



# Italians use abstract knowledge about lexical stress during spoken-word recognition

Simone Sulpizio<sup>a,\*</sup>, James M. McQueen<sup>b,c,d</sup>

<sup>a</sup> Department of Cognitive and Education Sciences, University of Trento, Trento, Italy

<sup>b</sup> Donders Institute for Brain, Cognition, and Behaviour, Centre for Cognition, Radboud University Nijmegen, Nijmegen, The Netherlands

<sup>c</sup> Behavioural Science Institute, Radboud University Nijmegen, Nijmegen, The Netherlands

<sup>d</sup> Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

## ARTICLE INFO

### Article history:

Received 12 April 2011

revision received 8 August 2011

Available online 3 September 2011

### Keywords:

Lexical stress

Prosodic knowledge

Suprasegmental abstraction

Spoken-word recognition

## ABSTRACT

In two eye-tracking experiments in Italian, we investigated how acoustic information and stored knowledge about lexical stress are used during the recognition of tri-syllabic spoken words. Experiment 1 showed that Italians use acoustic cues to a word's stress pattern rapidly in word recognition, but only for words with antepenultimate stress. Words with penultimate stress – the most common pattern – appeared to be recognized by default. In Experiment 2, listeners had to learn new words from which some stress cues had been removed, and then recognize reduced- and full-cue versions of those words. The acoustic manipulation affected recognition only of newly-learned words with antepenultimate stress: Full-cue versions, even though they were never heard during training, were recognized earlier than reduced-cue versions. Newly-learned words with penultimate stress were recognized earlier overall, but recognition of the two versions of these words did not differ. Abstract knowledge (i.e., knowledge generalized over the lexicon) about lexical stress – which pattern is the default and which cues signal the non-default pattern – appears to be used during the recognition of known and newly-learned Italian words.

© 2011 Elsevier Inc. All rights reserved.

## Introduction

As listeners recognize spoken words, they must combine acoustic–phonetic information in the speech signal with stored knowledge about the sound patterns of words. This much is uncontroversial. But which sources of information do listeners rely on, what knowledge do they have about how words sound, and when do they integrate information that has been extracted from the speech signal with stored knowledge? We ask here when and how Italian listeners recognize polysyllabic Italian words that differ in their stress patterns. Answers to these questions provide constraints on the nature of the lexical access process, and on the nature of the knowledge stored in the mental lexicon.

How words are accessed and stored in the lexicon is a matter of ongoing debate. Two extreme theoretical positions can be defined. According to the first approach, the mental lexicon consists of episodic traces. Each word is represented by multiple traces that consist of detailed acoustic representations of episodic encounters with those words (Goldinger, 1998; Pierrehumbert, 2002). Word recognition entails comparison of the current acoustically detailed input with those stored traces. There thus needs to be no phonological abstraction prior to lexical access. The second approach assumes that the mental lexicon contains phonologically abstract forms (Gaskell & Marslen-Wilson, 1997; McClelland & Elman, 1986; Norris & McQueen, 2008). Word recognition again entails comparison of the current input with stored lexical knowledge, but this requires a prelexical stage of phonological abstraction so that contact can be made with the abstract representations in the lexicon.

\* Corresponding author. Address: Department of Cognitive and Education Sciences (DiSCoF), Corso Bettini, 31, 38068 Rovereto, Trento, Italy.

E-mail address: [simone.sulpizio@unitn.it](mailto:simone.sulpizio@unitn.it) (S. Sulpizio).

Neither of these extreme positions is tenable. Strictly episodic models cannot explain evidence of prelexical abstraction about speech segments (McQueen, Cutler, & Norris, 2006), and strictly abstractionist models cannot explain evidence that episodic details are maintained in long-term memory (Goldinger, 1998). What is required, therefore, is a hybrid model with both episodic and abstractionist components (Cutler, Eisner, McQueen, & Norris, 2010; Goldinger, 2007). An important question to ask, therefore, is what the division of labor is between these two components in the word-recognition process. For example, do listeners have abstract knowledge not only about speech sounds (McQueen et al., 2006) but also about the prosodic structure of words (that is, about their lexical stress patterns and about other aspects of lexical prosody)? Is that knowledge the result of forming generalizations over the lexicon? Furthermore, can listeners use that knowledge during the lexical access process? We asked these questions here, with respect to knowledge about stress in Italian words.

Italian offers an especially interesting test of whether abstract prosodic knowledge is used in word recognition because it has a strongly asymmetrical distribution of lexical stress patterns. Consider three-syllable words. There are two main stress types (Krämer, 2009): an antepenultimate stress pattern (i.e., the first syllable bears stress, e.g., TAvolo 'table'; capital letters indicate stress), and a penultimate stress pattern (i.e., stress appears on the second syllable, e.g., coLOre 'color'). The only rule to assign stress in trisyllabic words refers to the weight of the penultimate syllable: If it is heavy – that is, if it ends in a consonant – then it must be stressed (Krämer, 2009). Nevertheless, there is a strong distributional bias toward the penultimate stress pattern. In fact, 80% of Italian tri-syllabic words have penultimate stress, 18% have antepenultimate stress, and 2% have stress on the last syllable (e.g., serviTU, 'servitude'; Thornton, Iacobini, & Burani, 1997). This distributional asymmetry may be reflected in how Italians recognize spoken words. If they have abstracted the knowledge (generalized over the relevant entries in the Italian lexicon) that a trisyllabic word will usually have penultimate stress, then they may assume (in the absence of evidence to the contrary) that this is the stress pattern of any trisyllabic word they hear. This assumption that there is a default stress pattern may apply both when Italians are recognizing known Italian words, and when they are recognizing newly-learned words. We tested both these possibilities in the present experiments.

Prior research has already indicated that Italian listeners are sensitive to lexical stress information (Tagliapietra & Tabossi, 2005). In a cross-modal priming paradigm, listeners performed a lexical decision task on visual targets preceded by spoken bi-syllabic primes. Responses were facilitated when the target (e.g., GOMito, 'elbow') was preceded by a fragment-prime with the same stress pattern (e.g., GOMi), in line with previous findings for Dutch (Cutler & Van Donselaar, 2001; van Donselaar, Koster, & Cutler, 2005) and Spanish (Soto-Faraco, Sebastian-Galles, & Cutler, 2001). Italian listeners thus appear to use lexical stress cues to recognize spoken words. It is not clear, however, how early in the recognition process knowledge and

information about stress in Italian are brought to bear. Dutch listeners use stress information very early (i.e., in words that are segmentally identical in their initial syllables, such as OCTopus, 'octopus', and okTOber, 'October', stress information is used prior to the segmental disambiguation point; Reinisch, Jesse, & McQueen, 2010). Since in Italian, as in Dutch, the difference between stressed and unstressed syllables is at the suprasegmental rather than the segmental level, we expect that Italian listeners can also take advantage of stress cues early in the recognition process. An open question, however, is whether the distributional bias toward the penultimate stress pattern in Italian can affect the earliest stages of word recognition.

Furthermore, although Tagliapietra and Tabossi's (2005) findings suggest that the word-recognition process in Italian benefits from stress information, it remains unclear what exactly that information is. Which acoustic cues specify the stress patterns of Italian words? In general, stressed vowels differ acoustically from unstressed vowels in pitch, duration, and intensity (Albano Leoni & Maturi, 1998). But it is not clear which of these acoustic cues Italian listeners pick up on. Some authors consider amplitude to be the main stress correlate (Albano Leoni & Maturi, 1998). Others argue that duration plays the main role (Alfano, 2006; Alfano, Savy, & Llisterra, 2009). An additional aim of the present study was therefore to establish which stress cues Italian listeners use during word recognition. We were especially interested in whether the bias toward the penultimate syllable stress pattern modulates the way the acoustic information that signals stress is processed. In fact, if Italian listeners have stored knowledge about the acoustic correlates of stress and about the asymmetrical distribution of the two stress patterns, then it is possible that their use of acoustic information about stress may also be asymmetric. In particular, they should be more sensitive to the acoustic cues specifying an antepenultimate stress pattern than to those specifying the penultimate pattern – because the latter pattern can be assumed to occur by default.

In summary, the present study investigated three related questions. First, when do Italians use knowledge and information about lexical stress in spoken-word recognition? Second, how does the distributional bias favoring penultimate stress in Italian affect the recognition process? Third, which acoustic cues are picked up by Italians as they detect stress position, and how do these cues interact with the distributional bias? Answers to these questions should inform the debate on the nature of lexical representation. Is lexical stress knowledge stored in an abstract way (i.e., are there generalizations made across the Italian lexicon), and is that knowledge available to assist in word recognition?

To address these questions, we examined how Italian listeners use lexical stress to recognize known and newly-learned words. In Experiment 2, an artificial-lexicon study, we examined recognition of newly-learned words. This allowed us to control for the amount of exposure to specific episodes of those words and test whether prior knowledge about prosodic structure (abstracted from earlier experience with real Italian words) can nonetheless be brought to bear during word recognition. Shatzman and McQueen (2006) used the same paradigm to test whether

Dutch listeners have abstract prosodic knowledge about syllable duration and whether they can use it in the recognition of new words. Shatzman and McQueen trained participants to associate spoken non-words with novel shapes (displayed on a computer screen). The critical materials were pairs of monosyllabic non-words (e.g., *bap*) and bisyllabic non-words which had the same syllable embedded in onset position (e.g., *baptoe*). The initial syllables in each pair (e.g., *bap*) had the same ambiguous duration during the training phase of the experiment. In the subsequent test phase, syllable duration was manipulated: It was longer, shorter or equal to the duration used during training. The results showed that participants tended to interpret shorter syllables as bisyllabic word onsets and longer syllables as monosyllabic words, as indeed tends to be the case in real Dutch words (Salverda, Dahan, & McQueen, 2003), even though the participants had heard the novel words with only ambiguous durations during the training phase. Dutch listeners thus appear to have abstract prosodic knowledge about syllable duration and they appear to be able to use this knowledge during the recognition of newly-learned words. Experiment 2 is based on Shatzman and McQueen (2006). We test there whether Italian listeners have abstract prosodic knowledge about lexical stress (about which pattern is the default and about the cues which specify a word's stress pattern) and whether they can use this knowledge to improve their ability to recognize novel words. Experiment 2 thus provides the critical test of whether stress knowledge in Italian is abstract.

