Pitch accent type matters for online processing of information status: Evidence from natural and synthetic speech*

AOJU CHEN, ELS DEN OS AND JAN PETER DE RUITER

Abstract

*Adopting an eyetracking paradigm, we investigated the role of H*L, L*HL, L*H, H*LH, and deaccentuation at the intonational phrase-final position in online processing of information status in British English in natural speech. The role of H*L, L*H and deaccentuation was also examined in diphone-synthetic speech. It was found that H*L and L*HL create a strong bias towards newness, whereas L*H, like deaccentuation, creates a strong bias towards givenness. In synthetic speech, the same effect was found for H*L, L*H and deaccentuation, but it was delayed. The delay may not be caused entirely by the difference in the segmental quality between synthetic and natural speech. The pitch accent H*LH, however, appears to bias participants' interpretation to the target word, independent of its information status. This finding was explained in the light of the effect of durational information at the segmental level on word recognition.*

* This research was supported by the COMIC project (IST-2001-32311) during the period from April 2004 to March 2005 at the Max Planck Institute for Psycholinguistics. We thank Dephine Dahan, Claudia Kuzla, Barbara Schmiedtová, and Michael White for helpful comments on the experimental design, Carlos Gussenhoven for recording the stimuli in natural speech, Rob Clark and Stefan Rossignol for their help with preparing the synthetic stimuli, Doug Davidson, Herbert Baumann, John Nagengast, Leah Roberts, Keren Shatzman, Johan Weustink, and Anna Zumach for their help with setting up the experiment, Pim Levelt and Antje Meyer for making it possible to conduct the experiments at the University of Birmingham, Linda Mortensen for her support during the testing and useful comments on an earlier version of the text, Yang Luo for his help with automating the data processing procedure, and Holger Mitterer for useful discussion on the statistical analyses.

The Linguistic Review 24 (2007), 317–344
DOI 10.1515/TLR.2007.012
0167–6318/07/024-0317

1. Introduction

Information conveyed by a sentence or a sentence constituent changes its status typically from new to given as discourse proceeds. Speakers can use intonation to signal changes in information status by varying the intonation of the corresponding lexical entities. It is generally accepted that in Germanic languages the placement of pitch accent is crucial for the marking of information status (Gussenhoven 2006). That is, new information tends to be accented, but given information deaccented. Appropriate intonational encoding of information status can facilitate processing of information while inappropriate intonational encoding of information status has the opposite effect (Cutler, Dahan, and van Donselaar 1997 and references therein).

In contrast, the role of *type of pitch accent* in processing information status is far from clear. Previous studies of the interaction between accent placement and information status in English (e.g., Birch and Clifton 1995; Dahan, Tanenhaus and Chambers 2002) and Dutch (e.g., Nooteboom and Terken 1982) often assumed that different types of pitch accents function in the same way and left the motivation to include one pitch accent type into the investigation, but not the other, unexplained. There is some empirical evidence in recent studies suggesting that different types of pitch accents are used to convey different types of information status. Specifically, in a perception experiment in which listeners judged the appropriateness of H*, H+L* and deaccentuation when followed by L-% in German in a context where the information status was varied, Bauman and Hadelich (2003) found that both H* and H+L* were considered appropriate in marking new information (i.e., the referent, defined as a noun depicting an object, was introduced neither visually nor auditorily earlier), with H* being more favored. Further, deaccentuation was judged to be most suitable for the signaling of given information (i.e., the referent was earlier introduced auditorily). Evidence in favor of H+L* as the "accessible" accent was provided in another perception experiment in which Bauman and Grice (2006) investigated the appropriateness of H+L*, H* and deaccentuation in the marking of inferentially accessible referents (i.e., referents who either constituted a part of an already mentioned whole or were predictable from the contextually given schema or frame) in German.

