Facilitated Auditory Detection for Speech Sounds

Carine Signoret, Etienne Gaudrain, Barbara Tillmann, Nicolas Grimault, Fabien Perrin

frontiers in PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 26 July 2011
doi: 10.3389/fpsyg.2011.00176

Facilitated auditory detection for speech sounds

Carine Signoret[1,2]*, Etienne Gaudrain[3,4], Barbara Tillmann[1,2], Nicolas Grimault[1,2] and Fabien Perrin[1,2]

[1] CNRS UMR5292, INSERM U1028, Lyon Neuroscience Research Center, Auditory Cognition and Psychoacoustics Team, Lyon, France
[2] Université de Lyon, Lyon, France
[3] Centre for the Neural Basis of Hearing, Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK
[4] Medical Research Council Cognition and Brain Sciences Unit, Cambridge, UK

Edited by:
Micah M. Murray, Université de Lausanne, Switzerland

Reviewed by:
Marianne Latinus, University of Glasgow, UK
Bruno Lucio Giordano, McGill University, Canada

***Correspondence:**
Carine Signoret, Cognition Auditive et Psychoacoustique, Lyon Neuroscience Research Center, 50 Avenue Tony Garnier, 69366 Lyon Cedex 07, France.
e-mail: carine.signoret@olfac.univ-lyon1.fr

If it is well known that knowledge facilitates higher cognitive functions, such as visual and auditory word recognition, little is known about the influence of knowledge on detection, particularly in the auditory modality. Our study tested the influence of phonological and lexical knowledge on auditory detection. Words, pseudo-words, and complex non-phonological sounds, energetically matched as closely as possible, were presented at a range of presentation levels from sub-threshold to clearly audible. The participants performed a detection task (Experiments 1 and 2) that was followed by a two alternative forced-choice recognition task in Experiment 2. The results of this second task in Experiment 2 suggest a correct recognition of words in the absence of detection with a subjective threshold approach. In the detection task of both experiments, phonological stimuli (words and pseudo-words) were better detected than non-phonological stimuli (complex sounds), presented close to the auditory threshold. This finding suggests an advantage of speech for signal detection. An additional advantage of words over pseudo-words was observed in Experiment 2, suggesting that lexical knowledge could also improve auditory detection when listeners had to recognize the stimulus in a subsequent task. Two simulations of detection performance performed on the sound signals confirmed that the advantage of speech over non-speech processing could not be attributed to energetic differences in the stimuli.

Keywords: speech detection effect, auditory threshold, recognition, model, knowledge

INTRODUCTION

Detection can be performed without involving any knowledge-based processing. Nevertheless, it has been shown that knowledge about a visual stimulus can influence its detection. The present study investigated knowledge-based influences on auditory detection by comparing three types of stimulus varying in their degree of phonological or lexical content. Our findings indicate that the knowledge-based processes, although not mandatory for the task, were automatically engaged when the relevant sounds were presented.

The influence of knowledge on stimulus processing was demonstrated for the first time by Cattell (1886). Words and non-words were visually presented for a short duration (5–10 ms) and the participants had to report as many letters as they could. The author reported a word superiority effect (WSE): target letters were better reported, identified, or recognized when they were part of a word than when they were part of either a pseudo-word (an orthographically legal pronounceable letter string) or a non-word (an orthographically illegal unpronounceable letter string). This finding, later replicated in numerous studies (Grainger and Jacobs, 1994; e.g., Reicher, 1969; Wheeler, 1970; McClelland, 1976; Grainger et al., 2003), suggests that short-term memory limitations can be more easily overcome for words than for pseudo-words or non-words because the lexical knowledge allows reconstructing the word and thus to the reporting of more letters. Similarly, a pseudo-word superiority effect (PWSE) has also been reported in the visual modality (Baron and Thurston, 1973; McClelland, 1976; McClelland and Johnston, 1977; Grainger and Jacobs, 1994). This effect refers to facilitated perception due to phonological features of the stimulus: target letters are better identified when they are part of a pseudo-word than when they are part of a non-word (e.g., McClelland and Rumelhart, 1981; Maris, 2002). To our knowledge, no study has investigated WSE and PWSE in the auditory modality. However, some studies have shown that listeners' lexical knowledge can also influence spoken word processing. When a phoneme in a spoken sentence was replaced by a non-speech sound, the participants' lexical knowledge filled in the missing speech sound: they were not aware of the missing phoneme and could not specify the location of the non-speech sound in the sentence they had just heard (Warren, 1970, 1984; Warren and Obusek, 1971; Warren and Sherman, 1974; see also Pitt and Samuel, 1993).

Whereas previous studies have shown that linguistic knowledge can influence tasks involving higher-level processing (such as letter identification or lexical discrimination), only two studies have investigated the possible influence of lexical or phonological knowledge on tasks relying on *more basic processes*. In the visual modality, Doyle and Leach (1988) and Merikle and Reingold (1990) have reported an advantage of words over non-words for detection, which has been called the word detection effect (WDE). In the study of Doyle and Leach (1988), participants more readily detected words than non-words that were briefly displayed on the screen. However, any difference in detection might be attributed to a difference in physical properties because the two sets of stimuli were not matched for the number of letters. Merikle and Reingold (1990) performed this control and also observed a WDE when the onset asynchrony between the visual target and the following

mask was so short that the participants had difficulties detecting the target stimulus. To our knowledge, no study has investigated this facilitation in the auditory modality. In particular, no study has investigated the influence of lexical or phonological knowledge on auditory detection, i.e., whether words, or more generally speech stimuli, are better detected than non-words (WDE) or non-speech stimuli (speech detection effect, SDE). Because a detection task does not require any form of knowledge, such an influence would indicate that knowledge is automatically activated when the relevant type of sound is presented.

In the present study, we investigated the influence of phonological and lexical knowledge on auditory detection and perception using three types of stimulus: words, pseudo-words, and non-phonological complex sounds. To minimize the differences in sensory processing, we matched the energetic properties of the three stimulus types as closely as possible (i.e., in terms of loudness, duration, temporal envelope, and average spectrum). This was further assessed by simulating the detection of the experimental material using models of the auditory system, which were missing knowledge-based influences. In addition, the words and the pseudo-words used in the experiments were composed of the same phonemes to minimize phonological differences. The participants performed a detection task (Experiments 1 and 2) that was followed by a two alternative forced-choice (2AFC) recognition task in Experiment 2. Therefore, any SDE, if observed, could only be related to speech being processed differentially due to the listeners' phonological and/or lexical knowledge.

EXPERIMENT 1

The aim of Experiment 1 was to investigate whether phonological and lexical knowledge influence auditory detection. Non-phonological complex sounds, pseudo-words, and words were randomly presented at sound levels that ranged from inaudible to audible. For each trial, the participants had to detect the presence or absence of a stimulus.

To test the potential influence of phonological and/or lexical knowledge on detection performance, we compared auditory detection performance for stimuli that were energetically matched while manipulating the amount of phonological or lexical content. We equalized all the stimuli by adjusting their presentation level to a common value. Detection performance was then measured over a range of presentation levels to obtain psychometric curves. Because the way these levels were adjusted may affect the shape of the psychometric function, two different equalization schemes were tested in Experiment 1: one based on the dB-sound pressure level (dB-SPL), and the other based on the dB-A levels. These equalization schemes differ by the weight given to each frequency in the signal when computing the overall level. The flat-weighting used in the dB-SPL equalized the stimuli in term of their physical energy. The A-weighting roughly mimics the external and middle ear transfer functions: the stimuli were equalized in the energy that reached the inner ear. Although dB-A has been reported to be better for equalizing isolated vowels (Kewley-Port, 1991), the presence of consonants in words broadens the long-term spectrum, which could yield a different outcome. Moreover, speech production does not imply constant loudness because some phonemes are naturally louder than others. The actual loudness variability in natural running speech must be between equal loudness (which can be approximated by dB-A) and equal production effort (which is close to dB-SPL)[1]. Therefore, in Experiment 1, the best equalization scheme would be the one that minimizes the detection variance across stimuli, i.e., the one that yields the steepest psychometric curve. However, the sought effects will be considered robust only if they appear with both equalization schemes as they would prove resilient to the natural variability of speech production across stimuli.

MATERIAL AND METHODS

Participants

Twenty students (age 22.9 ± 3.7 years, 16 females), right-handed on the "Edinburgh Handedness Inventory" (Oldfield, 1971), were included in Experiment 1. All were French native speakers and did not report any hearing problems or history of neurological disease. All participants had normal hearing, i.e., their pure tone thresholds (as described in ANSI, 2004) were below 15 dB-HL for frequencies between 250 and 8000 Hz. All participants provided written informed consent to the study, which was conducted in accordance with the guidelines of the Declaration of Helsinki and approved by the local Ethics Committee (CPPRB Léon Bérard, no. 05/026).