In Experiment 1, however, we first use real words to examine when Italian listeners use lexical stress information in spoken-word recognition and whether the distributional bias favoring penultimate stress affects the recognition process. Moreover, we investigated which acoustic cues Italians used to detect stress. The answers to these questions provide the basis for the further investigations in Experiment 2. Before asking if Italians use abstract knowledge about lexical stress in recognizing new words, we have to establish whether this knowledge exists and, if so, how it is used in the recognition of known words.

## Experiment 1

In Experiment 1, we used the printed-word eye-tracking paradigm (Huettig & McQueen, 2007; McQueen & Viebahn, 2007). Italian listeners heard spoken target words (e.g., *Canapa*, 'hemp') and had to identify the printed forms of those words on a computer screen from among an array of four alternative words which included a competitor with overlapping onset segments but a different stress pattern (e.g., *caNale*, 'channel'). Previous findings with this paradigm have shown that Dutch listeners use stress information as soon as it becomes available: The listeners preferred to fixate the targets before their spoken forms diverged segmentally from the competitors (Reinisch et al., 2010). We assume that a similar pattern of results will emerge for Italians. Two reasons make it plausible that Dutch and Italian will be treated similarly. First, lexical stress does not modify the segmental material in either language. In particular, unstressed vowels are not reduced

(as occurs, e.g., in English). Second, studies conducted in these two languages using the identity priming paradigm have shown similar results: Listeners benefit from stress information during word recognition (for Italian, see Tagliapietra & Tabossi, 2005; for Dutch, see Cutler & Van Donselaar, 2001; van Donselaar, Koster, & Cutler, 2005).

We thus hypothesize that Italian listeners will use stress information to constrain lexical access as soon as that information becomes available. Moreover, we expect that the asymmetry in the penultimate and antepenultimate stress distribution will affect word recognition. We hypothesize that listeners have knowledge about this distributional asymmetry and that they use this knowledge to optimize word recognition. If listeners know, when they hear a trisyllabic word, that it will usually have a penultimate-stress pattern (i.e., 80% of the time), then they can consider this pattern as a default. Acoustic cues may therefore not play a large role in the recognition of penultimate-stress words. In contrast, antepenultimate stress detection may be driven by the acoustic cues in the speech signal. Listeners could use this information to detect antepenultimate stress and hence to discard the default pattern. Acoustic cues may therefore be more important in the recognition of antepenultimate-stress words than in the recognition of penultimate-stress words.

In summary, we tested the following predictions. First, listeners should use the words' stress patterns to disambiguate segmentally identical fragments. They should tend to fixate targets (e.g., *Canapa*) and tend to ignore their segmentally overlapping competitors (e.g., *caNale*) before segmental disambiguation (e.g., the /p/ of *Canapa*) is available. Second, the distributional bias in Italian should affect performance. If Italians indeed assign penultimate stress by default, they should need to use acoustic cues to stress actively only when recognizing words with antepenultimate stress. Acoustic markers of stress should thus correlate with eye-movement behavior only for antepenultimate-stress targets. Testing this latter prediction should also allow us to identify which acoustic cues drive antepenultimate-stress detection; that is, we should be able to establish whether Italian listeners depend more on duration (Alfano, 2006; Alfano et al., 2009) or on amplitude (Albano Leoni & Maturi, 1998) in stress recognition.

## Method

### Participants

Thirty-two students (mean age: 26.3, *sd*: 6.2) from the University of Trento took part in the experiment. They received course credit for their participation. All participants were Italian native speakers with no known hearing problems and normal or corrected-to-normal vision.

### Materials

Thirty-two pairs of trisyllabic words were selected as experimental targets (see Appendix A). The words in each pair were segmentally identical in their first two syllables, but they differed in stress location. One word in each pair had stress on the penultimate (second) syllable and the other had stress on the antepenultimate (first) syllable. All pairs could be segmentally distinguished at the beginning

of the third syllable (e.g., *Canapa* and *caNAle*). Thirty-two distractor pairs were then selected, each coupled to one of the experimental stress pairs. Words in each distractor pair overlapped orthographically and phonologically on their first two syllables, and they did or did not differ in stress pattern (e.g., *GEnero* ‘son-in-law’, *GEnesi* ‘genesis’ for a stress-matched pair; *RUGine* ‘rust’, *rugGIto* ‘roar’ for a stress-mismatched pair). There were no semantic relationships among the four words in each set (i.e., an experimental pair plus a distractor pair). Twelve additional pairs were selected to be used in practice trials. Stress pairs as well as distractor pairs were matched on frequency (both  $t$ ’s  $< 1$ ) (CoLFIS database, Bertinetto et al., 2005), length in syllables (all were trisyllabic) and length in letters ( $t(31) = 1.13$ ,  $p = .14$ ). Acoustic measures of the stimuli are given below in Table 1.

A female native Italian speaker, naïve about the experiment’s purpose, recorded the stimuli in a sound-attenuated room (sampling at 44 kHz, 16 bit resolution, mono). Each word (i.e., each member of each of the 32 sets) was spoken at the end of the sentence “Clicca sulla parola” (“Click on the word”), with the sentence accent on the target word. Speaking rate was measured as the ratio per stress-type condition between total sentence duration and the number of syllables in the sentence. These ratios were matched across conditions (penultimate stress: 5.59 syllables per second; antepenultimate stress: 5.61 syllables per second;  $t < 1$ ).

### Procedure

Participants were seated about 50 cm in front of a computer screen (screen size 360 mm  $\times$  270 mm). The experiment had two parts: a familiarization task followed by the main eye-tracking experiment. During the first part, participants were familiarized with the stimuli. Because stimuli were all low-frequency words, all 64 experimental and distractor words were shown in lower-case letters in the middle of the screen, one by one in random order. Participants had to read them aloud: No word was found to be unknown to any participant, and all participants performed the task very well.

After this familiarization task, the eye-tracking experiment was run. Eye movements were recorded using a head-mounted Eyelink II System, at a sampling rate of

500 Hz, recording both eyes. The experimental section was composed of four blocks of 32 displays, each combined with a spoken instruction. In each display, four printed words were shown, one pair of experimental words plus one pair of distractor words. Each display of four words was shown in each block. Across blocks, different words from within each set of four were targets (i.e., were the words mentioned in the spoken instructions). In the first block, however, only words from the experimental stress pairs were selected as targets; half of them had penultimate stress and half had antepenultimate stress. In the subsequent blocks, the target could be the same word that was seen in the first block, its experimental competitor, or one of the distractor words. In this way, participants could not know which word they would hear when a given display was presented in any given block, because all four alternatives could occur as targets. Block order was counterbalanced across participants, and within each block trial order was randomized. The experiment was preceded by a small practice session using six displays; each display was shown two times, for a total of twelve practice trials. There were no breaks between the blocks.

Each trial started with a fixation cross in the center of the screen, displayed for 500 ms. Four words then appeared on the screen and remained there either until participants clicked the mouse button or for a maximum of 5000 ms. A white screen was used during the inter-stimulus interval of 480 ms. All words were presented in lower-case Lucida Sans Typewriter font, size 20. The four words were centered in the four quadrants of the screen. The auditory instructions (i.e., the carrier sentence plus each target word, e.g. “Clicca sulla parola *canapa*”) were played over headphones; the instructions began at the same time as the printed words appeared. Participants had to click the mouse on the target word that they heard at the end of the carrier sentence. Every eighth trial there was a drift correction to adjust for possible small head movements.

### Results

Three analyses of fixation behavior were performed. First, to test whether lexical stress information and/or knowledge is used early to distinguish between possible candidate words, a comparison between target and competitor fixations within each stress pattern (penultimate and antepenultimate) was run. Second, to test whether there was a distributional bias in the data, we performed an analysis comparing performance across the two stress patterns. Third, correlations between acoustic measures and behavioral data were run to establish which acoustic cues, if any, were used by listeners to detect the words’ stress patterns, and to ascertain whether this information was mainly used in the recognition of words with antepenultimate stress (i.e., the words with the non-default pattern), and less so (or not at all) in the recognition of penultimate-stress words.

Only trials in which participants clicked on the correct word were considered in the analyses (1% of all the trials were discarded for this reason). If a target was repeated during the experiment, only data from its first presentation were used. We considered fixations on a word as being all

**Table 1**

Mean acoustic measures and  $t$ -test comparisons for the first and second vowel of the words with each stress pattern in Experiment 1.

	First vowel	Second vowel	$t$ (29)	$p$ Value
<i>Antepenultimate stress</i>				
Duration (ms)	165	81	16.14	<.01
Pitch (Hz)	219	177	10.54	<.01
Amplitude (Pa)	.09	.04	8.03	<.01
Spectral tilt	.7	.03	5.19	<.01
<i>Penultimate stress</i>				
Duration (ms)	75	180	-18.2	<.01
Pitch (Hz)	238	204	4.43	<.01
Amplitude (Pa)	.09	.06	4.42	<.01
Spectral tilt	.3	.3	<1	n.s.

Note. Spectral tilt is expressed as a unitless ratio.



those that fell within a 6.3 cm square centered of the middle of each word: Thus, each fixation was coded as pertaining to the target, to the competitor, or to one of the two distractors. The proportion of fixations to each word over time (in 10 ms time intervals) was computed in each condition, by summing the number of fixations to each type of word and dividing it by the total number of fixations in the same time interval.