Against this backdrop, the present study set out to investigate the role of four nuclear (i.e., intonational phrase-final) pitch accent types, fall, rise-fall, rise, fall-rise, as well as deaccentuation in processing given vs. new information in Southern British English. Given information is defined as information conveyed by a referent that was mentioned previously in the discourse; new information is defined as information conveyed by a referent that was not previously mentioned or only indirectly touched upon (e.g., via semantic relatedness). The four types of pitch accents are known as H*L (fall), L*HL (rise-fall),

L*H (rise) and H*LH (fall-rise) in the Transcription of Dutch Intonation notation (ToDI) (Gussenhoven 2005). ToDI is adopted to describe pitch contours instead of ToBI (Tones and Break Indices – see Beckman and Ayers 1994) in this study for two reasons. First, the intonational grammars of British English and Dutch are very similar and the phonological categories proposed for Dutch in ToDI also exist in British English (Grabe 2004). Second, ToDI does not have leading tones in pitch accents and employs only one phrase-type, i.e., intonational phrase, both of which are rooted in the British English tradition (Gussenhoven 2005). The ToDI notation is therefore believed to reflect more closely than ToBI the nuclear tones in British English discussed in earlier analyses of intonational meaning, which serve as the starting point of our investigation. At the nuclear position, the boundary tone is the same as the trailing tone of the pitch accent in the case of H*L, L*HL and L*H. As for H*LH, the high trailing tone is realized as the boundary tone. The nuclear contours can thus be transcribed as H*L L%, L*HL L%, L*H H% and H*L H% in ToDI. Their counterparts in ToBI are H* L-L%, L*+H L-L%, L* H-H%, and H* L-H%, respectively (Gussenhoven 2005).[1] These four types of pitch accents were chosen for three reasons. First, H*L, L*H and H*LH are common nuclear pitch accent types in Southern Standard British English (Gussenhoven 1984, 2002; Grabe 2004). Second, they have been claimed to convey information status in theories of English intonational meaning. There is, however, no consensus on the exact functions of these pitch accents. Finally, the L*HL accent is claimed to function like an emphatic H*L (Brazil 1975; Gussenhoven 1984, 2002). Including L*HL allowed us to inspect the assumed gradient meaning difference between H*L and L*HL. The role of these pitch accents was examined in both natural speech (Experiment 1) and synthetic speech (Experiment 2). The use of synthetic speech was intended to find out whether effects of pitch accent type would be preserved when the segmental quality of the speech is limited.

In Section 2 we will first give a brief review of the postulated relations between pitch accent types (H*L, L*HL, L*H, and H*LH), and information status in theories of English intonational meaning, and then propose our hypotheses on the role of these pitch accents as well as deaccentuation in processing given vs. new information in British English. Experiment 1 will be reported in Section 3 and Experiment 2 in Section 4. A general discussion of findings from both experiments will be given in Section 5.

1. The four pitch accent types in ToDI are merged to three in ToBI (i.e., H*, L*+H, and L*). H*L and H*LH have the same starred tone in ToBI but differ in following phrasal tones.

2. Theoretical background and hypotheses

Four analyses of English intonational meaning will be reviewed in this section: Brazil (1975), Gussenhoven (1984, 2002), Pierrehumbert and Hirschberg (1990), and Steedman (2000). The British nuclear tone system was used to describe intonation in the first two analyses. Where possible, we give the ToDI labels of the intonation contours in brackets. The ToBI notation (Beckman and Ayers 1994) was used in the other two analyses and is maintained in the review.

According to Brazil (1975), the speaker makes a moment-by-moment assessment of the understanding he shares with the hearer, and "by choosing one intonation pattern rather than another, the speaker can affect what an utterance does towards achieving convergence" (1975: 3). Brazil proposed three speaker-options: (1) Proclaiming: the speaker presents what he says as new information; (2) Referring: the speaker makes references to features which he takes to be already present in the interpreting worlds of the speaker and the hearer; (3) Neutral: the speaker avoids proclaiming or referring, i.e., withdrawing himself from the interactive situation. These three options are signaled by five nuclear tones. Proclaiming tones are fall (H*L) and rise-fall (L*HL). Referring tones include fall-rise (H*LH) and high rise. Rise-fall and high rise have the effect of intensifying the meaning they signal. The neutral tone is low rise (L*H).

Following Brazil (1975), Gussenhoven (1984, 2002) argued that in a conversation, the speaker and the hearer strive towards some common understanding about a particular segment of the world and the speaker may achieve this goal in three ways: (1) Addition: adding the Variable (i.e., the information that the speaker contributes to the conversation) to the background, comparable to Brazil's proclaiming; (2) Selection: selecting a Variable from the background, comparable to Brazil's referring; or (3) Testing: choosing not to commit himself as to whether the Variable belongs to the background. Addition is conveyed by fall (H*L L%), selection by fall-rise (H*L H%), and testing by low rise (L*H H%). Note that Gussenhoven used nuclear tone to refer to both the pitch accent and the boundary tone. These tones were considered the basic nuclear tones of English. All the other tones are modifications of them. The modification relevant to us here is delay, i.e., postponing the association of the tone with the segment. This resulted in the delayed fall (L*HL L%), the delayed fall-rise (L*HL H%), and the delayed low rise (L* H%). Each delayed tone was claimed to signal the same meaning as the corresponding basic nuclear tone but with an extra meaning element, i.e., non-routineness.