Materials

Three types of stimulus were used: words, pseudo-words, and complex sounds. Words were selected from a French database (Lexique 2, New et al., 2004). They were common, singular monosyllabic nouns and contained two to seven letters and two to five phonemes. All words had a frequency of occurrence higher than 1 per million occurrences in books and movies (subtitles, New et al., 2007) and were uttered by the same female speaker. A list of monosyllabic pseudo-words was generated from words, by mixing all the phonemes of the words. Pseudo-words could be pronounced, but did not have any meaning[2]. The number of letters and phonemes were matched between words and pseudo-words. The pseudo-words were uttered by the same female speaker as the words. The average durations of the words, pseudo-words, and complex sounds (see below) were not significantly different [$F(2,358) = 0.73$; $p = 0.48$] and were 527.2 ms (SD = 103.7 ms), 542.6 ms (SD = 82.1 ms), and 531.3 ms (SD = 93.5 ms) respectively.

The complex sounds were created from the words and pseudo-words using the algorithm *Fonds sonores* (Perrin and Grimault, 2005; Hoen et al., 2007; see **Supplementary Material S2** for a

[1]We could not find any direct evidence supporting this claim in the literature. However, speech synthesizers are based on this assumption as the level is controlled at the source rather than by feedback from the output (e.g., Klatt, 1980). We also tested this assumption from long recordings of running speech and found that the variability of loudness (as estimated by the level in dB-A) was 10–30% greater than that of production energy (as estimated by the level in dB-SPL), thus confirming that natural utterances tend to be equalized in dB-SPL rather than in dB-A.

[2]In a pretest, five other participants (mean age 26 ± 2.1 years, 2 women) evaluated the phonological similarities of the pseudo-words to words. They had to judge if the pronounced pseudo-words sounded like a word, and if this was the case, they had to write down the corresponding word. All pseudo-words for which words have been indicated by at least two participants were eliminated. In a second part of the pretest, the participants judged the strength of semantic associations of pairs of words on a 5-point scale (from 0 = no association to 5 = very strong association). Word pairs with scores inferior to 2 were used in the second task of the Experiment 2 (i.e., recognition task) as distractors.

diagram of the sound processing method). This method is similar, at least in its principles, to other methods successfully used in neuroimaging studies (e.g., Davis and Johnsrude, 2003; Giraud et al., 2004). First, starting from a word or a pseudo-word, the overall phase spectrum was randomized while the overall magnitude spectrum of the phonological stimulus was preserved. Second, the slow temporal envelope (below 60 Hz) of the phonological stimulus was applied on the resulting signal. Consequently, the onsets and offsets of the complex sounds were matched to the original stimuli. This transformation preserved the average spectral content and preserved the slow time course of the amplitude. The excitation patterns (Moore and Glasberg, 1987) evoked by the original and transformed stimuli were almost identical. However, due to the phase-spectrum randomization, these stimuli sounded like different variations of noise, so they were not recognized as speech. In addition, although the phase randomization could alter the pitch strength of these sounds, the temporal envelope restoration resulted in a pitch strength equivalent for all three types of stimulus[3]. A diagram of the algorithm and sound samples of all stimuli categories are available in **Supplementary Materials S2 and S1**, respectively. These complex sounds were used rather than temporally reversed speech to avoid preserving phonological characteristics, with the goal to provide a stronger contrast with speech material. Speech segments that are steady-state, like vowels, are largely unaffected by time reversal. As a consequence, reversed speech is generally identified as speech, whereas the complex sounds used in the present study were not. In summary, these complex sounds had the same overall energetic properties as the words and pseudo-words – i.e., same average spectrum, slow temporal variations, duration, and periodicity – while not being recognized as speech.

Using the original (non-weighted) spectrum, the stimuli were equalized in dB-A, i.e., using the A-weighting, and in dB-SPL, i.e., using flat-weighting. Five levels of presentation were used (from inaudible to audible, with 5 dB steps).

Apparatus

Words and pseudo-words were recorded (32 bits, 44.1 kHz) using a Røde NT1 microphone, a Behringer Ultragain preamplifier, and a VxPocket V2 Digigram soundcard. The mean level of presentation was calibrated (ANSI, 1995) to reach 80 dB-A in a standard artificial ear (Larson Davis AEC101 and 824). The stimuli equalized with the flat-weighting scheme produced a level of presentation of 84 dB-SPL. Because dB-A and dB-SPL are different in nature, stimuli that have the same dB-A value may not have the same dB-SPL value and vice-versa. Therefore, only the average level over all stimuli could be compared across schemes: the stimuli equalized in dB-A had an average level of 83.3 dB-SPL. All stimuli were presented with the software Presentation 9.7 (Neurobehavioral Systems, Inc.,) using a soundcard (Creative Sound Blaster Audigy 2) followed by an analog attenuator (TDT PA4, one for each channel) that applied a fixed 40 dB attenuation. This attenuation was analog rather than digital to prevent acoustic distortion at low levels of presentation. All stimuli were binaurally presented to the participants through headphones (Sennheiser HD 250 Linear II) connected to a headphone buffer (TDT HB6).

Design and procedure

For each participant, 396 stimuli were randomly presented. The stimuli were words, pseudo-words, or complex sounds in 30.3% of the cases each. In 9.1% of the cases, there was no stimulus (i.e., a silence). The auditory stimuli were digitally attenuated to randomly obtain one of five selected levels of sound presentation: from 0 to +20 dB-A or to +4 to +24 dB-SPL with steps of 5 dB. A total of 120 words, 120 pseudo-words, and 120 complex sounds were presented in random order (resulting in 12 words, 12 pseudo-words, and 12 complex sounds per level and equalization or 24 words, 24 pseudo-words, and 24 complex sounds per level) together with 36 silences. The participants were told that the stimulus was sometimes replaced by a silence. Each stimulus, i.e., each word, pseudo-word, or complex sound, was presented only once to a participant, so each received only one presentation level and in only one equalization scheme (dB-A or dB-SPL). For example, for the same participant, if a given word was presented at level 0 in dB-A, it was not presented at any other level (in either equalization scheme) and was not presented at level 0 in dB-SPL either. Across all participants, each stimulus was presented at all levels and with all equalization schemes. The order of the stimuli was randomized for each participant.

After the stimulus presentation, the participants had to decide if they had detected an auditory item in a detection task by pressing *yes* or *no* answer keys, whose positions were counterbalanced across participants. The next stimulus occurred 500 ms after the response of the participants. A fixation cross appeared 100–500 ms before the presentation of the stimulus and remained until its end. The participants heard three blocks of 132 stimuli and short breaks were imposed between the three blocks. The duration of Experiment 1 was approximately 25 min.

Statistical analysis

From the proportion of "yes" responses Pr(yes), measures of detectability (d'_D) and criterion (k) were calculated for each participant as defined by the signal detection theory (SDT; Macmillan and Creelman, 2005). The constraints of the experimental design imposed that only one false-alarm (FA) rate was collected for each participant, i.e., common to all types of stimulus and all stimulus levels. Therefore, the criterion k calculated in this study was equal to $-z(FA)$, which represents the overall response bias (Macmillan and Creelman, 2005, p. 116). d'_D was analyzed with a three-way analysis of variance (ANOVA) with Type of Stimulus (words vs. pseudo-words vs. complex sounds), Stimulus Level (five levels from 0 to +20 dB-A and from +4 to +24 dB-SPL, with 5 dB steps), and Equalization (dB-A vs. dB-SPL) as within-participant factors. Separate ANOVAs on d'_D for the two Equalization schemes were also calculated with Type of Stimulus (words/pseudo-words/complex sounds) and Stimulus Level (five levels, from 0 to +20 dB-A and from +4 to +24 dB-SPL, with 5 dB steps) as within-participant factors. Because the d'_D measure represents the detectability of a stimulus by accounting for the participants' tendency to respond "yes" or "no," the performance could be separated into two measures: the *absolute*

[3]The pitch strength estimated based on a method similar to Ives and Patterson (2008) did not show any significant difference between the types of stimulus [$F(2,304) = 1.66$, $p = 0.19$].

sensitivity and the *absolute detectability*. The *absolute sensitivity* is captured by the slope of the psychometric function and is an unbiased measure of sensitivity. The *absolute detectability* is captured by the horizontal position of the psychometric function (defined by the abscissa of the point of the curve yielding 50% detection) and represents the detection bias related to each type of stimulus. These measures were estimated by fitting a cumulative Gaussian on the percent-correct detection data, using the maximum-likelihood method (as suggested in Macmillan and Creelman, 2005). These two measures were analyzed by two ANOVAs with Type of Stimulus as within-participant factor. An alpha level of 0.05 was used after Greenhouse–Geisser correction for all statistical tests.