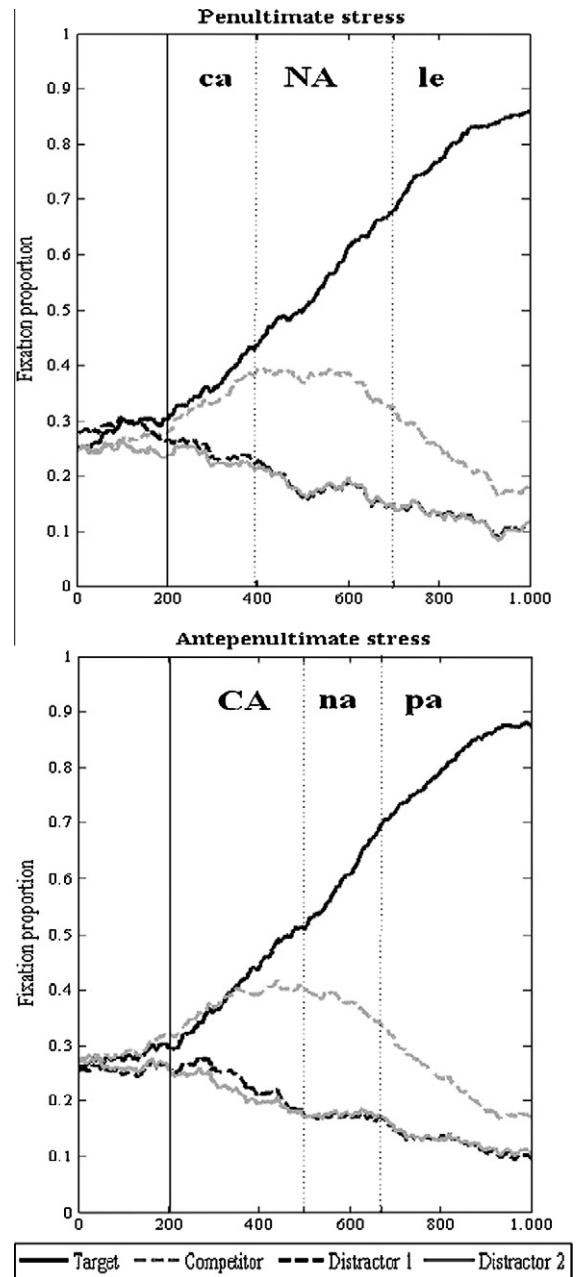
In all eye-tracking analyses, time windows were defined considering a delay of 200 ms as an estimate of the time needed to program and launch a saccade (Matin, Shao, & Boff, 1993). Thus, for example, when considering fixations in response to the first syllable of the words, a time window was defined as starting 200 ms after the acoustic onset of the syllable and ending 200 ms after the syllable's acoustic offset. Fig. 1 shows fixations on target, competitor and the two distractors over time for each stress pattern.

#### Comparing target and competitor fixations

Within each stress condition, a mixed-effects analysis (Baayen, Davidson, & Bates, 2008) was performed comparing fixation proportions on target and competitor words (e.g., *canapa* and *canale* when the spoken target was *canapa*). Fixation proportions were log transformed (Barr, 2008). Participants and items were treated as random factors, and stimulus type (target vs. competitor) was treated as a fixed factor. Models were fitted using R software (version 2.11; The R foundation for statistical computing) and *p*-values were calculated using the MCMC procedure, sampling 10,000 times (Baayen et al., 2008).

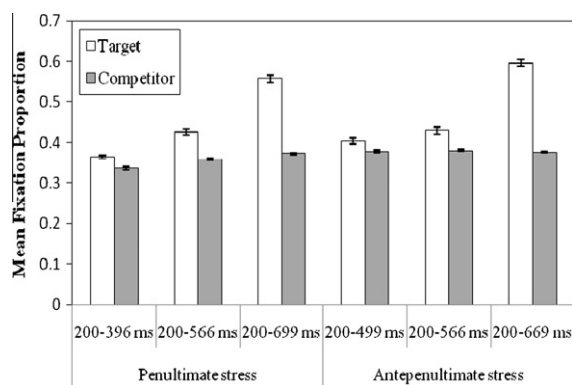
We ran three separate analyses considering the following time windows: first syllable, first 1.5 syllables, and first two syllables. Fig. 2 shows fixations proportions for targets and competitors in each of these three time windows. The first syllable and first two syllable windows were defined relative to the acoustic syllable boundaries, offset by 200 ms (i.e., 200–396 ms and 200–699 ms for penultimate-stress words, respectively, and 200–499 ms and 200–669 ms for antepenultimate-stress words). But the 1.5 syllable window was defined in absolute terms, and thus was of the same fixed length in both stress conditions (200–566 ms). The average duration of the first syllable (of both types of word) plus half of the average duration of the second syllable (again of both types of word) was 366 ms. The use of this time window thus allowed us to control for the differences across stress conditions in syllable duration, and thus also equated the amount of data used in the analysis in each condition. Although these syllable-duration differences are already controlled in the current within-item comparisons (target vs. competitor within stress type), they are not in the subsequent comparisons across stress types. Analyses of behavior in the 1.5 syllable window (along with those for the first syllable alone) also allowed us to ask whether eye movements were modulated by stress cues alone (i.e., before effects of the first consonant of the word's third syllable could influence behavior).

In the analyses of fixations in response to the first syllable, no differences between target and competitor were found (for penultimate stress,  $t < 1$ ; for antepenultimate stress,  $t = 1.01$ ). The analysis on fixations in response to



**Fig. 1.** Experiment 1: Fixation proportions to targets, competitors, and distractors over time (in ms on the abscissa). The solid vertical lines show the beginnings of the time windows starting 200 ms after the words' average onsets; the dotted lines indicate the ends of the time windows aligned to the average offsets of the first and the second syllables respectively, each again delayed by 200 ms.

the first 1.5 syllables revealed a difference between target and competitor fixations, for words with penultimate stress ( $\beta = 0.346$ ,  $t = 5.91$ ,  $p < .01$ ) and for words with antepenultimate stress ( $\beta = 0.467$ ,  $t = 6.96$ ,  $p < .01$ ). Participants looked at the target more than at the competitor before segmentally disambiguating information (at the onset of the third syllable) became available. The analysis on fixations in response to the first and second syllables showed



**Fig. 2.** Experiment 1: Mean fixation proportions to targets and competitors in each time window: the first syllable (200–396 ms or 200–499 ms, respectively, for penultimate and antepenultimate stress words), the first syllable plus half of the second syllable (200–566 ms for both types of word), and the first and second syllables (200–699 ms or 200–669 ms, respectively, for penultimate and antepenultimate stress words). Error bars are standard errors.

the same pattern for both penultimate ( $\beta = 0.597$ ,  $t = 10.87$ ,  $p < .01$ ) and antepenultimate stress ( $\beta = 0.466$ ,  $t = 8.39$ ,  $p < .01$ ).

#### Comparing penultimate- and antepenultimate-stress words

A mixed-effects analysis was run to see whether there was a difference in the proportion of fixations between penultimate- and antepenultimate-stress targets. Fixation proportions were again log transformed. The time windows included the first 1.5 syllables (as already noted, this window of fixed duration controls for durational differences across stress types) and the first two syllables (where amount of information in terms of the number of segments is controlled, and the durational difference across stress types is only 30 ms on average). Analyses of the first syllable alone were not included because of the large durational difference between antepenultimate- and penultimate-stress words (103 ms on average). Fixations on target words were used as the dependent variable, with stress type (penultimate vs. antepenultimate) as fixed factor and participants and items as random factors. No effect of stress type was found in either time window ( $t$ 's  $< 1$ ). To test whether the amount of competition varied across stress patterns, we conducted further analyses using the difference in the proportion of fixations to target and competitor as dependent variable and the stress type (penultimate vs. antepenultimate) as fixed factor. Again, no effect of stress type was found in either time window ( $t$ 's  $< 1$ ).

#### Correlation analyses

Acoustic measures of the first two vowels of the target words were performed in order to explore which information listeners used to determine the words' stress patterns. For the first and the second vowel of each target word, we measured pitch (in Hz), duration (in ms), amplitude (in Pa), and spectral tilt (calculated as a ratio between energy in high and low frequency band; see Cutler, Wales, Cooper, & Janssen, 2007). These analyses revealed that, in antepenultimate-stress words, the first vowel (i.e., the stressed vo-

wel) was longer, higher, and louder, and had more high frequency energy than the unstressed second vowel. In contrast, in the penultimate-stress words, the second (stressed) vowel was longer but it was also lower and weaker than the unstressed initial vowel, and the spectral tilt of the two vowels did not differ (see Table 1).

Two types of correlations were then performed comparing the acoustic measures of the first and second vowels with fixation behavior: within and between stress types. The within stress-type comparison indicates whether listeners used the differences between the vowels within words to detect the words' stress pattern. The between stress-type comparison provides an index of whether use of cues in the recognition of penultimate-stress words differs from that in the recognition of antepenultimate-stress words. In both cases, only significant correlations are reported.

*Within stress types.* For each stress pattern, correlations were performed on difference measures: For each acoustic measure, the difference between fixation proportions on the stressed and unstressed vowels was compared to the acoustic difference between the first and second vowels. A significant correlation was found only between the behavioral data for antepenultimate-stress targets and the amplitude difference between the vowels of those words ( $r = .46$ ,  $t(29) = 2.57$ ,  $p < .05$ ). As the difference between the first and second vowels became larger, listeners looked more at the target words. In a backward regression model, with fixation difference as dependent variable and the acoustic difference measures as predictors, amplitude was the only significant predictor in the model:  $t(29) = 2.57$ ,  $p < .05$ . ( $R^2 = .186$ , adjusted  $R^2 = .158$ ).

*Between stress types.* Correlations were also performed comparing the difference in fixation proportions between antepenultimate- and penultimate-stress words to the differences in the acoustic measures between antepenultimate- and penultimate-stress words. Based on the earlier analyses on fixation proportions, the first and second syllable time window was chosen. Behavioral and acoustic measures of the target words (e.g., *CaNapa*) were subtracted from the respective measures of their competitor words, that is, the words with the opposite stress pattern (e.g., *caNale*). The data showed a significant correlation between the difference in fixations between the antepenultimate-stress targets and their penultimate-stress competitors and the corresponding difference in spectral tilt ( $r = -.46$ ,  $t(29) = -2.81$ ,  $p < .01$ ). As the difference in spectral tilt between the second vowels of penultimate- and antepenultimate-stress words decreased, listeners looked less to antepenultimate-stress targets. A backward regression model with the fixation difference between antepenultimate targets and their competitors as dependent variable and the acoustic difference measures as predictors revealed that spectral tilt was the only significant predictor:  $t(29) = -2.18$ ,  $p < .01$  ( $R^2 = .215$ , adjusted  $R^2 = .188$ ).

#### Discussion

The results of Experiment 1 show that Italians use acoustic cues to lexical stress and lexical-stress knowledge during spoken-word recognition. In line with results ob-

tained in Dutch (Reinisch et al., 2010), Italian listeners use the acoustic information about stress in the speech signal as soon as it becomes available. They can thus distinguish between two trisyllabic words with segmentally-identical first and second syllables but different stress patterns (e.g., *Canapa* and *caNAle*) before the segmental disambiguation (the [p] or the [l]) is available to them.