In line with Brazil (1975) and Gussenhoven (1984, 2002), Pierrehumbert and Hirschberg (hereafter P&H) (1990) proposed that the choice of pitch contour largely conveys how the speaker evaluates his contribution to the discourse with respect to some mutual beliefs between the speaker and the hearer(s). Different from Brazil (1975) and Gussenhoven (1984, 2002), P&H's analysis assumed

strong compositionality in the meaning of the pitch contour, according to which each type of components (i.e., pitch accent, phrase accent, and boundary tone) of the pitch contour is interpreted with respect to its distinct phonological domain and contributes a distinct type of information to the overall interpretation of a contour. The type of components that conveys information about the status of individual discourse referents is pitch accent. Here we mention briefly the postulated functions of H*, L*, and L*+H, which are relevant to the pitch accents under investigation (see footnote 1). Pitch accents consisting of H* mark lexical items that should be treated as new in the discourse. Pitch accents consisting of L* mark lexical items that are not to be treated as new, but nevertheless are salient in the discourse. Within this group of pitch accents, L*+H signals a lack of speaker commitment to a scale that links the accented item to other items salient in the hearer's mutual beliefs. Phrase accent and boundary tone convey the degree of relatedness between intermediate phrases and intonational phrases respectively, with the high tone emphasizing relatedness and the low tone independence.

Different from the three previous analyses, Steedman (2000) divided an utterance into theme and rheme. A theme is what the speaker and the hearer(s) have agreed to talk about, the part of the sentence that ties it to the previous discourse; a rheme is the speaker's new contribution on the subject of the theme. Both the theme and the rheme can be marked or unmarked. Marked information is either new (in the case of rheme) or contrastive (in the case of theme); unmarked information is neither. Marked words in rhemes generally receive H*, but can also receive L*, and possibly H*+L, and H+L*. Marked words in themes generally receive L+H*, and possibly L*+H in responses where contradiction is involved. Following P&H (1990), Steedman argued that the meanings of the pitch accents remain the same when followed by different phrasal tones.

As may have become clear, these theories make different claims on the functions of the nuclear pitch accents at issue in marking information status. According to the theories rooted in the British English tradition (i.e., Brazil 1975; Gussenhoven 1984, 2002), H*L and L*HL mark new information whereas H*LH marks given information. The two analyses differ in the meaning of L*H, which is "neutral" according to Brazil (1975) but "testing" according to Gussenhoven (1984, 2002). Opposite predictions can be derived for L*HL (L*+H L-L% in ToBI) and H*LH (H* L-H% in ToBI) from P&H (1990) operating on the ToBI system. As phrasal boundary tones are considered irrelevant to the conveyance of the information status of individual discourse referents, pitch contours with the same pitch accents but different phrasal tones have the same meaning (e.g., H* L-L% and H* L-H%) and pitch contours with different pitch accents but the same phrasal tones have different meanings (e.g., H* L-L% and L*+H L-L%). It follows that L*HL marks given information and H*LH marks new information. Steedman (2000) differed from P&H (1990) in

claiming that L*H (L* H-H% in ToBI) conveys new information and that L*HL (L*+H L-L% in ToBI) conveys given information involving contradiction.

Because we are dealing with pitch accent types in British English, we derived the following working hypotheses on the role of nuclear H*L, L*HL, L*H, and H*LH in processing information status mainly from Brazil (1975) and Gussenhoven (1984, 2002):

a. H*L triggers the interpretation of newness. (All the theories reviewed)
b. L*HL triggers the interpretation of newness with more emphasis than H*L. (Brazil 1975; Gussenhoven 1984, 2002)
c. L*H triggers the interpretation of givenness, like deaccentuation. (P&H 1990)
d. H*LH triggers the interpretation of givenness, like deaccentuation. (Brazil 1975; Gussenhoven 1984, 2002)

3. Experiment 1 – natural speech

3.1. Method

The eye-tracking paradigm used in Dahan, Tanenhaus and Chambers (2002) was adopted to evaluate our hypotheses in natural speech. Dahan, Tanenhaus and Chambers examined the role of accent placement in reference resolution by monitoring eye fixations to lexical competitors (e.g., *coat* and *comb*) as participants followed pre-recorded instructions to move objects displayed on a computer screen using a computer mouse. Each display contained four objects and four geometric shapes, as illustrated in Figure 1. It was found that the effect of accent placement was reliably reflected in the proportion of fixations to the referent and its lexical competitor in a selected time window. The eye-tracking paradigm may thus offer a measure of the effect of pitch accent type on the processing of information status.