The differences between stimulus types could occur on the slope of a psychometric curve, but not at the lowest and highest presentation levels where the performance of all stimuli was minimal (0–5%) or maximal (95–100%; i.e., floor and ceiling effects). For our data, a ceiling effect appeared for levels 4 and 5 ($p = 0.97$). This suggested that the performance between the three types of stimulus could only differ between levels 1 and 3, i.e., on the slope of the psychometric curve. Based on this *a priori* hypothesis, planned comparisons (local contrasts) were performed on these levels for significant interaction between Stimulus Level and Type of Stimulus or Equalization. For significant interaction between Type of Stimulus and Equalization, *post hoc* two-tailed paired *t*-tests, with Tukey correction (Howell, 1998), were performed.

Finally, we investigated the correlation between the word detection performance and their occurrence frequencies. These frequencies were calculated separately for the film and for the book corpus as the cumulative frequency over all homophones within each corpus (Lexique database, New et al., 2004).

RESULTS

Overall percentages of correct responses to silence (i.e., correct rejections) were high (mean = 96.11%, SD = 0.74%).

For d'_D analysis (see **Figure 1**), the main effect of Type of Stimulus [$F(2,38) = 18.17$, $p < 0.001$], the main effect of Stimulus Level [$F(4,76) = 218.15$, $p < 0.001$], and the interaction between Stimulus Level and Type of Stimulus [$F(8,152) = 2.67$, $p = 0.009$] were significant. For levels 1–3 (i.e., 0 to +10 dB-A or +4 to +14 dB-SPL), the planned comparisons revealed that words were detected better than complex sounds [$F(1,19) = 21.43$, $p < 0.001$] and pseudo-words were detected better than complex sounds [$F(1,19) = 26.79$, $p < 0.001$]. However, no difference was observed between words and pseudo-words [$F(1,19) = 0.7$, $p = 0.41$]. The ANOVA also revealed a significant main effect of Equalization [$F(1,19) = 7.41$, $p = 0.014$] and a significant interaction between Stimulus Level and Equalization [$F(4,76) = 5.43$, $p < 0.001$], as confirmed by the analyses on *absolute sensitivity* and *detectability* presented below. The planned comparisons revealed that the participants obtained better detection performance [$F(1,19) = 16.70$, $p < 0.001$] for dB-SPL than for dB-A at level 1 only (0 dB-A or +4 dB-SPL). In addition, the interaction between Equalization and Type of Stimulus was significant [$F(2,38) = 3.49$, $p = 0.04$]. *Post hoc* analysis revealed that the participants obtained better detection performance for dB-SPL than for dB-A for complex sounds only ($p = 0.03$), but not for words ($p = 0.07$), or pseudo-words ($p = 0.99$). Finally, the interaction between Equalization, Type of Stimulus and Stimulus Level was not significant [$F(8,152) = 1.36$, $p = 0.22$]. The criterion k was equal to 1.81 (SD = 0.32).

Separate ANOVAs on d'_D for the two Equalization schemes revealed a significant effect of Type of Stimulus in both schemes [$F(2,38) = 13.21$, $p < 0.001$ for dB-A and $F(2,38) = 10.21$, $p < 0.001$ for dB-SPL]. *Post hoc* revealed that both words and pseudo-words were better detected than complex sounds ($p < 0.001$ and $p = 0.002$ for dB-A and dB-SPL, respectively). The effect of Stimulus Level was also significant [$F(4,76) = 205.32$, $p < 0.001$ for dB-A; $F(4,76) = 177.36$, $p < 0.001$ for dB-SPL] but not its interaction with Type of Stimulus [$F(8,152) = 2.05$, $p = 0.08$ for dB-A; $F(8,152) = 1.90$, $p = 0.10$ for dB-SPL].

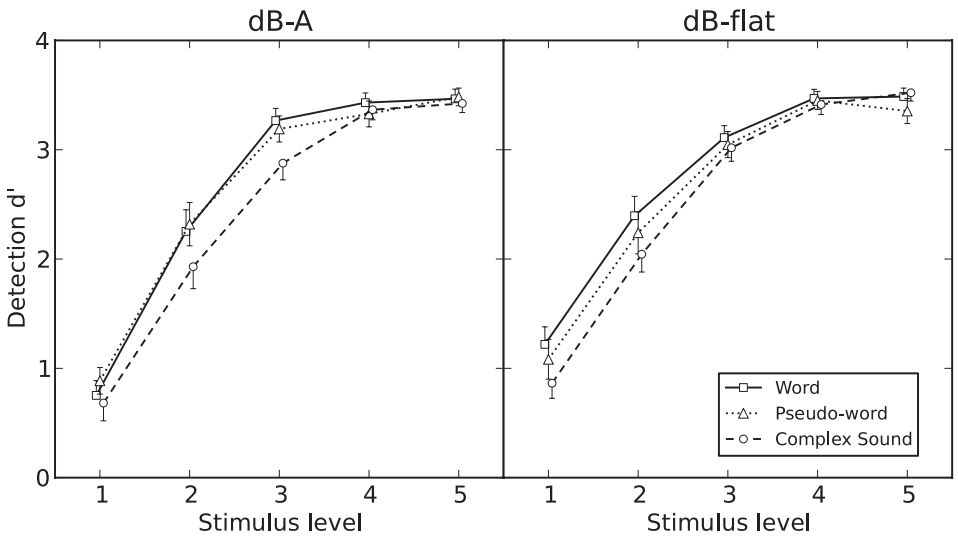

FIGURE 1 | Detection performance in Experiment 1. *Left panel*: d'_D presented for dB-A Equalization as a function of Stimulus Level (1–5; 0 to +20 dB-A) and Type of Stimulus (word/pseudo-word/complex sound). The error bars are the SE. *Right panel*: d'_D presented for dB-SPL Equalization as a function of Stimulus Level (1–5; +4 to +24 dB-SPL) and Type of Stimulus (word/pseudo-word/complex sound).

For *absolute sensitivity* (related to the slope of the psychometric function), the ANOVA revealed only a significant main effect of Equalization [$F(1,19) = 6.25$, $p = 0.022$]: the absolute sensitivity was on average 1.6 times larger for dB-A than for dB-flat. No effect of Type of Stimulus was found [$F(2,38) = 1.05$, $p = 0.35$]. For the *absolute detectability* (related to the horizontal position of the psychometric function), the ANOVA revealed a significant main effect of Type of Stimulus [$F(2,38) = 5.17$, $p = 0.01$]. The speech stimuli (words and pseudo-words) were both better detected than the complex sounds (both $p < 0.03$). A significant main effect of Equalization [$F(1,19) = 4.80$, $p = 0.04$] was also observed for *absolute detectability*: the stimuli equalized in dB-SPL were better detected than those equalized in dB-A. No interaction was observed between Equalization and Type of Stimulus, for the *absolute sensitivity* [$F(2,38) = 0.12$, $p = 0.89$] or the *absolute detectability* [$F(2,38) = 0.18$, $p = 0.84$].

No significant correlation was observed between the word detection performance and word occurrence frequencies ($r = -0.09$, $p = 0.71$ in films; $r = -0.07$, $p = 0.77$ in books), probably because of the limited variability of these word properties in the present material.

DISCUSSION

Experiment 1 showed that, near the auditory threshold (between 0 and +10 dB-A or between +4 and +14 dB-SPL), the detection performance was better for speech stimuli (words and pseudo-words) than for non-speech stimuli (complex sounds). This result suggests that, when auditory stimuli were difficult to detect, the listeners' knowledge facilitated the detection of phonological sounds over meaningless non-phonological sounds. This reveals a SDE in the auditory modality. It is important to note that as words and pseudo-words were both pronounced by a natural human voice, whereas the complex sounds were synthetic constructions (based on the same recordings), the SDE observed in this experiment could also be explained by a Voice Detection Effect. This point is further discussed in the Section "General Discussion."

The effect of equalization on d'_D and *absolute detectability* indicates that the stimuli equalized in dB-SPL were perceived louder than those equalized in dB-A which was consistent with the 0.7-dB shift described in the Section "Apparatus." The greater *absolute sensitivity* observed for dB-A than for dB-SPL equalization suggests that the dispersion of loudness across the stimuli was narrower in dB-A than in dB-SPL equalization. As expected, the dB-A equalization proved to be better adapted to equalize the level of complex stimuli, such as speech. Nevertheless, most importantly for the goal of our study, there was no interaction between the equalization types and stimulus types on *absolute sensitivity* and *absolute detectability*. As also confirmed by the two separate analyses, the effect of stimulus type was observed for both equalization schemes, supporting the consistency of the SDE. These findings indicate that the effect of knowledge on detection does not depend on the equalization method, i.e., it was not influenced by natural variations in loudness, thus attesting to the robustness of this effect.

Although the phonological content seemed to improve the detection (in comparison with the complex sounds), we did not observe a WDE as previously reported in the visual modality (Merikle and Reingold, 1990). In Merikle and Reingold (1990, Experiment 4), the detection task was immediately followed by a recognition task in each experimental trial. This task could have engaged the participants in lexical processing even during the detection task, stressing the importance of differences between words and non-words. To investigate a potential WDE in the auditory modality, a recognition task was added following the detection task in Experiment 2.