The correlation analyses, however, suggest that Italians pick up on acoustic cues only when detecting antepenultimate stress: They used primarily intensity information (Albano Leoni & Maturi, 1998) to identify antepenultimate stress patterns, but appeared not to use acoustic cues when recognizing words with penultimate stress (despite the presence of such cues in the speech signal). The primary acoustic signal that listeners detect appears to be a marked decrease in amplitude of the second vowel compared to the first vowel of words with antepenultimate stress (note that there is a smaller decrease in amplitude across the first two vowels in words with penultimate stress, see Table 1). It might be assumed that listeners could use a further criterion, based on the amplitude of the first syllable: If the amplitude is higher than a threshold value, listeners could start to assume the word has antepenultimate stress. The analyses on fixations in response to the first syllable, however, did not show that the competition between target and competitor was already resolved at this point in time. Furthermore, as Table 1 shows, there is no difference in mean first syllable amplitude across conditions. These observations suggest that first syllable amplitude alone is not enough to recognize antepenultimate stress. But it is possible that listeners might use both criteria – the amplitude of the first vowel and the amplitude difference between the first and second vowels – to identify that a word has antepenultimate stress.

Even though there was no evidence of listener sensitivity to the acoustic cues signaling the penultimate stress pattern, penultimate-stress words were recognized just as quickly as antepenultimate-stress words. This suggests that Italians were using knowledge that penultimate stress is the much more frequent pattern, and so were recognizing penultimate-stress words by default. In short, it appears that Italian listeners assume that trisyllabic words will have stress on their penultimate syllables and hence will recognize sequences such as *cana-* as being the onset of *canale*, unless acoustic evidence (primarily a more marked decrease in amplitude in the second vowel relative to the first vowel) indicates that the antepenultimate syllable is stressed, and hence that they must be hearing *canapa*.

These findings thus indicate that Italians have knowledge about the stress-pattern distribution in the Italian lexicon. They know that penultimate stress is the most frequent pattern in trisyllabic Italian words, and they exploit this knowledge to optimize word recognition. They assign this more frequent pattern by default, and detect words with antepenultimate stress using the intensity information contained in the signal. To further test Italians' knowledge about the use of this distributional bias, and the interaction between this knowledge and the use of acoustic cues to stress, we ran a second experiment using an artificial lexicon. Critically, the use of newly-acquired words allowed us to test if stored prior knowledge about

lexical stress can be used by Italians when the number of exposures to those new words was controlled, and hence whether that knowledge is abstracted away from memories of specific lexical episodes.

## Experiment 2

In Experiment 2, we investigated how listeners use prosodic knowledge about stress to recognize newly-acquired words. We tested whether Italians apply their stored knowledge about default stress patterns and about the acoustic cues to stress when they are recognizing words that they have never heard before the experiment began. As discussed above, Shatzman and McQueen (2006) found evidence that Dutch listeners use stored prosodic knowledge about word duration to recognize newly-learned words. In keeping with this finding, we hypothesize that Italian listeners will use prior knowledge about lexical stress when recognizing new words. Following the distributional bias, Italian listeners may assign penultimate stress by default and identify only novel words with antepenultimate stress on the basis of the acoustic information in the speech signal.

To test these assumptions, we used an artificial-lexicon eye-tracking paradigm (Creel, Tanenhaus, & Aslin, 2006; Magnuson, Tanenhaus, Aslin, & Dahan, 2003; Shatzman & McQueen, 2006). Previous studies have shown that lexical access with an artificial lexicon works in a similar way to lexical access with a real lexicon: Participants' eye movements show the same kinds of effects as those observed with real words (Magnuson et al., 2003). In addition, the recognition of artificial-lexicon words appears to be relatively unaffected by their similarity to specific real words (i.e., there is effectively no competition from words belonging to the lexicon of the participant's native language; Magnuson et al., 2003). The use of an artificial lexicon therefore allowed us to investigate, in a controlled fashion, the involvement of stored prosodic knowledge and signal-based prosodic information during word recognition.

We trained participants to associate non-objects (non-sense shapes) with spoken non-words. The non-objects' names formed minimal pairs that were segmentally identical and differed only in stress placement (e.g., *TOlaco* vs. *toLAcO*). In the training phase, participants heard acoustically reduced versions of the non-words as they learned the object-word associations. Differences in two acoustic stress cues in the original natural utterances – amplitude and duration differences – were neutralized. In the test phase, participants heard stimuli in both reduced- and full-cue versions (i.e., with and without the acoustical manipulations, though note that the full-cue versions were also edited tokens and hence were not the original recordings). In the test phase, participants had to recognize the corresponding objects. If participants use their stored prosodic knowledge about lexical stress – that penultimate stress is the default, and that primarily amplitude cues signal words with antepenultimate stress – then there should be a difference in fixation behavior between the reduced- and full-cue versions only for antepenultimate stress words. If penultimate stress words such as *toLAcO* are

recognized by default, the addition of amplitude (and duration) cues should not influence their recognition. But the addition of these cues in the test phase should allow participants to perform better when they hear the full-cue versions of antepenultimate non-objects' names such as *Tolaco*. Critically, if this benefit for the full-cue versions is found, it must reflect prior abstract, not word-specific knowledge about antepenultimate stress cues (i.e., knowledge that is generalized over the real Italian lexicon and hence is not specific to the newly-learned words). It cannot reflect memories for specific episodic encounters with the newly-learned words, since, prior to the test phase, the participants will never have heard these words with amplitude (or duration) differences between their first two vowels.

In summary, in Experiment 2 we investigated whether Italian listeners exploit stored abstract knowledge about lexical stress to optimize the recognition of newly-learned words. Such a finding would suggest that prosodic knowledge should be considered part of the listener's abstract phonological knowledge about spoken words, knowledge which, alongside that about individual segments, is used during lexical access (Cho, McQueen, & Cox, 2007; Gaskell & Marslen-Wilson, 1997; McQueen et al., 2006; Norris and McQueen, 2008; Shatzman & McQueen, 2006).

### Method

#### Participants

Twenty-two students (mean age: 27.9, *sd*: 5.1) from the University of Trento took part. They received course credit for their participation. They were all Italian native speakers with no known hearing problems and normal or corrected-to-normal vision. None had participated in the previous experiment.

#### Materials

Twelve trisyllabic non-words were created (*binulo*, *canvilo*, *confuro*, *curfino*, *desico*, *goliso*, *patuco*, *pencilo*, *pindumo*, *tefubo*, *tolaco*, and *tudero*). Each non-word was recorded twice, once with penultimate stress (e.g., *toLaco*), and once with antepenultimate stress (e.g., *TOlaco*) by a female Italian speaker in a sound-attenuated room (sampling at 44 kHz, 16 bit resolution, mono). Each non-word was spoken at the end of the sentence "Clicca sul" ("Click on the"). As in Experiment 1, speaking rate was controlled across conditions (penultimate stress: 5.06 syllables per second; antepenultimate stress: 4.92 syllables per second;  $t = 1.16$ ,  $p = .25$ ). In this way, we obtained twelve critical pairs. Each critical pair was composed of two segmentally identical non-words that differed only in stress placement (*toLaco* vs. *TOlaco*). Twenty-four line drawings of nonsense objects were randomly selected from a database of non-objects (Non-existing Objects Database, [www-server.mpi.nl/experiment-pictures/production-pictures/](http://www-server.mpi.nl/experiment-pictures/production-pictures/); see Fig. 3). The nonsense objects were randomly assigned to the non-words.

We created a modified version of each non-word. Based on the results of Experiment 1, we neutralized one main stress cue (the amplitude of the first two vowels) and one secondary stress cue (the duration of these vowels).

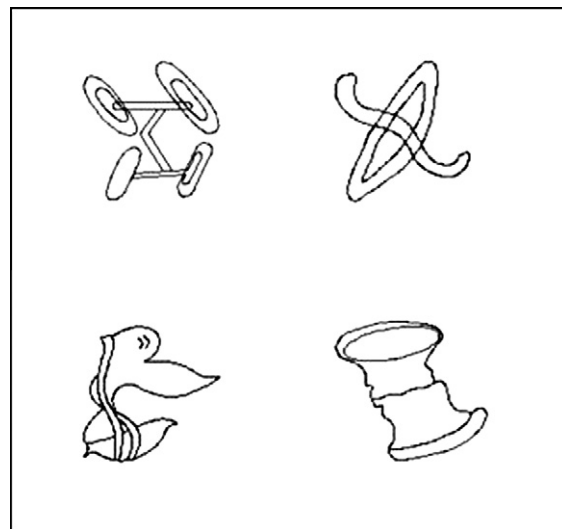


Fig. 3. Experiment 2: Examples of non-objects displayed in a 4-alternative trial.

For each non-word, we calculated the average amplitude and the average duration of its first and second vowels. Then, using the PSOLA algorithms in Praat (Boersma & Weenink, 2007), we replaced the original duration values of both vowels in each non-word with the average duration of those two vowels in each non-word. The amplitude of the first two vowels in each nonword was set to the average value of those vowels. In this way, we replaced the original acoustics of the two first two vowels of each non-word. Note that the full-cue versions were also obtained by editing the natural versions of the stimuli. That is, we applied the same adjustment procedures used when creating the reduced-cue versions, but replaced the original values with those same values. This meant that the overall duration and amplitude values in the full-cue materials remained the same as in the original recordings, but also ensured that the stimuli had nonetheless been passed through the same procedures, so that the reduced- and full-cue versions did not differ in their overall quality. In this way, we had two versions of each non-word: the full-cue version and the reduced-cue version, in which the acoustic cues to lexical stress pattern were partially neutralized (see Table 2). Both the full- and reduced-cue versions of all non-words were spliced back into the carrier sentence ("Clicca sul"). The same token of this sentence was used throughout.

For each stimulus, a feedback sentence for use in the training phase was also recorded by the same speaker, with the stimuli uttered at the end of the sentence (e.g., "Ora puoi vedere di nuovo il *Tolaco*", "Now you can see the *Tolaco* again"). One token of this feedback sentence, without the final non-word, was selected and each reduced-cue non-word was spliced onto the end of it.