3.1.1. Experimental design. On each experimental trial, two of the objects had names that were phonemically related, i.e., sharing the same stressed syllable (e.g., *candle* vs. *candy*), or the same onset-peak cluster (e.g., *comb* vs. *coat*) if the words were monosyllabic. One served as the target (e.g., *comb*) and the other as the competitor (e.g., *coat*). Each trial consisted of two consecutive instructions (see Table 1). The second instruction always mentioned the target (e.g., *now put the comb below the diamond*). The first instruction mentioned either the target (e.g., *Put the comb below the triangle)*, marking the target at the onset of the second instruction as "given" but the competitor as "new", or the competitor (e.g., *Put the coat below the triangle*), marking the target at the onset of the second instruction as "new" but the competitor as

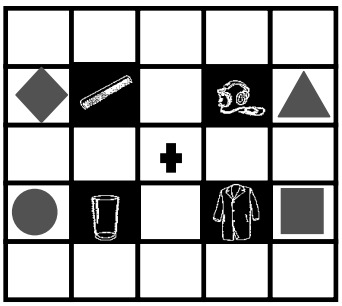

Figure 1. Example of a visual display. Geometric shapes were blue.

"given". The target noun in the second instruction was temporarily ambiguous during the segments it has in common with the competitor noun. At that stage, both the target and competitor nouns were potential candidates for selection, and participants were expected to make use of intonation to identify the noun (Dahan, Tanenhaus and Chambers 2002). The intonation of the first instruction was kept the same throughout the experiment; the intonation of the second instruction was varied by having the target noun produced with H*L, L*HL, L*H, H*LH and deaccentuation with an intonational phrase boundary after the noun. Combining the two types of information status of the target/competitor during the second instruction and the five accent conditions gave us ten experimental conditions, as illustrated in Table 1.

3.1.2. Predictions. The patterns of fixations to the competitor picture and the target picture from the target word onset to the identification of the target word during the second instruction have been used as indicators to how intonation affects the interpretation of information status (Dahan, Tanenhaus and Chambers 2002). In line with this method, we arrived at the following predictions:

(1) When the target/competitor is new (i.e., not previously mentioned), accent conditions conveying newness (e.g., H*L, L*HL) will trigger more fixations to the target/competitor picture than accent conditions conveying givenness (e.g., L*H, H*LH, deaccentuation).

(2) When the target/competitor is given (i.e., previously mentioned), accent conditions conveying givenness (e.g., L*H, H*LH, deaccentuation) will trigger more fixations to the target/competitor than accent conditions conveying newness (e.g., H*L, L*HL).

Table 1. Illustration of the ten experimental conditions in Experiment 1

First instruction	Second instruction	Information status
Put the *coat* above the triangle (competitor)	Now put the *comb* below the diamond (target) H*L L% L*HL L% L*H H% H*L H% deaccentuation L%	New target Given competitor
Put the *comb* above the triangle (target)		Given target New competitor

3.1.3. Materials. Twenty pairs of phonemically similar nouns served as the materials for experimental trials, of which eighteen pairs were also used in Dahan, Tanenhaus and Chambers (2002). As pitch accents are realized differently in monosyllabic words than in disyllabic words, to minimize effects related to phonetic realization of pitch accent, we included twelve pairs of monosyllabic words and eight pairs of disyllabic words. One member of each pair was assigned the role of target, the other the role of competitor. In the case of the monosyllabic pairs, care was taken to have a similar distribution of voiced codas and voiceless codas in the targets and competitors. The mean lexical frequencies of the targets (33.6 per million) and competitors were similar (27.5 per million), as reported in Francis and Ku_era (1982). Each of the 20 target-competitor pairs was associated with two distractor nouns, resulting in four pictures on each display (see Figure 1). Two target-competitor pairs were assigned to each experimental condition by means of a Latin Square. This led to ten lists of experimental stimuli.