EXPERIMENT 2

In Experiment 2, the procedure of Experiment 1 was modified for three reasons. First, to investigate the effect of stimulus type and in particular, to focus on a potential detection advantage of words over pseudo-words, the detection task was followed by a 2AFC recognition task (as in Merikle and Reingold, 1990, for the visual modality). The 2AFC recognition task allowed us to investigate a potential WSE and/or speech superiority effect (SSE) in the auditory modality. Whereas the WSE and PSWE for visual stimuli have been reported previously (for example, Grainger and Jacobs, 1994), to our knowledge, no study has investigated these effects for sounds. Second, to investigate the psychometric curves from none to complete detection with a better precision, we used a larger range of presentation levels and steps of 3 dB (smaller than the 5 dB steps of Experiment 1). In addition, as Experiment 1 showed that the dB-A equalization minimized the detection variance across stimuli, i.e., that it was better adapted to equalize the physical energy of complex stimuli such as speech, only the dB-A equalization was used in Experiment 2.

The 2AFC recognition task also allowed us to investigate the dissociation between auditory detection and recognition. Whereas the dissociation between detection and higher level processing has been observed for the visual modality using different experimental designs and methods (e.g., Reingold and Merikle, 1988; Merikle and Reingold, 1990; Dehaene et al., 1998; Naccache and Dehaene, 2001), only a few studies have investigated these effects for the auditory modality. Using masking paradigms, Shipley (1965) did not observe any dissociation between detection and recognition for tones, whereas Lindner (1968) did observe this dissociation when indicating to the participants that recognition was possible even without detection. Using time-compressed and masked primes, Kouider and Dupoux (2005) suggested dissociation between categorization and semantic processing for speech sounds. They observed repetition priming (but no phonological or semantic priming) while the participants were unable to categorize the prime as word or non-word (but they were probably able to detect the presence of the prime). Thus, to our knowledge, no study has previously investigated a potential dissociation between detection and recognition of speech and non-speech stimuli in the auditory modality.

MATERIAL AND METHODS

Participants

Nineteen students (mean age 21.2 ± 2.1 years, 14 females) participated in Experiment 2. They were selected with the same criteria as described in Experiment 1.

Materials

A subset of 108 stimuli from Experiment 1 was added to another set of stimuli with the same properties to form a larger set of 462 stimuli. In this set, the average duration of words,

pseudo-words, and complex sounds were not significantly different [$F(2,459) = 2.58$, $p = 0.08$] and were 521.5 ms (SD = 115.5 ms), 539.2 ms (SD = 87.8 ms), and 546.6 ms (SD = 92.1 ms), respectively. To reduce the differences in the perceived loudness, all stimuli were equalized to the same dB-A level.

Apparatus

The same apparatus as in Experiment 1 was used. The mean level of presentation of the stimuli equalized in dB-A was calibrated (ANSI, 1995) to reach 80 dB-A in a standard artificial ear (Larson Davis AEC101 and 824).

Design and procedure

For each participant, 504 trials were presented in random order using Presentation 9.7. Within each trial, three stimuli from the same category were presented (see **Figure 2**).

The first stimulus was a word, a pseudo-word, or a complex sound in 30.6% of trials respectively, and in 8.2% of trials there was no stimulus (i.e., a silence). A digital attenuation was randomly applied to the first stimulus from 15 to 45 dB by steps of 3 dB to reach 11 levels of presentation from −5 to +25 dB-A. A total of 154 words, 154 pseudo-words, and 154 complex sounds were presented in random order (resulting in 14 words, 14 pseudo-words, and 14 complex sounds per level) along with 42 silences. The participants were told that the stimulus was sometimes replaced by a silence. Each stimulus was presented only once to a participant (i.e., at one given level of presentation). Across participants, each stimulus was presented at a different presentation level. As in Experiment 1, each stimulus was presented only once to a participant, but across participants, each stimulus was presented at all presentation levels. The order of the stimuli was randomized between participants. The participants had to decide whether they detected an auditory stimulus (detection task) by pressing *yes* or *no* answer keys, whose position was counterbalanced across participants.

Two hundred milliseconds after the response to the detection task, the second and third stimulus were presented at an audible level (+40 dB-A), with the third stimuli occurring 200 ms after the second one. One of the two stimuli was the same as the detection stimulus (repetition relationship) and was randomly and equally presented in the first or second interval over stimuli and participants. The other stimulus was a distractor of the same category (154 words, 154 pseudo-words, and 154 complex sounds), which was not presented in the detection task and which appeared only once in the experimental session. For words and pseudo-words, the distractor was neither phonologically nor semantically related to the first stimulus [as evaluated by a pretest (see text footnote 1)]. The number of letters and phonemes was the same for the items within each pair of words or pseudo-words. For the trials where the first stimulus was a silence, a pair of words, pseudo-words, or complex sounds was randomly selected (in total, 14 pairs of words, 14 pairs of pseudo-words, and 14 pairs of complex sounds were presented). After having listened to the pair of stimuli, the participants had to choose whether the stimulus presented in the detection task was similar to the first or to the second stimulus of the recognition pair (2AFC task). They were asked to respond as quickly as possible (but no timeout was imposed) even if they had indicated that they had not heard anything in the detection task. The next trial appeared 1000 ms after the participants' response. A visual fixation cross appeared 100–500 ms before the onset of the first stimulus and remained on the screen until its offset. The participants heard six blocks of 84 trials in a randomized order. Short breaks were imposed between the six blocks. The duration of Experiment 2 was approximately 1 h.

Statistical analysis

An alpha level of 0.05 after Greenhouse–Geisser correction was used for all statistical tests.

Detection task. Similar to Experiment 1, d'_D, *absolute detectability* and *absolute sensitivity* were analyzed with a two-way ANOVAs with Type of Stimulus (words/pseudo-words/complex sounds) and Stimulus Level (11 levels from 1 to 11, i.e., from −5 to +25 dB-A with 3 dB steps) as within-participant factors. A floor effect was observed between levels 1 and 3 (no significant differences were observed between levels 1 and 2, 2 and 3, 1 and 3, all $p > 0.07$), and a ceiling effect appeared between levels 9 and 11 (no significant differences were observed between 9 and 10, 10 and 11, 9 and 11, all $p > 0.37$). This suggested that the performance between the three types of stimulus could only differ between levels 4 and 8, i.e., on the slope of the psychometric curve.

Recognition task. Performance (percentage of correct recognitions) was analyzed with a two-way ANOVA with Type of Stimulus (words/pseudo-words/complex sounds) and Stimulus Level (11 levels numbered 1–11, ranging from −5 to +25 dB-A with 3 dB steps) as within-participant factors.

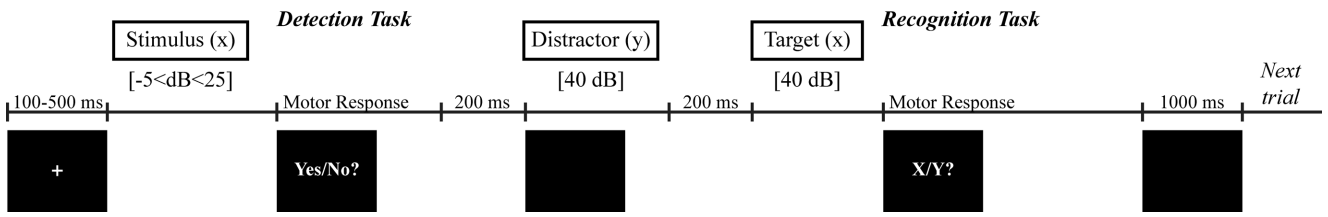

FIGURE 2 | Time course of a trial in Experiment 2. After displaying a fixation cross, a stimulus (word, pseudo-word, or complex sound) or a silence was presented (from –5 to 25 dB-A) and the participants had to perform a detection task by pressing yes or no answer keys. At 200 ms after the response, two stimuli separated by 200 ms were presented at audible level (one being the same as the stimulus presented in the detection task – the target – and the other being a distractor of the same category, presented in random order). The participants had to choose in a 2AFC recognition task the stimulus that was the same as the stimulus presented in the detection task by pressing one of two answer keys (first or second stimulus).