#### Procedure

The experiment was composed of three phases: two training phases plus a test phase. Because previous research has shown that the lexicalization of newly-acquired words



**Table 2**

Original values of duration, amplitude, pitch and spectral tilt and their *t*-test comparisons, for the first and second vowel of the novel words with each stress pattern in Experiment 2.

	First vowel	Second vowel	<i>t</i> (11)	<i>p</i> Value	Mean
<i>Antepenultimate stress</i>					
Duration (ms)	165	82	8.26	<.001	121
Amplitude (Pa)	.05	.02	5.19	<.001	.03
Pitch (Hz)	196	181	4.36	<.001	–
Spectral tilt	.3	.03	2.41	<.05	–
<i>Penultimate stress</i>					
Duration (ms)	64	165	–10.69	<.001	211
Amplitude (Pa)	.04	.02	3.36	<.005	.03
Pitch (Hz)	244	186	20.23	<.001	–
Spectral tilt	.1	.1	<1	n.s.	–

Notes. The mean values used in creating the manipulated versions of these stimuli are also reported. Spectral tilt is expressed as a unitless ratio.

is associated with nocturnal sleep (Davis, Di Betta, Macdonald, & Gaskell, 2009; Dumay & Gaskell, 2007), we decided to run the experiment over two consecutive days. This choice increased the chance that the new words would be learnt well, but note that even words learned over 2 days with the benefits of overnight consolidation should not be considered to be equivalent to existing words. On the first day, participants completed the first training phase. On the second day, they returned to do the second training phase and the test phase. During the test phase we recorded participants' eye-movements using a head-mounted Eye-link II System, at a sampling rate of 500 Hz and recording both eyes. Participants performed both the training and the test phase sitting approximately 50 cm in front of a computer screen (screen size 360 mm × 270 mm).

The first training phase was composed of five blocks. Within each block, each stimulus was presented four times, for a total of 96 trials for each block (24 non-words × 4 repetitions each). In Block 1, participants had to choose between two non-objects displayed on the screen; they never saw both non-objects that formed a critical pair on the same screen (e.g., we displayed TOLaco and biNUlo, but never TOLaco and toLAcO). In Block 2, participants had to choose between 2 objects that did form critical pairs (e.g., we displayed TOLaco and toLAcO). In Block 3, participants had to choose among four non-objects, and, as in Block 1, no critical pairs were displayed together (e.g., we displayed TOLaco, biNUlo, CANvilo, and deSIco). Block 4 was the same as Block 2 (two objects forming a critical pair). Finally, in Block 5 participants had to choose among 4 non-objects and, as in Blocks 2 and 4,

the displayed stimuli formed critical pairs. This procedure is summarized in Table 3.

As is also shown in Table 3, the second training phase was composed of four blocks. The procedures for Blocks 6 and 7 corresponded respectively to Blocks 1 and 2 from Day 1. Blocks 8 and 9 corresponded to Block 5. Note that we included two blocks where both members of a critical pair appeared on the same screen because a previous study on lexical learning showed that Italians build stress information into new lexical representations only when they are explicitly encouraged to do so (Sulpizio & McQueen, *in press*). Including trials with minimal pairs forced participants to attend to stress differences.

We used the same timing procedures in the two training phases. Each trial started with a fixation cross in the center of the screen, displayed for 500 ms. Then two or four non-objects appeared on the screen and remained there until participants clicked the mouse button. At the same time as the visual stimuli appeared, the auditory instruction (carrier sentence plus target word, e.g. "Clicca sul TOLaco") was played over headphones. Participants had to click the mouse on the target non-object that corresponded to the non-word they heard at the end of the carrier sentence. At the same time the mouse was clicked, a sentence was played to indicate if the response was correct (*giusto*, 'right') or not (*sbagliato* 'wrong'). Then the target non-object was displayed again, centered on the screen and the feedback sentence (e.g., "Ora puoi vedere di nuovo il TOLaco" 'Now you can see the TOLaco again') was played. In the two training phases participants heard the non-words only in their reduced-cue versions.

**Table 3**

Training block structure in Experiment 2 and percentage accuracy per block.

Day 1				Day 2			
Blocks	Objects	Stress pairs	Accuracy (%)	Block	Objects	Stress pairs	Accuracy (%)
B1	2	No	75	B6	2	No	98
B2	2	Yes	65	B7	2	Yes	86
B3	4	No	90	B8	4	Yes	87
B4	2	Yes	78	B9	4	Yes	90
B5	4	Yes	83				

Notes. Blocks = block number; objects = number of objects displayed per screen; stress pairs: yes if stress pairs were shown in the same display; accuracy = percentage of correct responses.

The test phase followed the second training phase. Before the test, the eye-tracker was mounted and calibrated. The test phase was composed of 2 blocks. Within each block, each trial was repeated two times. For each trial, participants heard a target non-word (e.g., *Tolaco*) and they had to select the corresponding non-object among four possible alternatives displayed on the screen. The four possible choices belonged to two critical pairs (e.g., *Tolaco* and *tolAco*; *Binulo* and *biNulo*). In Block 1, participants heard stimuli only in their reduced-cue versions, whereas in Block 2 they heard the non-words only in their full-cue versions. The two blocks were run one after the other, with no break between them. Stimuli were randomized within each block.

In the test phase each trial was structured as follows. First a fixation cross was displayed, centered on the screen, for 500 ms. Then four non-objects appeared on the screen (see Fig. 3) and remained there either until participants clicked the mouse button or for a maximum of 5000 ms. A white screen was used during the inter-stimulus interval of 480 ms. The four non-objects were centered in the four quadrants of the screen. The auditory instructions (carrier sentence plus target non-word, e.g., *Clicca sul Tolaco*) were played over headphones (starting when the non-objects appeared). Participants had to click the mouse on the target non-object whose name was heard at the end of the carrier sentence. During the test phase, participants did not receive any feedback. Every eighth trial there was a drift correction to adjust for possible small head movements.

## Results

During the training phases, participants successfully learned the non-object names. At the end of the first training phase, object identification accuracy reached 83%, whereas at the end of the second training phase it reached 90% (for details, see Table 3). Training phase data were not analyzed further.

Two analyses were performed on the results from the test phase. First, we ran a  $2 \times 2$  analysis, comparing performance on the two stress types (penultimate and antepenultimate) for each of the two acoustic versions of each newly-learned word (reduced-cue and full-cue versions, with data from Blocks 1 and 2 of the test phase respectively). In this way, we tested whether there was evidence that the distributional bias favoring penultimate stress in the Italian lexicon influenced the behavioral data and more specifically whether this bias affected the recognition of the reduced- and full-cue stimuli. Second, correlations between duration and amplitude measures of the full-cue stimuli (i.e., the cues that had been neutralized during training) and the fixation behavior on these stimuli were run in order to establish whether listeners, in this condition, used those cues to identify the words' stress patterns.

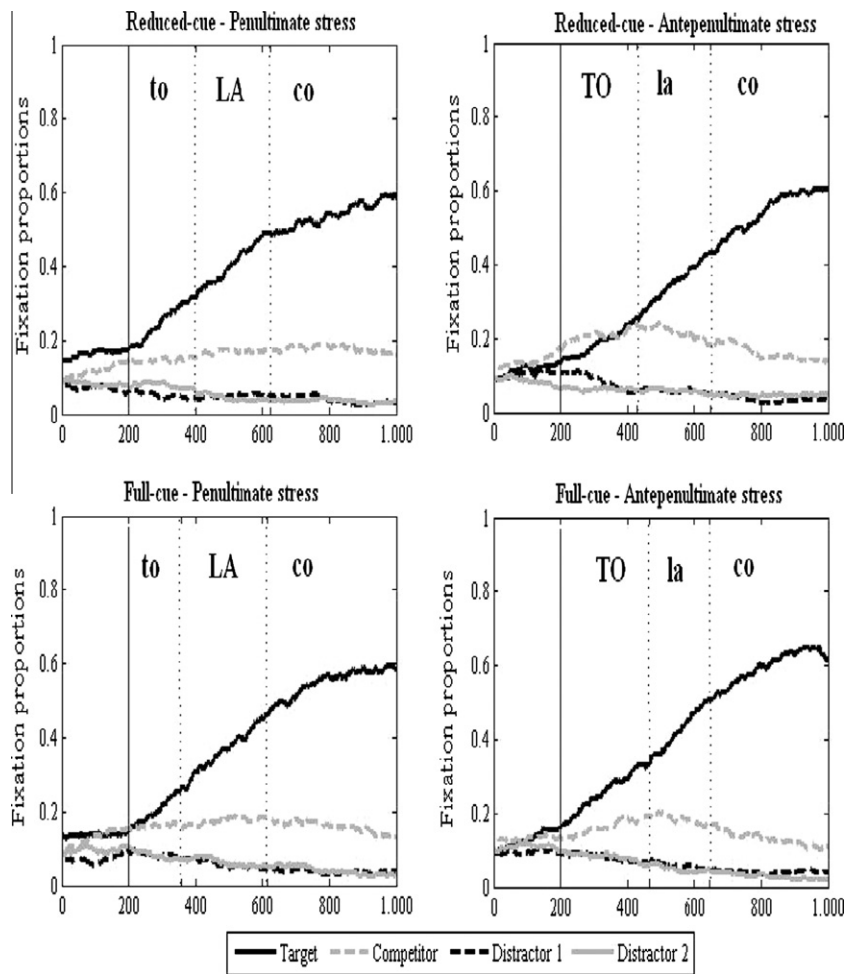
Only trials in which participants clicked on the correct non-object were considered in these analyses (10% of all observations were discarded for this reason). We considered fixations on a non-object as being all those that fell within a 6.3 cm square centered on the middle of each non-object. Thus, each fixation was coded as being made

to the target non-object, to its competitor, or to one of the distractor non-objects. Fixation proportions were computed in the same way as in Experiment 1, and, also as before, three time windows (first syllable, an absolute window corresponding to the grand average duration of the first 1.5 syllables, and the first two syllables) were defined (again with an offset of 200 ms for programming and making a saccade). Fig. 4 shows fixations on target non-objects, their competitors and the two distractors over time for each experimental condition. Fig. 5 shows fixation proportions to targets in each condition in each of the time windows. As in Experiment 1, mixed-effect analyses of log transformed data were performed with participants and items as random factors.