In addition to the 20 experimental trials, 48 filler trials were constructed to prevent participants from developing the expectation that pictures with phonemically similar names were likely to be moved in either instruction.

Combining the ten lists of experimental stimuli and the fillers gave us 10 stimulus lists. To minimize order effects, two stimulus orders were created for each stimulus list.

The 272 (20 experimental trials $\times$ 4 + 48 filler trials $\times$ 4) pictures were selected from Snodgrass and Vanderwart's (1980) picture database and the picture database of the Max Planck Institute for Psycholinguistics (MPI). All were black and white line drawings.

The spoken instructions were recorded by a prosodically trained male speaker of Southern Standard British English at 48 kHz sampling rate in the soundproof studio at the MPI. The speaker read the instructions from printed recording script (see (3) for an example). The intonation for each instruction was transcribed in the ToDI notation. The speaker was an expert on ToDI and familiar with producing pitch contours on request. Figure 2 shows example f_0 tracks for the target word *comb* produced in all five accent conditions.

(3) Put the comb above the square; now put the comb below the diamond.
%L H* H*L H*L L% %LH* H*L H% H*L H*L L%

The rise in L*HL and the fall in H*LH may be realized largely on the segments present in both the target and the competitor. If L*HL were found to have the same effect as L*H, and H*LH were found to have the same effect as H*L, this might have been caused by the ambiguity in the realization of the pitch

Table 2. f_0 values of the shared part of the pitch accents

	H*L	H*LH	L*H	L*HL
Mean maximal F0 (Hz)	*166*	*179*	156	158
Mean minimal F0 (Hz)	80	78	*79*	*98*

accents. To establish that L*H was acoustically distinguishable from L*HL in the rise, and H*L was acoustically distinguishable from H*LH in the fall, we measured maximal f_0 and minimal f_0 in the rise and the fall. As can be seen in Table 2 and Figure 2, L*H rose from a significantly lower pitch point than L*HL ($t = 4.669$, $df = 12$, $p < .005$), whereas H*LH fell from a significantly higher pitch point than H*L ($t = 4.128$, $df = 10$, $p < .005$). [2]

3.1.4. Procedures. Twenty-four undergraduates and two postgraduates from the School of Psychology at the University of Birmingham participated in the experiment. They all spoke Southern British English as their only native language. None of them reported to have hearing problems. They received either course credits or a small fee for their participation. The experiment took about 10 minutes.

Participants were tested individually. The experimenter described first briefly to the participants what they were supposed to do and then gave them the written instructions on the experimental task to read. An example of the visual display was also included in the written instructions. Participants were seated at a comfortable distance from the computer screen in a quiet room. The eye tracker was mounted and calibrated. Eye movements were monitored with a portable SR EyeLink II eye-tracking system. Spoken instructions were presented to the participants through headphones. The structure of a trial was as follows: first, a central fixation point appeared on the screen for 500 ms. Then, a 5×5 grid with four pictures and four geometric shapes appeared on the screen, as the auditory presentation of the first instruction was initiated. The positions of the pictures were randomized across four fixed positions of the grid, while the geometric shapes appeared in fixed positions on every trial. As soon as a picture was moved after the first instruction ended, the second instruction was initiated. Once the participant moved a picture following the second instruction, the next trial began. The position of the mouse cursor on the computer screen was sampled and recorded, along with the eye-movement data. A cen-

2. The analyses were performed on maximal f_0 obtained from 11 target words and minimal f_0 obtained from 13 target words. For the other target words, no reliable measurements could be taken because the rise and the fall were only partially visible in f_0 tracks.

