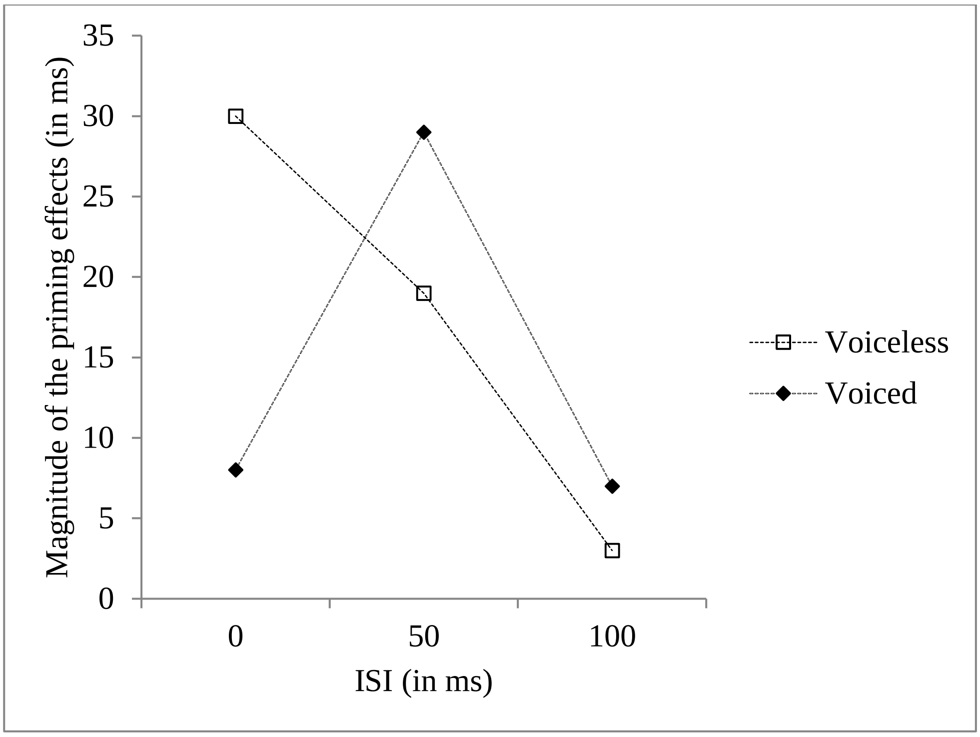


Figure 7: Time-course of the semantic priming effect for the voiceless and voiced words.



Appendix: Primes and targets used in the Experiments

Voiced words

Congruent condition			Incongruent condition		
Control Primes	**Related Primes**	**Target Words**	**Control Primes**	**Related Primes**	**Target Words**
grossi	briguer	MANDAT	grossi	briguer	NETTOYER
jumelle	désert	SABLE	jumelle	désert	CHOCOLAT
bandit	revue	MAGAZINE	bandit	revue	NON
obus	abat	VIANDE	obus	abat	POISSON
pincer	gaver	OIE	pincer	gaver	ERREUR
lingot	cajou	NOIX	lingot	cajou	BONBON
entraver	aggraver	PIRE	entraver	aggraver	FEUILLES
douleur	modèle	MANNEQUIN	douleur	modèle	HOTEL
outil	égaux	PAREILS	outil	égaux	MONTAGNE
héro	envie	DESIR	héro	envie	COURS
guidon	badaud	PASSANT	guidon	badaud	MER
talent	radeau	MEDUSE	talent	radeau	PELLE
fumer	caser	RANGER	fumer	caser	BRISER
éduquer	embaucher	TRAVAIL	éduquer	embaucher	ARGENT
abouti	avaler	DEGLUTIR	abouti	avaler	CANAPE
tampon	combat	BOXE	tampon	combat	MATHS
quittance	pédale	VELO	quittance	pédale	FLEUR
impliquer	embraser	FEU	impliquer	embraser	BISOU
pinceau	cabot	CHIEN	pinceau	cabot	VOITURE
verrue	condé	FLIC	verrue	condé	FROMAGE

Appendix continuation

Voiceless words

Congruent condition			Incongruent condition		
Control Primes	**Related Primes**	**Target Words**	**Control Primes**	**Related Primes**	**Target Words**
grossi	briquer	NETTOYER	grossi	briquer	MANDAT
jumelle	dessert	CHOCOLAT	jumelle	dessert	SABLE
bandit	refus	NON	bandit	refus	MAGAZINE
obus	appât	POISSON	obus	appât	VIANDE
pincer	gaffer	ERREUR	pincer	gaffer	OIE
lingot	cachou	BONBON	lingot	cachou	NOIX
entraver	agrafer	FEUILLES	entraver	agrafer	PIRE
douleur	motel	HOTEL	douleur	motel	MANNEQUIN
outil	écho	MONTAGNE	outil	écho	PAREILS
héro	amphi	COURS	héro	amphi	DESIR
guidon	bateau	MER	guidon	bateau	PASSANT
talent	râteau	PELLE	talent	râteau	MEDUSE
fumer	casser	BRISER	fumer	casser	RANGER
éduquer	empocher	ARGENT	éduquer	empocher	TRAVAIL
abouti	affaler	CANAPE	abouti	affaler	DEGLUTIR
tampon	compas	MATHS	tampon	compas	BOXE
quittance	pétale	FLEUR	quittance	pétale	VELO
impliquer	embrasser	BISOU	impliquer	embrasser	FEU
pinceau	capot	VOITURE	pinceau	capot	CHIEN
verrue	comté	FROMAGE	verrue	comté	FLIC