## Fixation analyses

**Target analysis.** Fixation proportions on targets was the dependent variable, and stress type (penultimate vs. antepenultimate), acoustic version (reduced- or full-cue), and their interaction were fixed factors. The analysis on fixations in response to the first-syllable revealed that participants fixated more penultimate-stress than antepenultimate-stress targets ( $\beta = -0.81$ ,  $t = -3.19$ ,  $p < .01$ ). An interaction between the two factors revealed that, compared to the reduced-cue version, listeners improved their performance when they heard the full-cue versions of the newly-learned words, but this happened only for stimuli with antepenultimate stress ( $\beta = 0.77$ ,  $t = 2.16$ ,  $p < .05$ ). No main effect of acoustic version was found ( $t = -1.6$ ). This suggests that addition of acoustic cues benefited recognition only of newly-learned words with antepenultimate stress. But the duration of the first syllables differed across conditions (and hence the amount of data contributing to the different cells of the analysis was not controlled). The same results were found, however, in the other two analyses, where durational differences were controlled (across stress types and acoustic versions for the 1.5-syllable window, and across acoustic versions for the 2-syllable window). For responses to the first 320 ms of the stimuli (the 200–520 ms time window, i.e., the first 1.5 syllables), there was a main effect of stress type ( $\beta = -0.97$ ,  $t = -3.38$ ,  $p < .01$ ), a significant interaction between stress type and acoustic version ( $\beta = 0.87$ ,  $t = 2.13$ ,  $p < .05$ ) and no main effect of acoustic version ( $t = -1.54$ ). For responses to the first two syllables, there was again a main effect of stress type ( $\beta = -0.98$ ,  $t = -3.046$ ,  $p < .01$ ), a significant interaction between stress type and acoustic version ( $\beta = 0.86$ ,  $t = 1.98$ ,  $p < .05$ ) and no main effect of acoustic version ( $t = -1.6$ ).

**Target-competitor analysis.** Further analyses compared the amount of competition in the four conditions. We used the difference in fixation proportions on target and competitor as dependent variable and stress type and acoustic version as fixed factors. We selected the same three time windows as before. The analysis in the first-syllable time window revealed more competition for antepenultimate-stress than for the penultimate-stress targets ( $\beta = -1.20$ ,  $t = -2.55$ ,  $p < .01$ ), presumably because, in the former case, the (default) penultimate stress words were the competitors. Moreover, the interaction between the two factors revealed that competition decreased when listeners heard



**Fig. 4.** Experiment 2: Fixation proportions to targets, competitors, and distractors over time (in ms on the abscissa) in each of the four experimental conditions. Fixations in response to newly-learned words with penultimate (e.g., toLAco) and antepenultimate stress (e.g., TOlaCO) are shown on the left and right respectively. Fixations to reduced-cue tokens (those heard during the learning phase) are given in the upper panels; those to the full-cue tokens are shown in the lower panels. The solid vertical lines show the beginnings of the time windows starting 200 ms after the words' average onsets; the dotted lines indicate the ends of the time windows aligned to the average offsets of the first and the second syllables respectively, each again delayed by 200 ms.

the full-cue versions of the stimuli, but only when the targets had antepenultimate stress ( $\beta = 1.16$ ,  $t = 1.78$ ,  $p = .08$ ). The other two analyses (with durational differences controlled) showed the same pattern of results, with a main effect of stress type (first 1.5 syllables:  $\beta = -1.58$ ,  $t = -3.03$ ,  $p < .01$ ; first two syllables:  $\beta = -1.63$ ,  $t = -2.949$ ,  $p < .01$ ) and a significant interaction between stress type and acoustic version (first 1.5 syllables:  $\beta = 1.53$ ,  $t = 2.06$ ,  $p < .05$ ; first two syllables:  $\beta = 1.52$ ,  $t = 2.029$ ,  $p < .05$ ).

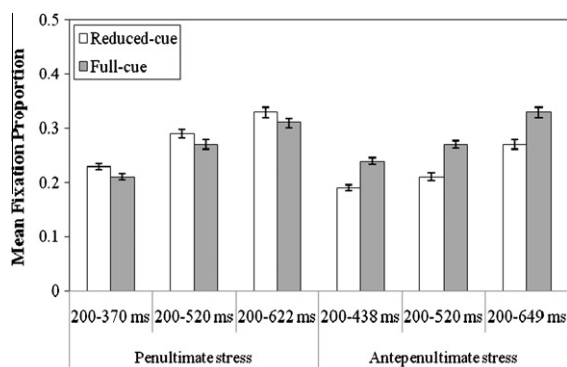
#### Correlation analysis

We performed correlation analyses to test whether, in the full-cue condition, amplitude and duration cues drove the observed improvement in antepenultimate-stress detection. We did not run these correlations for penultimate-stress targets because recognition of penultimate-stress targets in the full-cue condition did not improve. Considering first the initial syllable and then the first two syllables, correlations were performed across words, comparing the difference in fixation proportions between pairs

of antepenultimate- and penultimate-stress words to the duration differences between these pairs of words, and then again for the corresponding amplitude differences. Behavioral and acoustic measures of the target words were subtracted from the respective measures of their competitors (those with the opposite stress pattern). When we used the first syllable as the time window, we found a marginal correlation between fixations and duration ( $r = -.50$ ,  $t(10) = -1.83$ ,  $p < .1$ ). When we used the first two syllables as the time window, we found a marginal correlation between fixations and amplitude ( $r = -.41$ ,  $t(10) = -1.43$ ,  $p < .1$ ). In both correlations, as the difference in duration or amplitude increased, listeners tended to look more at the antepenultimate-stress targets.

#### Discussion

In Experiment 2, we found that Italian listeners applied two kinds of stored knowledge about lexical stress when recognizing newly-learned words. First, they appeared to



**Fig. 5.** Experiment 2: Mean fixation proportions to penultimate- and antepenultimate-stress targets for both the reduced- and the full-cue conditions. The mean values are given (in ms) for all three time windows: the first syllable (200–370 ms or 200–438 ms, respectively, for penultimate and antepenultimate stress words), the first syllable plus half of the second syllable (200–520 ms for both types of word), and the first and second syllables (200–622 ms or 200–649 ms, respectively, for penultimate and antepenultimate stress words). Error bars are standard errors.

have knowledge about the distributional stress bias in Italian trisyllabic words, and specifically that penultimate stress is the default, because the acoustic manipulation of stress cues did not affect how they recognized non-objects with penultimate stress. Moreover, penultimate-stress targets were recognized earlier than antepenultimate-stress targets: Listeners appear to assume that the penultimate pattern is the default. Second, they appeared to know about the acoustic cues that normally signal words with antepenultimate stress. Unlike in Experiment 1, they used not only amplitude, but also duration when detecting antepenultimate stress (though both effects were statistically weak). This difference across experiments suggests that Italians have knowledge about the variety of acoustic cues that are used to signal stress, and that they can use them to different degrees in different situations. In the normal situation (i.e., with the real words tested in Experiment 1), amplitude information appears to be enough to establish that the current word does not have the default stress pattern. But in the situation where listeners are attempting to recognize new words that differ only in stress (as in Experiment 2), they may appeal to all available acoustic cues (i.e., amplitude and duration information) to detect the non-default pattern.

Both of these findings reflect the use of abstract (i.e., not word-specific) knowledge about lexical stress. Listeners heard the antepenultimately and penultimately stressed newly-learned words in equal proportions, so there was nothing in their experience with these specific words that indicated that they should be treated differently. Furthermore, when listeners heard the full-cue versions of the newly-learned words, this improved their recognition of the antepenultimate-stress targets, even though they had learned those stimuli through hearing acoustically-different (reduced-cue) versions. That is, there was nothing in their prior experience with these new words that indicated they should have particular durational or amplitude properties. There are therefore two different types of knowl-

edge about stress that Italian listeners have abstracted and stored: the phonological patterns related to stress, and the relative frequency of those patterns in the Italian lexicon. Both of these types of knowledge appear to be used during the recognition of newly-learned words.

### General discussion

We investigated how Italians use lexical stress in spoken word recognition and whether they use abstract knowledge about lexical stress when recognizing spoken words. In Experiment 1, in line with previous results for Dutch (Reinisch et al., 2010), we found that Italian listeners used stress information in word recognition as soon as it became available, and prior to segmental disambiguation. Listeners considered penultimate stress (the most common pattern) to be the default, and picked up on acoustic cues to stress only when recognizing words with antepenultimate stress. In Experiment 2, we found two main results. First, we found further evidence for default assignment of penultimate stress: Novel words with this stress pattern were recognized more quickly than those with antepenultimate stress, and the addition of stress cues to penultimate-stress newly-learned words did not improve their recognition. Second, we found that Italians used their knowledge about the acoustic cues normally associated with antepenultimate stress to help them recognize new antepenultimate-stress words. This prior knowledge appears to be abstract knowledge about how stress is normally cued in Italian (i.e., not knowledge specific to the newly-learned words) since, prior to the test phase, these cues had not been associated with the novel words.

These results shed new light on several issues. First, they provide information on the acoustic cues that Italians use as they detect stress position and on how these cues interact with the distributional bias. Our results show that Italian listeners use mainly amplitude to identify a word's stress pattern (Albano Leoni & Maturi, 1998), but only when those words have antepenultimate stress. We also found that duration may be involved in the detection of antepenultimate stress. This happened only in Experiment 2, during the recognition of newly-learned words (and there only weakly). In such cases, amplitude information may not be sufficient to signal antepenultimate stress: To detect the correct stress pattern of a newly-learned word accurately, listeners may tend to use all the acoustic information that they find in the signal. These results are partially in contrast with those of Alfano (2006). In her experiments, Alfano manipulated the vowel duration and pitch of words belonging to minimal pairs (e.g., *PAgano* 'they pay' vs. *paGAno* 'pagan'), but she did not investigate amplitude. She asked participants to listen to the manipulated words and to identify which syllable bore stress; then, listeners had to judge whether the two words of the minimal pairs had the same pattern or not. She showed that duration was the main acoustic cue that the listeners used to identify stress. A possible explanation for these different findings is that listeners are able to use more or fewer cues depending on the amount of information found in the signal. Thus, when amplitude information is not sufficient, Italians may also use other available cues to detect stress.