Recognition without detection. As in previous studies using visual materials (Haase and Fisk, 2004; Holender and Duscherer, 2004; Reingold, 2004; Snodgrass et al., 2004a,b; Fisk and Haase, 2005), dissociation between detection and recognition was analyzed using a subjective threshold approach (e.g., Merikle and Cheesman, 1986) and an objective threshold approach (e.g., Greenwald et al., 1995). The subjective threshold approach supposes a dissociation between detection and recognition such that under stimulus conditions where the participants do not report awareness of the stimuli, they can nevertheless perform above chance on the perceptual discrimination tasks (e.g., Cheesman and Merikle, 1984, 1986; Merikle and Cheesman, 1986). The subjective threshold approach tests whether correct recognition performance exceeded the chance level (0.50). Single sample *t*-tests (two-tailed) were used to test whether recognition performance was above chance level for missed stimuli. These tests were performed at the stimulus level where each participant reached maximum recognition (over missed and detected stimuli) while having a minimum of 15% of misses in the detection task (as in Fisk and Haase, 2005), i.e., for level 5. The objective threshold approach is based on an index of sensitivity (d'_D) on the awareness variable (i.e., performance is at chance with a direct measure of detection) that is used as an indicator of null awareness (e.g., Snodgrass et al., 1993; Greenwald et al., 1995). Recognition was modeled using methods that were based on the SDT (Macmillan and Creelman, 2005) as described by Greenwald et al. (1995). Recognition sensitivity was expressed as d'_R, calculated with the $\sqrt{2}$ correction because two response choices were possible (Macmillan and Creelman, 2005). The d'_D and d'_R variables were compared at each stimulus level using paired *t*-tests (two-tailed): d'_D was always significantly greater than d'_R ($p > 0.05$). To assess the possibility of recognition without detection, the value of d'_R when d'_D was close to zero needed to be evaluated. Since the distribution of individual d'_D was not centered on zero, the value of d'_R when $d'_D = 0$ was extrapolated using a linear regression as indicated by Greenwald et al. (1995).

RESULTS

Detection task

The overall percentages of correct responses for silence (i.e., correct rejections) were high (mean = 97.62%, SD = 0.70%).

For the analysis of d'_D (see **Figure 3**), the main effect of Type of Stimulus [$F(2,36) = 29.93$, $p < 0.001$], the main effect of Stimulus Level [$F(10,180) = 1817.3$, $p < 0.001$] and the interaction between these two factors [$F(20,360) = 2.26$, $p < 0.001$] were significant. For levels 4–8 (i.e., from +4 to +16 dB-A), the planned comparisons revealed that words were better detected than pseudo-words [$F(1,18) = 7.01$, $p = 0.016$] and complex sounds [$F(1,18) = 59.80$, $p < 0.001$], and pseudo-words were better detected than complex sounds [$F(1,18) = 17.28$, $p < 0.001$]. On average, the criterion *k* was equal to 1.99 (SD = 0.32). When compared to the criterion observed in Experiment 1, a *t*-test (two-tailed) revealed that the difference was not significant [$t(37) = 1.81$, $p = 0.08$].

For the *absolute sensitivity*, this analysis revealed no significant effect [$F(2,36) = 0.51$, $p = 0.61$] whereas for the *absolute detectability*, the effect of Type of Stimulus was significant [$F(2,36) = 24.76$, $p < 0.001$]: words and pseudo-words yielded better *absolute detectability* than complex sounds (both $p < 0.001$). See details in **Table 2**, **Supplementary Material S3**.

As in Experiment 1, no significant correlation was observed between the word detection performance and word occurrence frequencies ($r = -0.19$, $p = 0.43$ for films; $r = -0.30$, $p = 0.20$ for books).

Recognition task

The ANOVA showed a significant main effect of Stimulus Level on recognition performance [$F(10,180) = 110.77$, $p < 0.001$]: correct recognition was more likely for high than for low levels (**Table 2**, **Supplementary Material S3**). There was no significant effect of Type of Stimulus [$F(2,36) = 0.95$, $p = 0.40$] and no significant interaction between Stimulus Level and Type of Stimulus [$F(20,360) = 1.56$, $p = 0.06$]. No significant correlation was observed between detection and recognition performance [$r(17) = 0.004$; $p = 0.96$ for

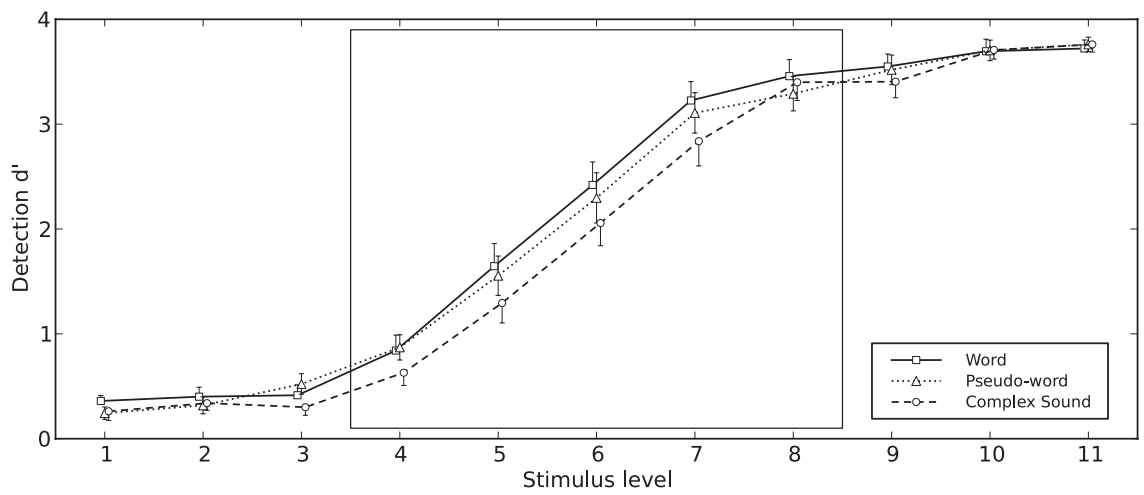

FIGURE 3 | Detection performance in Experiment 2. d'_D presented as a function of Stimulus Level (1–11, i.e., **–**5 to **+**25 dB-A) for each Type of Stimulus (word/pseudo-word/complex sound). The error bars show the SE.

words; $r(17) = 0.014$; $p = 0.86$ for pseudo-words; $r(17) = 0.020$; $p = 0.80$ for complex sounds]. Moreover, the target position in the recognition interval did not influence recognition for the three sets of stimuli [$t(18) = 0.67$, $p = 0.51$ for words; $t(18) = 0.51$, $p = 0.62$ for pseudo-words; $t(18) = 0.47$, $p = 0.64$ for complex sounds].

Recognition without detection

For the subjective approach, the analysis was conducted at level 5 (+7 dB-A), by analyzing recognition scores of trials where the participants had responded "no" in the detection task. At the level 5, the recognition performance was significantly above chance [$t(18) = 2.29$, $p < 0.05$]. The percentage of correct recognition was above chance for words [$t(16) = 2.65$, $p = 0.017$], but not for pseudo-words [$t(18) = 0.49$, $p = 0.62$], or complex sounds [$t(18) = 1.69$, $p = 0.11$].

For the objective threshold approach (**Figure 4**), the analysis was conducted for levels where individual d'_D were distributed near zero (i.e., levels 1–3) as the contribution of better detected conditions would bias the regression (Miller, 2000). A vertical-intercept greater than zero indicates recognition without detection and a slope greater than zero indicates a correlation between recognition and detection performance. The intercepts from the regression lines were never significantly greater than zero for words, pseudo-words and complex sounds [$y = -0.004$, $t(51) = -0.041$; $p > 0.97$; $y = 0.064$, $t(55) = 0.586$; $p > 0.56$; $y = -0.010$, $t(55) = -1.249$; $p > 0.22$, respectively]. The slopes of the regressions were never significantly greater than zero for words, pseudo-words and complex sounds [$x = -0.045$, $t(51) = -0.237$; $p > 0.81$; $x = 0.070$, $t(55) = 0.333$; $p > 0.74$; $x = 0.172$, $t(55) = 0.941$; $p > 0.35$, respectively].

DISCUSSION

Speech and word detection effects

Experiment 2 confirmed the main result of Experiment 1: close to the auditory threshold, phonological stimuli (words and pseudo-words) were better detected than non-phonological stimuli (complex sounds). Furthermore, Experiment 2 showed better detection of meaningful phonological stimuli (words) over meaningless ones (pseudo-words and complex sounds). The addition of the recognition task after the detection task allowed to reveal this WSE in the auditory modality, strengthening the result of Merikle and Reingold (1990) who also used this succession of the two tasks in the visual modality. However, this latter effect was not as strong as the SDE (the difference between average d'_D was 0.10 between words and pseudo-words whereas it was 0.23 between speech and non-speech). Overall, these findings suggest that when auditory stimuli were difficult to detect, (1) phonological knowledge facilitated the detection of stimuli, i.e., suggesting the existence of a SDE in the auditory modality; and further showed that (2) lexical knowledge facilitated auditory detection, i.e., a WDE, when participants are engaged in tasks requiring lexical processing.

Word superiority effect

In the 2AFC recognition task (second task), the data of Experiment 2 did not show a WSE, as previously observed for written letter recognition (e.g., Cattell, 1886; Reicher, 1969), or a SSE. At first surprising, the absence of the WSE and SSE might be explained by the specificity of our experimental paradigm. In contrast to studies that demonstrated a WSE, the recognition task of our study was both time-delayed and disrupted by another task (the detection task). The temporal decay of the WSE and its sensitivity to task interference has not been previously studied, but it has been shown that linguistic facilitations are time-limited (e.g., Neely, 1977), and so could be the SSE. Further studies without an interfering detection task would be necessary to investigate the presence or absence of WSE and/or SSE in the 2AFC recognition task.