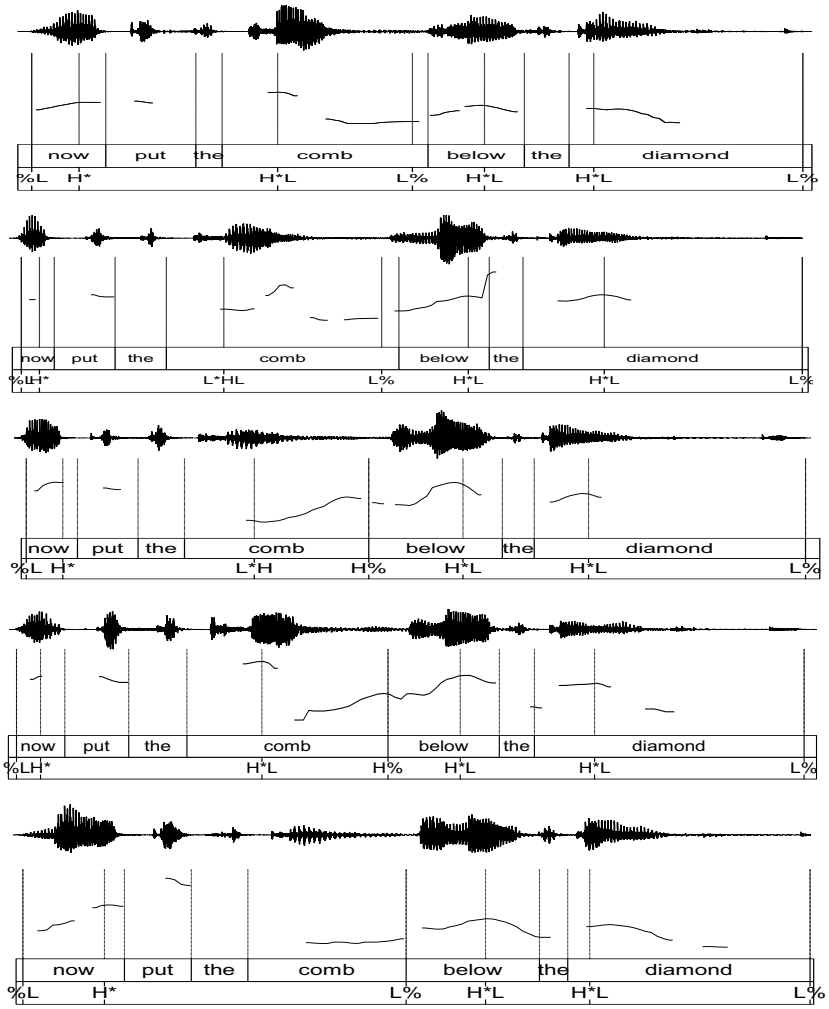

Figure 2. f_0 values of the shared part of the pitch accents

tral fixation point appeared on the screen after every five trials, which allowed automatic drift correction in the calibration.

Two participants were randomly assigned to each of the 10 stimulus lists; one of them received the stimulus list in stimulus order 1 and the other in stimulus order 2. In six cases, the eye movement data were not properly sampled due to technical failure (4 participants), incorrect interpretation of instructions

(1 participant), and difficulty in recognizing the picture due to poor eyesight (1 participant). For each of these six cases, a new participant was recruited and tested. The total number of participants thus amounted to 26. At the end of the experiment, participants were asked to fill out a questionnaire on their language background.

3.1.5. Coding procedure. Data from the above-mentioned six participants were excluded from coding. Data from the other 20 participants were coded for fixations. For 16 of these participants, data from the right eye were coded; for four of these participants, data from the left eye were coded because of calibration problems with the right eye. On each trial, the duration of a fixation was established relative to the onset of the target word in the second instruction. The graphical analysis software SUSI developed at the MPI was used to do the mapping between the position of fixations, the mouse movements, and the pictures presented on each trial, and to display them simultaneously. Each fixation was represented by a dot associated with a number, indicating the order in which the fixations occurred. The onset and duration of fixation were specified for each fixation point.

For each experimental trial, fixations were coded from the onset of the target word in the second instruction (including closure for initial voiceless consonants) to the moment when participants clicked on the target picture with the mouse, which was taken to reflect participants' confident identification of the target word (Salverda, Dahan, and McQueen 2003). Fixations directed to the target picture, to the competitor picture, to the distractor pictures, and to any other locations on the screen were coded. Fixations falling within the cell of the grid in which a picture was presented or on the edge of that grid were coded as fixations to that picture.

3.2. Results

The coded data from two participants were excluded from further analysis because few fixations were launched before the end of the target word. The proportion of fixations to each location (i.e., target picture, competitor picture, distractor pictures, and elsewhere) was calculated in 33 ms time intervals (Dahan, Tanenhaus and Chambers 2002) for each condition and each of the 18 participants, by dividing the total number of trials in which a location was fixed during a specific time interval by the total number of trials in which a fixation was launched to any location in this time interval (Salverda, Dahan, and McQueen 2003). As the minimal latency to plan and launch a saccade is 200–300 ms in tasks like visual search (Hallett 1986; Viviani 1990), fixations realized in the first 300 ms of the target word were likely to be related to speech input preced-

ing the target word. Because the part of the target word that overlapped with the competitor was between 290 ms (in the deaccentuation condition) and 360 ms (in the H*LH condition) long on average, the effects of accent conditions were expected to be observable in the time window from 300 ms to 700 ms.