The way Italians exploited acoustic information for recognition of newly-learned antepenultimate-stress words suggests that they have abstract knowledge about the acoustic cues related to stress. This means not only that Italian listeners are able to analyze spoken words into their component phonological parts (segments and suprasegmental attributes), but also that they are able to form abstractions about those components. In particular, they appear to have knowledge that words that have a particular stress pattern tend to have particular acoustic properties. Besides this acoustic–phonetic knowledge, however, Italians also have other knowledge available for use in the recognition of both novel and well-known words. The listeners combined their knowledge about acoustic cues with knowledge about the biased distribution of penultimate and antepenultimate stress in Italian (Thornton et al., 1997). These two sources of stored knowledge appear to interact with each other to optimize stress detection. This could work very efficiently, at least in the situation where it is known that the target word will have three syllables, as in the present experiments. Italian listeners could assign penultimate stress by default (because they know that 80% of trisyllabic words will have this pattern) and then test the validity of this assumption by checking the phonetic information in the speech signal. If the first two syllables of a trisyllabic word contain antepenultimate stress cues (e.g., a marked reduction in amplitude going from the first to the second syllable), then listeners would need to change this default assumption, but otherwise they could maintain the hypothesis that they are hearing a word with penultimate stress. The assumption that penultimate stress is the default pattern might extend beyond trisyllabic words, however. This is because, across word lengths, penultimate stress is the most common pattern in the Italian lexicon (Krämer, 2009). Italians might thus exploit knowledge about this bias in words of all lengths.

This study also addressed a temporal question: When do Italians use stress information in spoken-word recognition? Lexical stress is a source of information that could help to resolve the lexical competition process: Stress information can reduce the number of possible competitors that the listener needs to consider during word recognition. The results of Experiment 1 show that Italians take advantage of stress information to modulate the lexical competition process as soon as that information comes available. In this situation, they exploited lexical stress information (and knowledge about the distributional bias) before segmentally disambiguating material became available (as Dutch listeners also appear to do with respect to signal-based stress cues; Reinisch et al., 2010). More generally, however, Italian listeners are likely to use segmental and suprasegmental information at the same time during word recognition. For instance, in distinguishing between *CaNaPa* and *caNaLe* as they unfold over time, Italians appear to use the stress information in the first two syllables to help resolve the temporary ambiguity between these two candidates, but it is likely that they are also using the segmental information in these syllables to rule out other candidates – those that do not begin /kana/.

Three points should be made concerning this perspective on how stress is processed over time in Italian. First, the present findings offer further support for the view the listening to speech is incrementally optimal (Norris & McQueen, 2008; Reinisch et al., 2010; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Warren & Marslen-Wilson, 1987), that is, that listeners use all incoming information as soon as it becomes available to form an optimal interpretation of the currently unfolding utterance. Second, the current findings are consistent with the hypothesis that segmental and suprasegmental properties of the speech signal are decoded in parallel (Cho et al., 2007; Tagliapietra & McQueen, 2010). Cho et al. (2007) proposed a Prosody Analyzer that is responsible for the computation of suprasegmental information. Working in parallel with prelexical mechanisms responsible for the extraction of segmental information, the Prosody Analyzer extracts suprasegmental information (including that which signals lexical stress) from the speech input, and builds a prosodic representation of the current utterance. This representation then constrains the word recognition process, along with segmental representations of the input. Because the segmental and suprasegmental analyzers use the same source of acoustic information, they are interconnected, the results of the two processes are inter-dependent, and the two types of representation are computed at the same time. Italian listeners thus appear to be processing segmental and stress information simultaneously.

Third, however, use of stress cues in lexical access is language specific. For example, it appears that English listeners do not depend heavily on suprasegmental stress information during spoken-word recognition (Cooper, Cutler, & Wales, 2002; Creel et al., 2006). English listeners certainly make use of other prosodic cues in the speech signal (e.g., information about the location of Intonational Phrase boundaries; Cho et al., 2007), but suprasegmental stress information appears to be relatively unimportant in English word recognition because stress in English is signaled segmentally. Unstressed vowels in English are usually reduced, and hence English listeners focus primarily on the segmental distinction between full vowels and reduced vowels (usually schwa; Cooper et al., 2002; Fear, Cutler, & Butterfield, 1995). Since segmental information is enough to recognize words efficiently, English listeners do not rely on suprasegmental lexical stress cues (Cooper et al., 2002; Cutler, 2005; Fear et al., 1995). In contrast, in a language such as Italian, where suprasegmental stress cues can be temporarily more informative than segmental cues, lexical stress information is used to optimize the recognition of spoken words.

Our findings also shed new light on the ongoing debate about how words are stored in the mental lexicon. We found that Italians have abstract knowledge about lexical stress – about the acoustic cues signaling antepenultimate stress and about the distributional bias favoring penultimate stress – and that this knowledge is used during lexical access. This first set of findings could be explained either in a model in which lexical representations are phonologically abstract (McClelland & Elman, 1986; Gaskell & Marslen-Wilson, 1997; Norris & McQueen,

2008) or in a model in which lexical representations are episodic traces (Goldinger, 1998; Pierrehumbert, 2002). In both cases, stress knowledge could modulate the word-recognition process (e.g., knowledge about the distributional bias could be derived over time either from abstract or episodic lexical representations, and then used on-line to influence lexical selection). Critically, however, the other findings from Experiment 2 cannot be explained by a purely episodic model of lexical representation. Recognition of antepenultimate-stress targets improved when full-cue versions of the newly-learned words were presented, even though the participants had never heard those acoustic versions before. A purely episodic model of the lexicon predicts that listeners should recognize a word better if the acoustically detailed input perfectly matches previously stored traces of that word, and hence that the reduced-cue versions of the new words – which had each been heard 72 times during the exposure phase – would be recognized better than previously unheard full-cue versions. Our results show that this was not the case.

A possible response in defense of the episodic position might be that the full-cue advantage for the antepenultimate-stress words arises because listeners recognize the new words by comparing them to episodic traces of known words with the same stress pattern – traces which do contain amplitude and duration cues. But this seems to be an unlikely possibility. First, as Magnuson et al. (2003) have shown, recognition of newly-learned words in the artificial-lexicon paradigm appears to be relatively unaffected by the similarity of those new words to specific well-known words. The fact that the artificial lexicon may be considered self-contained and not affected by competition from words in the listeners' native lexicon does not mean that the way newly-learned words are accessed cannot be affected by knowledge about the segmental and supra-segmental phonology of the native language (cf. Magnuson et al., 2003). But the effects of specific words on the recognition of words in an artificial lexicon appear to be limited. Second, if word recognition was done through a process of comparison to episodic traces, one would still have to predict that the strongest analogies would be between the current input and previous traces of the same word. That is, one would still expect the exposure episodes of the novel words to dominate in the comparison, and hence that there would be an advantage in the recognition of the reduced-cue versions of these words.

Alternatively, one might argue that listeners do not store episodes of words at the lexical level, but instead store episodes of fine-grained, sub-lexical phonetic details. These components, if available at a prelexical level, could then be used in the recognition of both well-known and newly-learned words, and in particular it would be possible for the listener to use generalizations made over prior episodes of known words (e.g., about the acoustic-phonetic properties of words with antepenultimate stress) in the recognition of new words. But this is not a theory about lexical representation.

The current results thus constitute a challenge for the view that the lexicon is composed solely of specific stored episodes of words. It would instead appear to be the case

that the lexicon is comprised of phonologically abstract representations. As new words are learned, knowledge about the phonological content of those words – including their segmental make-up and their stress pattern – comes to be stored in the mental lexicon. Specifically, listeners appear to be able to label a novel word as having particular phonemes and a certain stress pattern, and then use stored knowledge about the acoustic properties of other (existing) words which have phonological components that are labeled in the same way. It is this analytic capability (the ability to form abstract representations of the components of spoken words and use those representations in word recognition) that strictly episodic models of the lexicon lack (Cutler et al., 2010; McQueen et al., 2006). In Italian, it appears that new words with penultimate stress (like existing words of this type) are coded as having the dominant stress pattern, and hence that they can be recognized by default. For words with antepenultimate stress, however, they are coded during learning as such, and hence, when additional cues associated with this pattern are present in the input, and even though those cues have never been heard before in those words, those cues can nevertheless be used to facilitate word recognition. On this view, knowledge about the cues signaling antepenultimate stress is abstract too – it needs to be general (i.e., not word-specific) knowledge about antepenultimate stress for it to be applied to other words which share that structure.

It is important to note, however, that we are not advocating a strictly abstractionist model of spoken-word recognition. As we argued in the Introduction, models in which all episodic details are lost cannot explain the evidence that such details are retained in long-term memory (Goldinger, 1998; Nygaard, Sommers, & Pisoni, 1994). Creel, Aslin, and Tanenhaus (2008), for instance, have recently shown that listeners may use information about the talker's voice in word recognition. Using the same paradigm as we adopted in Experiment 2, they found that listeners fixated targets more when a target and its competitor had been spoken, during the learning phase, by different talkers than when the same stimuli had been spoken by the same talker. These results suggest that episodic details (such as those concerning talker voice) play an important role in word recognition, and even in the recognition of newly-acquired words.

The best framework to interpret our results is thus a hybrid model with both episodic and abstractionist components (Cutler et al., 2010; Goldinger, 2007). In such a model, an episodic memory system would store the idiosyncrasies of specific speech episodes. This system could then interact with both prelexical and lexical abstract representations and could be involved in the consolidation of new traces into abstract forms (Goldinger, 2007). On this view, then, word recognition is based on phonological abstraction, but that process is supported by an episodic memory system. The evidence that memories of episodic detail can influence word recognition (e.g., Creel et al., 2008; Goldinger, 1998; Nygaard et al., 1994) arises in this account not because those details are stored in the mental lexicon, but rather because they are stored elsewhere, in a manner that they can nevertheless influence word recognition.

We suggest that this episodic influence has its effect at the prelexical level. Previous research has suggested that, with respect to segmental information, abstraction is a prelexical process, such that abstract representations of speech sounds mediate between the speech signal and the mental lexicon (Cutler et al., 2010; McQueen et al., 2006; Sjerps & McQueen, 2010). These representations are flexible, allowing listeners to learn about idiosyncratic pronunciations (through exposure to talker-specific details; Eisner & McQueen, 2005). These representations may thus be based on episodic memories, and may be modulated by experience with specific talkers. Importantly, because they are prelexical and abstract, they support generalization of speech learning over the lexicon. Abstract knowledge about the component segments of words can thus benefit word recognition.