Dissociation between detection and recognition

The present study is the first to investigate a potential dissociation between detection and recognition of speech and non-speech stimuli. The subjective approach suggests that participants could

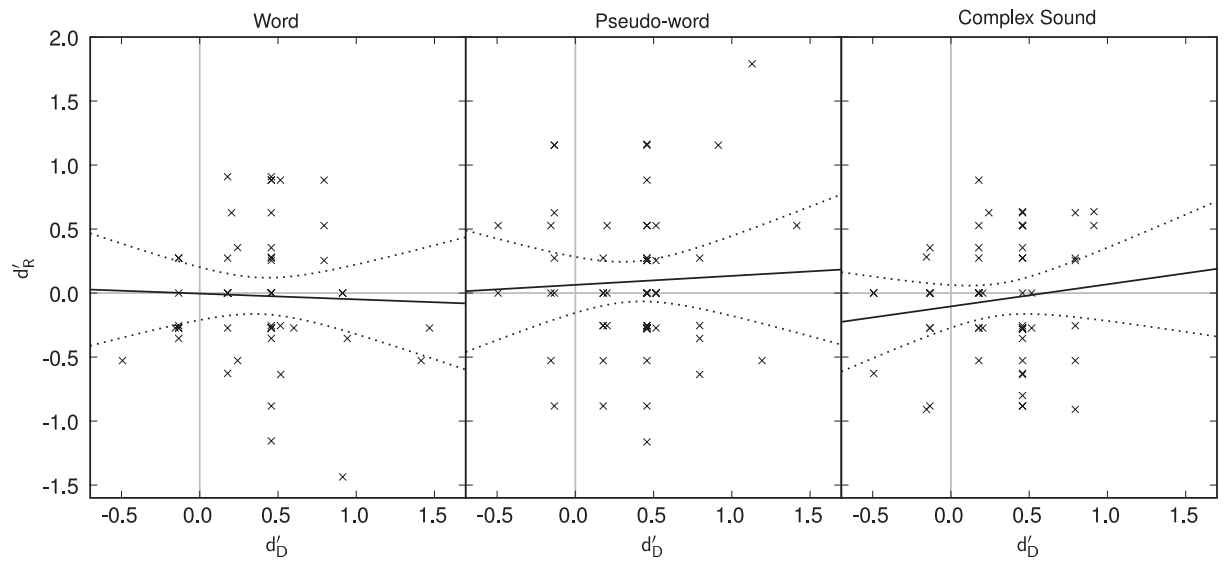

FIGURE 4 | Linear regression between d'_R and d'_D for each Type of Stimulus. The linear regression is the black line and the confidence interval is represented by the dotted lines.

recognize auditory stimuli even when they had previously said they could not detect any sound in the trial. This was demonstrated by recognition performance above chance level for words at 7 dB-A (level 5 in Experiment 2). Previously, Merikle and Reingold (1990, Experiments 2 and 3) observed a dissociation between detection and recognition for visual stimuli using the subjective threshold approach. When participants did not detect the stimulus, the words were recognized, but non-words were not. The authors interpreted this finding as unconscious processing and concluded that only familiar stimuli could be perceived unconsciously. Our study further suggests that (1) the recognition of familiar stimuli without subjective detection could also be observed in the auditory modality and (2) unfamiliar stimuli could not be "unconsciously perceived" even when they have a phonetic structure. Future experiments should investigate whether participants are able to perform lexical decision without subjective auditory detection, similar to what has been previously reported in the visual modality by Merikle and Reingold (1990). In their study, the participants performed a detection task immediately followed by a categorization task. Using this experimental design, the retrieval environment was not disrupted by the presentation of two stimulus alternatives. If unconscious perception was a domain-general phenomenon, we would predict that comparable patterns of results should be found with a lexical decision task in the auditory modality, as has been previously reported in the visual modality (Merikle and Reingold, 1990, Experiment 4).

The objective approach derived from the SDT did not suggest dissociation between detection and recognition at lower levels of detection (from −5 to 1 dB-A). The recognition performance was never greater than the detection performance, suggesting that decisions for detection did not necessarily entail recognition. As dissociation was observed with the subjective approach for a level where the average detection was greater than zero, one may suggest that participants need at least some degree of stimulus awareness to perform a correct recognition response following an absence of stimulus detection. Merikle and Reingold (1990) have argued that the qualitative difference between the "detect" (words and non-words recognition was observed after stimulus detection) and the "non-detect" state (only word recognition was observed after an absence of stimulus detection) supports the validity of the subjective measure of conscious awareness (Sandberg et al., 2010). This would suggest that auditory dissociation could be observed only at the subjective detection threshold, as was previously reported for the visual modality (Cheesman and Merikle, 1984, 1986).

Context effects

Te detection performance differed for a given level in Experiments 1 and 2. For example, at 0 dB-A, the percentage of detection for stimuli equalized in dB-A was approximately 15% in Experiment 1 but approximately 5% in Experiment 2. This effect was probably due to differences in the ranges of stimulus levels used in the two experiments: Experiment 1 restricted the investigation to a smaller range of stimulus levels than did Experiment 2, leading to smaller contrasts between the highest and the lowest levels in Experiment 1. The increased ability to detect sounds in a small range of levels was consistent with the study of Luce and Green (1978) that showed better detection performance when the range spanned over 10 to 20 dB than when the range was larger (more than 20 dB, as in Experiment 2 of our study).

SIMULATING DETECTION USING AUDITORY MODELS

To further assess the potential role of acoustic cues for detection, auditory models can be used to predict the detection scores for the stimuli used in the present study. The proposed simulations were based on the Signal Detection Theory: detection performance is driven by a continuous internal variable *via* a decision rule. The internal variable is related to loudness (or detectability) and is directly driven by the physical properties of the stimulus, independent of the listener's behavior. Two different auditory models were used to obtain internal variables (see **Supplementary Material S4** for a detailed description). The outputs of these models were then converted into detection scores using a decision model (also detailed in **Supplementary Material S4**).

AUDITORY MODELS

Time-varying loudness model

For a large variety of sounds, including complex sounds, detection can be directly related to loudness (Moore et al., 1997; Buus et al., 1998). Kewley-Port (1991) successfully predicted detection thresholds for stationary isolated vowels using the loudness model proposed by Moore et al. (1997). In loudness models, the estimated loudness is derived from the internal excitation pattern that can be viewed as the repartition of energy in the cochlea. This excitation pattern can then be transformed into a *specific loudness* pattern that accounts for thresholds effects and for the compressive nature of loudness: a large increase in intensity evokes a smaller increase in loudness at high intensities than at lower ones. The specific loudness is integrated over frequencies to yield a single loudness level value (Moore et al., 1997). For time-varying signals, such as speech, this instantaneous loudness also needs to be integrated over time. In their time-varying loudness (TVL) model, Glasberg and Moore (2002) proposed temporal integration functions that successfully mimic behavioral loudness judgments for a variety of time-varying sounds (Rennies et al., 2010). The model produces a short-term loudness (with a time constant of about 20 ms) and a long-term loudness (with a time constant of about 100 ms). We assumed that only a short burst was necessary for detection and hence used the maximum of the short-term loudness over time as the internal variable for detection.

Auditory image model

The TVL model is based on cochlea representations of the sounds. Although the spectro-temporal excitation pattern is integrated both in time and frequency, which implies the existence of a slightly higher process overlooking this representation, there is no specific feature extraction. In particular, temporal regularity, such as the one that gives rise to a pitch percept is not accounted for. Another relative weakness of this model is that its sensitivity to phase differences throughout the spectrum is unrealistic. This results in the incapacity to properly exploit potential coincidences across frequencies. This is important because the algorithm used to generate the complex sound stimuli in the present study specifically manipulated the phase relationship between the frequency

components of the signal. Phase re-alignment and pitch extraction are believed to occur at later stages of the auditory processing and models for these processes have been proposed, such as the auditory image model (AIM; Patterson et al., 1992; Bleeck et al., 2004).

RESULTS

The simulated probabilities of observing a difference in d'_D equal to zero (two-tailed) are presented in **Table 1**. This probability is comparable to the *p*-value of a two-tailed *t*-test testing the null hypothesis "equal to zero." None of the models demonstrated any significant advantage for the speech stimuli. The AIM predicts an advantage for the complex sounds over the linguistic stimuli. A figure showing the simulated d'_D for the two models is provided in the **Supplementary Material S4**.