Figure 3 (top) presents the mean proportions of fixations to the competitor (e.g., *coat*) when it was a *given* entity (e.g., *Put the coat above the square; now put the comb below the diamond.*) at the onset of the second instructions in 33 ms time intervals from 0 to 1023 ms after the onset of the target word in the second instructions. Over the 0–300 ms time window, the mean proportions of fixations were generally low ($0.1 \sim 0.25$) and only differed marginally in different accent conditions. Over the 300–700 ms time window, in the L*H L accent condition, the mean proportion of fixations remained low ($0.18 \sim 0.23$) before starting to decrease around 560 ms. This pattern is consistent with the hypothesis that L*HL conveys newness and therefore triggers fewer fixations to the "given" competitor. In contrast, in the L*H, deaccentuation and H*L conditions, the mean proportion of fixations increased considerably before it started to decrease. At first sight, this pattern seems to accord with our hypothesis that the L*H and deaccentuation conditions convey givenness and therefore trigger an increase in fixations to the "given" competitor, but it does not agree with our hypothesis that H*L conveys newness. However, a close inspection of the time point of the decrease in the three conditions revealed an unexpected temporal difference in the patterns of fixation proportions. In particular, the decrease began observably earlier in the H*L condition (at 430 ms) than in the L*H (at 529 ms) and deaccentuation conditions (at 562 ms). As the decrease in fixations to the competitor indicates the recognition of the target word (e.g., *comb*), an earlier decrease thus means an earlier recognition of the target word. The accent conditions (i.e., L*H and deaccentuation), creating a bias towards givenness, would keep the participants' attention longer to the "given" competitor and delay the recognition of the target word longer than the 'new' accent condition (i.e., H*L). This is exactly what we found here. Our data are thus consistent with the hypothesis that H*L conveys newness whereas L*H and deaccentuation convey givenness. Note that the decrease started comparatively later in the deaccentuation condition than in the L*H condition. This may suggest that deaccentuation creates a stronger bias towards givenness than L*H.

The effect of the hypothetical "given" pitch accent H*LH is, however, unexpected. The proportion of fixations started to increase around 264 ms but did not go higher than 0.34. It began to decrease as early as at 364 ms. The overall relatively low proportion of fixations and the early decrease suggest that H*LH functioned like a newness accent instead.

Figure 3 (bottom) presents the mean proportions of fixations to the competitor (e.g., *coat)* when it was a *new* entity (e.g., *Put the comb above the square; now put the comb below the diamond.*). As can be seen, the mean