The current findings, along with those on the use of prior knowledge about the durational properties of prosodic words (Shatzman & McQueen, 2006), suggest that similar prelexical abstraction processes apply to the suprasegmental properties of the speech signal. Our findings show that listeners have abstract knowledge not only about the form of prosodic words (the relative durations of syllables in monosyllabic vs. polysyllabic words; Shatzman & McQueen, 2006), but also about other prosodic properties (lexical stress patterns). Because Italians have acquired abstract knowledge about stress in Italian – the penultimate-stress bias in trisyllabic words, and the acoustic properties associated with antepenultimate stress – they can bring that knowledge to bear when recognizing newly-learned words. As we suggested earlier, these cues could be extracted from the speech signal by a Prosody Analyzer (Cho et al., 2007), working in parallel with the prelexical segmental abstraction process.

We draw three related conclusions. First, Italian listeners have abstract knowledge about lexical stress. They

know that a distributional bias in trisyllabic words exists which favors penultimate stress, and they use that knowledge during the recognition of well-known and newly-learned spoken words. Moreover, they know that the uncommon pattern (antepenultimate stress) is revealed by specific sources of acoustic–phonetic information in the speech signal, and again they use that knowledge in the recognition of known and new words. Second, it appears that listeners extract and compute prosodic information at the same time as they compute segmental information. These two processes seem to occur in parallel, as the speech signal unfolds over time. This means, as we have shown, that stress information can sometimes be used to disambiguate Italian words before segmental disambiguation is available. Third, prosodic knowledge about lexical structure appears to be phonologically abstract rather than word-specific, suggesting in turn that lexical representations are abstract rather than episodic in nature. Prelexical processing may thus involve abstraction processes not only for segmental material (McQueen et al., 2006) but also for suprasegmental material.

### Acknowledgments

We thank Eva Reinisch for help with stimulus construction, data analysis, and helpful comments on a previous version of the manuscript, Holger Mitterer for help with data analysis, three anonymous reviewers for constructive feedback, and Giusy Turco for recording the stimuli.

### Appendix A. Target stress pairs used in Experiment 1.

	Frequency	No. letters		Frequency	No. letters
Abaco	0	5	aBAte	1.1	5
abacus			abbot		
Acaro	0	5	aCAcia	0	6
mite			acacia		
Acero	1	5	aCEto	1.3	5
maple			vinegar		
ALLuce	0.7	6	aLLUme	0.3	6
big toe			alum		
Asino	1	5	aSIllo	1.7	5
donkey			kindergarten		
ATtico	0.7	6	aTIguo	0.4	7
penthouse			adjacent		
COLlera	1.4	7	colLEga	2.2	7
anger			colleague		
CAlamo	0	6	caLAta	1.3	6
quill			invasion		

(continued on next page)

## Appendix A (continued)

	Frequency	No. letters		Frequency	No. letters
CANapa	0	6	caNAle	2.1	6
hemp			channel		
CANdido	1.3	7	canDlto	0.4	7
candid			candied		
CElebre	2.1	7	ceLEste	1.3	7
famous			pale blue		
CODice	2.3	6	coDIno	1.1	6
code			ponytail		
COMico	1.7	6	coMlzio	1.3	7
funny			meeting		
Eremo	0.9	5	eREde	1.8	5
hermitage			heir		
Estero	2.3	5	eSTEta	0	5
foreign			aesthete		
FEdera	0	6	feDEle	2	6
pillow case			faithful		
FORbice	0.6	7	forBlto	0.4	7
scissor			polished		
FRAgola	0.7	7	fraGOre	0.9	7
strawberry			uproar		
IMpeto	1.2	7	imPEro	1.9	7
impetus			empire		
LATtice	0.6	7	latTIna	0.8	7
Latex			can		
LOculo	0	6	loCUsta	0	7
burial niche			locust		
MACabro	1.1	7	maCAco	0	6
gruesome			macaque		
MAStice	0.6	7	maSTIno	0.3	7
putty			mastif		
MISsile	1.3	7	misSlva	0.9	7
missile			missive		
MONito	1.2	6	moNlle	0	6
warning			jewel		
PANico	1.8	6	paNIno	1.1	6
panic			sandwich		
PROtesi	1.6	7	proTEsta	2.2	7
prothesis			complaint		
REmora	0.3	6	reMOto	1.4	6
hesitation			remote		
SAlice	0.8	6	saLlva	1.4	6
willow			spittle		
SEnape	1.2	6	seNAto	2.3	6
mustard			senate		
TONaca	0.9	6	toNAle	0.9	6
habit			tonal		
ZIgomo	0	6	ziGote	0	6
cheekbone			zigote		

Note. The stressed syllables are in capital letters. Frequency is log transformed.

## References

- Albano Leoni, F., Maturi, P. (1998). *Manuale di fonetica*. Roma, Italy: Carocci.
- Alfano, I. (2006). La percezione dell'accento lessicale: Un test sull'italiano a confronto con lo spagnolo. In R. Savy & C. Crocco (Eds.), *Atti del II AISV. Associazione Italiana di Scienze della Voce* (pp. 632–656). Padova, Italy: EDK Editore.
- Alfano, I., Savy, R., & Llisterri, J. (2009). Sulla realtà acustica dell'accento lessicale in italiano ed in spagnolo: La durata vocalica in produzione e percezione. In L. Romito, V. Galatà, & Lio (Eds.), *Atti del IV AISV. Associazione Italiana di Scienze della Voce* (pp. 22–39). Torriana, Italy: EDK Editore.



- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412.
- Barr, D. J. (2008). Analyzing 'visual world' eye tracking data using multilevel logistic regression. *Journal of Memory and Language*, 59, 457–474.
- Bertinetto, P. M., Burani, C., Laudanna, A., Marconi, L., Ratti, D., Rolando C., & Thornton, A. M. (2005). *Corpus e Lessico di Frequenza dell'Italiano Scritto (CoLFIS)* (CoLFIS. Corpus and Frequency Lexicon of Written Italian). <[www.istc.cnr.it/material/database/](http://www.istc.cnr.it/material/database/)>.
- Boersma, P., & Weenink, D. (2007). PRAAT, a system for doing phonetics by computer (version 5.1.32) (computer program). <<http://www.praat.org>>.
- Cho, T., McQueen, J. M., & Cox, E. A. (2007). Prosodically-driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics*, 35, 210–243.
- Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech*, 45, 207–228.
- Creel, S., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition*, 106, 633–664.
- Creel, S. C., Tanenhaus, M. K., & Aslin, R. N. (2006). Consequences of lexical stress on learning an artificial lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 15–32.
- Cutler, A. (2005). Lexical stress. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 264–289). Oxford: Blackwell.
- Cutler, A., Eisner, F., McQueen, J. M., & Norris, D. (2010). How abstract phonemic categories are necessary for coping with speaker-related variation. In C. Fougere, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory phonology* (Vol. 10, pp. 91–111). Berlin, Germany: de Gruyter.
- Cutler, A., & Van Donselaar, W. (2001). Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech*, 44, 171–195.
- Cutler, A., Wales, R., Cooper, N., & Janssen, J. (2007). Dutch listeners' use of suprasegmental cues to English stress. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the XVI international congress of phonetic sciences* (pp. 1913–1916). Dudweiler, Germany: Pirrot.
- Davis, M. H., Di Betta, A. M., Macdonald, M. J. E., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience*, 21, 803–820.
- Dumay, N., & Gaskell, G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, 18, 35–39.
- Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, 67, 224–238.
- Fear, B. D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. *Journal of the Acoustical Society of America*, 97, 1893–1904.
- Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes*, 12, 613–656.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251–279.
- Goldinger, S. D. (2007). A complementary-systems approach to abstract and episodic speech perception. In *Proceedings of XVI international congress on phonetic sciences* (pp. 49–54). Dudweiler, Germany: Pirrot.
- Huetig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, 57, 460–482.
- Krämer, M. (2009). *The phonology of Italian*. New York: Oxford University Press.
- Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, 132, 202–227.
- Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, 53, 372–380.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86.
- McQueen, J. M., Cutler, A., & Norris (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, 30, 1113–1126.
- McQueen, J. M., & Viebahn, M. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology*, 60, 661–671.
- Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115, 357–395.
- Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech-perception as a talker-contingent process. *Psychological Science*, 5, 42–46.
- Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology VII* (pp. 101–140). Berlin, Germany: Mouton de Gruyter.
- Reinisch, E., Jesse, A., & McQueen, J. M. (2010). Early use of phonetic information in spoken word recognition: Lexical stress drive eye movements immediately. *The Quarterly Journal of Experimental Psychology*, 63, 772–783.
- Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, 90, 51–89.
- Shatzman, K. B., & McQueen, J. M. (2006). Prosodic knowledge affects the recognition of newly-acquired words. *Psychological Science*, 17, 372–377.
- Sjerps, M. J., & McQueen, J. M. (2010). The bounds on flexibility in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 36, 195–211.
- Soto-Faraco, S., Sebastian-Galles, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language*, 45, 412–432.
- Sulpizio, S., & McQueen, J. M. (in press). When two newly-acquired words are one: New words differing in stress alone are not automatically represented differently. In *Proceedings of interspeech 2011*, Florence, Italy.
- Tagliapietra, L., & Tabossi, P. (2005). Lexical stress effects in Italian spoken word recognition. In *Proceeding in XXVII conference of the cognitive science society* (pp. 2140–2144).
- Tagliapietra, L., & McQueen, J. M. (2010). What and where in speech recognition: Geminate and singletons in spoken Italian. *Journal of Memory and Language*, 63, 306–323.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634.
- Thornton, A. M., Iacobini, C., & Burani, C. (1997). *BDVBD Una base di dati sul vocabolario di base della lingua italiana* (BDVBD: A database for the Italian basic dictionary). Roma, Bulzoni.
- van Donselaar, W., Koster, M., & Cutler, M. (2005). Exploring the role of lexical stress in lexical recognition. *Quarterly Journal of Experimental Psychology*, 58, 251–273.
- Warren, P., & Marslen-Wilson, W. D. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics*, 4, 262–275.