DISCUSSION

The TVL model predicts an advantage of words over pseudo-words, but this difference was not significant. This suggests that energetic differences could have partially contributed to the WDE without being sufficient to make it significant. Finding where these differences exactly originate from is beyond the scope of this article. However, one could venture that a potential explanation might lay in the fact that the pseudo-words were more likely to contain uncommon phonotactic arrangements than the words. This could have effects on the intrinsic acoustic structure of the pseudo-word, or on the way it was effectively pronounced by the speaker.

However, despite their complexity and recognized validity for a wide variety of sounds, both auditory models failed to predict the detection facilitation demonstrated by the participants for words and pseudo-words compared to the complex sounds. In particular neither the temporal regularity nor the spectro-temporal coincidence simulated in the AIM explained this facilitation effect. Both models predicted that the complex sounds would be better detected than the pseudo-words (and this difference reached significance for the AIM), which is opposite to the behavioral data. These results indicate that the previously described SDE could not be due to differences in the sound representations at the lower levels of the auditory system.

Table 1 | Statistical analyses of the simulated results using the two auditory models TVL and AIM.

Differences	*Behavioral data*	TVL	AIM
Word – Pseudo-word	*t(18) = 2.65, p = 0.016**	*p* = 0.173 (>0)	*p* = 0.9872 (<0)
Word – Complex sound	*t(18) = 7.73, p < 0.0001****	*p* = 0.808 (>0)	*p* = 0.010 (<0)
Pseudo-word – Complex sound	*t(18) = 4.16, p < 0.0001****	*p* = 0.269 (<0)	*p* = 0.009 (<0)

The behavioral data are presented in italics. For each comparison, the estimated probability of the "equal to zero" hypothesis to be true is reported with the sign of the difference between brackets. Negative differences are opposite to the behavioral effects.

GENERAL DISCUSSION

The main aim of our study was to investigate whether phonological and lexical knowledge could facilitate a task requiring only lower-level processing, such as auditory signal detection. Words, pseudo-words, and complex sounds were energetically matched as closely as possible and were presented from inaudible to audible levels. The participants performed a detection task (Experiments 1 and 2) that was followed by a 2AFC recognition task in Experiment 2. Experiments 1 and 2 showed a SDE: near the auditory threshold, phonological stimuli (words and pseudo-words) were better detected than non-phonological stimuli (complex sounds). In addition, in Experiment 2 where participants were also engaged in a second task (recognition task), phonological and meaningful stimuli (words) were better detected than phonological and meaningless stimuli (pseudo-words), i.e., we showed a WDE in the auditory modality. This suggests that the recognition task may have affected auditory detection, the second task encouraging a lexical processing during the detection task. To our knowledge, such cognitive facilitation effects on auditory detection have not been previously reported and further investigations should be conducted to specifically assess the WDE. Regarding the SDE, the observed differences might also be interpreted as a Voice Detection Effect (words/non-words vs. complex sounds). Interestingly, voice specific processes are also likely the result of long-term specialization and could be categorized as knowledge-based. Future research is needed to further estimate their contribution to the effects observed here.

The SDE and WDE did not appear as differences in the slope of the psychometric functions, characterizing the sensitivity, but as differences in the horizontal shift along the stimulus level axis. This means that the observed differences on d'_D were due to differences in absolute detectability. Under the hypothesis that the internal criterion was constant within the experimental session[4], a difference in absolute detectability suggests that the type of stimulus modulated the amount of internal noise and its effect on sensory representations. For our study, the difference in absolute detectability cannot be explained by systematic energetic differences between the items of the three stimulus types. The stimuli were carefully designed so that the energetic features matched as closely as possible between the three categories. Indeed, intensity, duration, temporal envelope, and spectrum, as well as phonemes for words and pseudo-words, were on average as similar as possible across the sets of stimuli. Moreover, models estimating the loudness of the stimuli on the basis of the energetic properties of the sounds showed that the loudness differences could not explain the difference in detectability between the speech and non-speech stimuli, and between words and pseudo-words. Even when short-term regularities in the sound were exploited (in the AIM), the auditory models (TVL, AIM) failed to reproduce the observed effects. The differences in detection performance could be explained by differences in processing at higher processing levels rather than at sensory processing levels. The influence of knowledge on auditory detection is in agreement with Merikle and Reingold (1990) who showed a WDE with a visual

[4]With this experimental design, it was not possible to obtain a value of the internal criterion for each type of stimuli because there was only one condition – the silence condition – to calculate the false-alarms.

subliminal detection paradigm. Our study extended their results to the auditory modality and suggests that linguistic knowledge could facilitate lower level tasks for both modalities. Our findings could be integrated into models that simulate processing facilitation with perceivers' knowledge. Two classes of models have been proposed to explain facilitation due to knowledge-related influences. The first class of word recognition models in the auditory modality suggests that the information carried by the word onset activates a group of candidate words with the influence of lexical and semantic contextual information arising only at a later decision stage, as in the MERGE model (McQueen et al., 1994; Norris et al., 2000). The second class of models explains lexical and phonological facilitations by top-down processing. McClelland and collaborators (McClelland and Elman, 1986; McClelland et al., 2006) explained that the lexical and semantic processes can influence lower-level acoustic and phonetic processes, by combining bottom-up information and top-down feedback from the lexical level down to the phonemic level. Samuel and colleagues (Samuel, 1997; Rapp and Samuel, 2002; Samuel and Pitt, 2003; Kraljic and Samuel, 2005) have shown that linguistic-based facilitation can affect the phonemic level of word processing. Our results are compatible with the TRACE and MERGE models, and extend these previous findings by suggesting that knowledge is automatically activated, even when no form of knowledge is required by the task. Indeed, the ability to detect more easily phonological stimuli, which might carry lexical information, than non-phonological stimuli might be important for human communication because it could be helpful to react quickly when a speech stimulus, which presents a social interest, emerges in our environment.

CONCLUSION

Our study investigated the detection of speech and non-speech sounds in the auditory modality. The observed influence of phonological knowledge on detection, as reflected in the SDE (and more weakly in the WDE), indicates that all levels of auditory processing, from early encoding in the auditory nerve to phonological and lexical processing, are involved in a detection task. This might be mediated by either top-down connections or long-term adaption of the bottom-up pathways. Moreover, data from Experiment 1 in which the task did not require any speech-specific processes showed that processes related to the SDE could be automatically engaged in the presence of speech stimuli. More generally, the results suggest that detection depends on the nature of the sound and specifically on the potential relevance of the stimulus one have to detect. Hence, this approach could be used to explore the relevant features that are unconsciously extracted from sounds to construct conscious percepts. In particular, this method could be used to identify the currently missing links between sensory low-level auditory models (e.g., Glasberg and Moore, 2002; Bleeck et al., 2004) and knowledge-based higher-level sound perception models (e.g., Kiebel et al., 2009). This would help to explain how higher-level tasks, such as recognition, are influenced by unconscious basic processing when the stimuli are familiar, as has been suggested in our study for subjective unconscious perception of words for the auditory modality and in the study of Merikle and Reingold (1990) for the visual modality. Future research will have to assess this effect to confirm that stimuli containing relevant information are more likely to be consciously perceived and if not perceived, are more likely to be unconsciously processed.

ACKNOWLEDGMENTS

The work described in this paper was supported by a doctoral grant from the "Ministère de l'Éducation Nationale et de la Recherche" of France related to the grant ACI "Junior Research Team," as well as by the grant ANR "PICS" and a grant of the UK-Medical Research Council (G9900369). The authors wish to thank Samuel Garcia for providing technical assistance, Neil A. Macmillan, Eyal M. Reingold, and Arthur G. Samuel for their helpful comments on this work, and Brian C. J. Moore for suggesting to apply his loudness model to the stimuli.

SUPPLEMENTARY MATERIAL

The Supplementary Materials S1 to S4 for this article can be found online at http://www.frontiersin.org/auditory_cognitive_neuroscience/10.3389/fpsyg.2011.00176/abstract

REFERENCES

ANSI. (1995). *S3.7-1995 (R2003), Methods for Coupler Calibration of Earphones*. New York: American National Standard Institute.

ANSI. (2004). *S3.21-2004, Methods for Manual Pure-Tone Threshold Audiometry*. New York: American National Standard Institute.

Baron, J., and Thurston, I. (1973). An analysis of the word-superiority effect. *Cogn. Psychol.* 4, 207–228.

Bleeck, S., Ives, T., and Patterson, R. D. (2004). Aim-mat: the auditory image model in MATLAB. *Acta Acustica* 90, 781–787.

Buus, S., Müsch, H., and Florentine, M. (1998). On loudness at threshold. *J. Acoust. Soc. Am.* 104, 399–410.

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind* 11, 63–65.

Cheesman, J., and Merikle, P. M. (1984). Priming with and without awareness. *Percept. Psychophys.* 36, 387–395.

Cheesman, J., and Merikle, P. M. (1986). Distinguishing conscious from unconscious perceptual processes. *Can. J. Psychol.* 40, 343–367.

Davis, M. H., and Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *J. Neurosci.* 23, 3423–3431.

Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., van de Moortele, P. F., and Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature* 395, 597–600.

Doyle, J., and Leach, C. (1988). Word superiority in signal detection: barely a glimpse, yet reading nonetheless. *Cogn. Psychol.* 20, 283–318.

Fisk, G. D., and Haase, S. J. (2005). Unconscious perception or not? An evaluation of detection and discrimination as indicators of awareness. *Am. J. Psychol.* 118, 183–212.

Giraud, A. L., Kell, C., Thierfelder, C., Sterzer, P., Russ, M. O., Preibisch, C., and Kleinschmidt, A. (2004). Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. *Cereb. Cortex* 14, 247–255.

Glasberg, B. R., and Moore, B. C. J. (2002). A model of loudness applicable to time-varying sounds. *J. Audio. Eng. Soc.* 50, 331–342.

Grainger, J., Bouttevin, S., Truc, C., Bastien, M., and Ziegler, J. (2003). Word superiority, pseudoword superiority, and learning to read: a comparison of dyslexic and normal readers. *Brain Lang.* 87, 432–440.

Grainger, J., and Jacobs, A. M. (1994). A dual read-out model of word context effects in letter perception: further investigations of the word superiority effect. *J. Exp. Psychol. Hum. Percept. Perform.* 20, 1158–1176.

Greenwald, A. G., Klinger, M. R., and Schuh, E. S. (1995). Activation by marginally perceptible ("subliminal") stimuli: dissociation of unconscious from conscious cognition. *J. Exp. Psychol. Gen.* 124, 22–42.

Haase, S. J., and Fisk, G. D. (2004). Valid distinctions between conscious and unconscious perception? *Percept. Psychophys.* 66, 868–871; discussion 888–895.

Hoen, M., Meunier, F., Grataloup, C., Pellegrino, F., Grimault, N., Perrin,

F., Perrot, X., and Collet, L. (2007). Phonetic and lexical interferences in informational masking during speech-in-speech comprehension. *Speech Commun.* 49, 905–916.

Holender, D., and Duscherer, K. (2004). Unconscious perception: the need for a paradigm shift. *Percept. Psychophys.* 66, 872–881; discussion 888–895.

Howell, D. (1998). *Méthodes Statistiques en Sciences Humaines*. Paris: De Boeck Université.

Ives, D. T., and Patterson, R. D. (2008). Pitch strength decreases as F0 and harmonic resolution increase in complex tones composed exclusively of high harmonics. *J. Acoust. Soc. Am.* 123, 2670–2679.

Kewley-Port, D. (1991). Detection thresholds for isolated vowels. *J. Acoust. Soc. Am.* 89, 820–829.

Kiebel, S. J., von Kriegstein, K., Daunizeau, J., and Friston, K. J. (2009). Recognizing sequences of sequences. *PLoS Comput. Biol.* 5, e1000464. doi: 10.1371/journal.pcbi.1000464

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *J. Acoust. Soc. Am.* 67, 971.

Kouider, S., and Dupoux, E. (2005). Subliminal speech priming. *Psychol. Sci.* 16, 617–625.

Kraljic, T., and Samuel, A. G. (2005). Perceptual learning for speech: is there a return to normal? *Cogn. Psychol.* 51, 141–178.

Lindner, W. A. (1968). Recognition performance as a function of detection criterion in a simultaneous detection-recognition task. *J. Acoust. Soc. Am.* 44, 204–211.

Luce, R. D., and Green, D. M. (1978). Two tests of a neural attention hypothesis for auditory psychophysics. *Percept. Psychophys.* 23, 363–371.

Macmillan, N. A., and Creelman, C. D. (2005). *Detection Theory: A User's Guide,* 2nd Edn. Mahwah NJ: Lawrence Erlbaum Associates.

Maris, E. (2002). The role of orthographic and phonological codes in the word and the pseudoword superiority effect: an analysis by means of multinomial processing tree models. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 1409–1431.

McClelland, J. L. (1976). Preliminary letter identification in the perception of words and nonwords. *J. Exp. Psychol. Hum. Percept. Perform* 2, 80–91.

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

McClelland, J. L., and Johnston, J. C. (1977). The role of familiar units in perception of words and nonwords. *Percept. Psychophys.* 22, 249–261.

McClelland, J. L., Mirman, D., and Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends Cogn. Sci. (Regul. Ed.)* 10, 363–369.

McClelland, J. L., and Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychol. Rev.* 88, 375–407.

McQueen, J. M., Norris, D., and Cutler, A. (1994). Competition in spoken word recognition: spotting words in other words. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 621–638.

Merikle, P. M., and Cheesman, J. (1986). Consciousness is a "subjective" state. *Behav. Brain Sci.* 9, 42.

Merikle, P. M., and Reingold, E. M. (1990). Recognition and lexical decision without detection: unconscious perception? *J. Exp. Psychol. Hum. Percept. Perform.* 16, 574–583.

Miller, J. (2000). Measurement error in subliminal perception experiments: simulation analyses of two regression methods. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 1461–1477.

Moore, B. C. J., and Glasberg, B. R. (1987). Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns. *Hear. Res.* 28, 209–225.

Moore, B. C. J., Glasberg, B. R., and Baer, T. (1997). A model for the prediction of thresholds, loudness, and partial loudness. *J. Audio. Eng. Soc.* 45, 224–240.

Naccache, L., and Dehaene, S. (2001). The priming method: imaging unconscious repetition priming reveals an abstract representation of number in the parietal lobes. *Cereb. Cortex* 11, 966–974.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: roles of inhibitionless spreading activation and limited-capacity attention. *J. Exp. Psychol. Gen.* 106, 226–254.

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

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

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

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113.

Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. (1992). "Complex sounds and auditory images," in *Auditory Physiology and Perception,* eds Y. Cazals, K. Horner, and L. Demany (Oxford: Pergamon Press), 429–446.

Perrin, F., and Grimault, N. (2005). *Fonds Sonores (Version 1.0)* [Sound samples]. Available at: http://olfac.univ-lyon1.fr/unite/equipe-02/FondsSonores.html

Pitt, M. A., and Samuel, A. G. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. *J. Exp. Psychol. Hum. Percept. Perform.* 19, 699–725.

Rapp, D. N., and Samuel, A. G. (2002). A reason to rhyme: phonological and semantic influences on lexical access. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 564–571.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *J. Exp. Psychol.* 81, 275–280.

Reingold, E. M. (2004). Unconscious perception and the classic dissociation paradigm: a new angle? *Percept. Psychophys.* 66, 882–887; discussion 888–895.

Reingold, E. M., and Merikle, P. M. (1988). Using direct and indirect measures to study perception without awareness. *Percept. Psychophys.* 44, 563–575.

Rennies, J., Verhey, J. L., and Fastl, H. (2010). Comparison of loudness models for time-varying sounds. *Acta Acustica* 96, 383–396.

Samuel, A. G. (1997). Lexical activation produces potent phonemic percepts. *Cogn. Psychol.* 32, 97–127.

Samuel, A. G., and Pitt, M. A. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *J. Mem. Lang.* 48, 416–434.

Sandberg, K., Timmermans, B., Overgaard, M., and Cleeremans, A. (2010). Measuring consciousness: is one measure better than the other? *Conscious. Cogn.* 18, 1069–1078.

Shipley, E. (1965). Detection and recognition: experiments and choice models. *J. Math. Psychol.* 2, 277–311.

Snodgrass, M., Bernat, E., and Shevrin, H. (2004a). Unconscious perception at the objective detection threshold exists. *Percept. Psychophys.* 66, 888–895.

Snodgrass, M., Bernat, E., and Shevrin, H. (2004b). Unconscious perception: a model-based approach to method and evidence. *Percept. Psychophys.* 66, 846–867.

Snodgrass, M., Shevrin, H., and Kopka, M. (1993). The mediation of intentional judgments by unconscious perceptions: the influences of task strategy, task preference, word meaning, and motivation. *Conscious. Cogn.* 2, 169–193.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science* 167, 392–393.

Warren, R. M. (1984). Perceptual restoration of obliterated sounds. *Psychol. Bull.* 96, 371–383.

Warren, R. M., and Obusek, C. J. (1971). Speech perception and phonemic restorations. *Percept. Psychophys.* 9, 358–362.

Warren, R. M., and Sherman, G. L. (1974). Phonemic restorations based on subsequent context. *Percept. Psychophys.* 16, 150–156.

Wheeler, D. D. (1970). Processes in word recognition. *Cogn. Psychol.* 1, 59–85.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 13 April 2011; accepted: 14 July 2011; published online: 26 July 2011.
Citation: Signoret C, Gaudrain E, Tillmann B, Grimault N and Perrin F (2011) Facilitated auditory detection for speech sounds. Front. Psychology ***2****:176. doi: 10.3389/fpsyg.2011.00176*
This article was submitted to Frontiers in Auditory Cognitive Neuroscience, a specialty of Frontiers in Psychology.