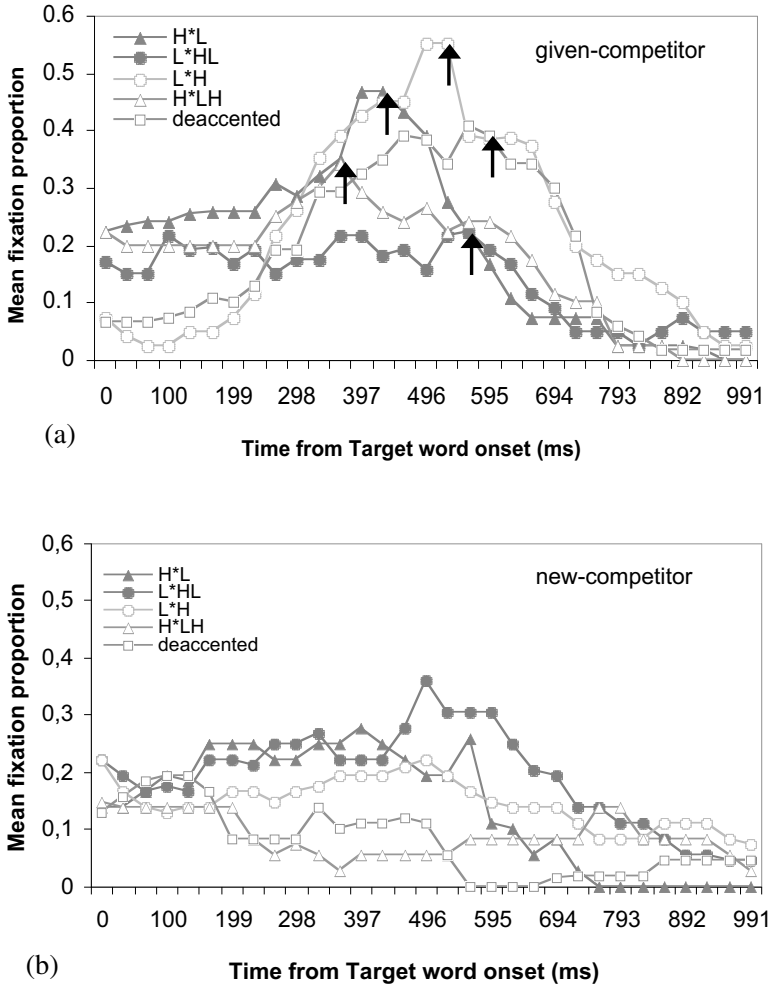

Figure 3. Proportions of fixations (averaged across 18 participants) to the competitor picture from the onset of the target word in the second instruction in the 'given' competitor condition (top) and in the 'new' competitor condition (bottom). The arrows mark the time point of the decrease in fixation proportion in each accent condition.

proportion of fixations to the *new* competitor was relatively low in all accent conditions (< 0.27) in the time window from 300 ms to 700 ms, possibly

caused by a general bias towards a given entity (e.g., *comb*) in the participants. There were, nevertheless, noticeable differences in different conditions: H*L – 0.199, L*HL – 0.266, L*H – 0.172, H*LH – 0.064, and deaccentuation – 0.069. The higher mean proportion of fixations in the H*L and L*HL conditions than in the L*H, H*LH and deaccentuation conditions is consistent with the hypothesis that H*L and L*HL mark newness, but L*H and H*LH, like deaccentuation, mark givenness. Further, the pattern of fixation proportions changed across time in different conditions. Largely, there was a decrease trend in the proportion of fixation in the time window from 400 ms to 600 ms. This trend had different temporal properties in different accent conditions, though it was not straightforward to establish where the descending trend exactly started.

To evaluate the observed effects of accent conditions within a certain time interval and across time, we conducted two repeated measures ANOVAs with two variables at a significance level of 0.05: Time Interval (12 levels: 300 ms to 700 ms in 33 ms interval), and Accent Condition (5 levels: H*L, L*HL, L*H, H*LH, deaccentuation). The dependent variables were the mean proportion of fixations to the "given-competitor" and the mean proportion of fixations to the "new-competitor" respectively. When the competitor was a given entity, the analysis revealed a main effect of Time Interval ($F(11,187) = 3.588$, $p < .05$, partial $\eta^2 = .174$) and a significant interaction of Accent Condition $\times$ Time Interval ($F(44,748) = 1.81$, $p < .05$, partial $\eta^2 = .096$). These results confirm that the patterns of fixations in different accent conditions changed across time, and they differed mainly in the time point of a sharp decrease in the fixation proportion to the "given-competitor". When the competitor was a new entity, the analysis only revealed a main effect of Accent Condition ($F(4,68) = 2.881$, $p < .05$, partial $\eta^2 = .145$). This result confirms that the mean proportion of fixations differed in different conditions. The nonsignificance of the effect of the variable Time Interval indicates that the temporal differences in the patterns of fixations in the new-competitor condition were probably not consistent across trials and participants. The asymmetry in the results between the "given-competitor" condition and the "new-competitor" condition may be accounted for by the fewer fixations launched in general when the competitor was new (because of the general bias towards a given entity).

A temporal difference in the patterns of fixations across accent conditions similar to that found in the "given-competitor" condition was also observable in the time point of a sharp increase in the fixation proportion to the target in both the "new-target" condition and the "given-target" condition. The difference was however not statistically significant. This asymmetry in results may be related to the fact that the phonemically similar part between the target and the competitor came from the target word. This might have biased its lexical interpretation towards the target word because of coarticulation effects. Con-

sequently, the effect of accent condition and its interaction with time interval became less strong on fixation proportions to the target.

3.3. Discussion

The results show that pitch accent type plays a role in the online processing of information status at the intonational phrase-final position in natural speech. The effect of pitch accent type can clearly be seen in the pattern of fixations to the competitor picture in the selected time window from 300 ms to 700 ms. It can be reflected in the mean proportion of fixations to the competitor, as predicted. Specifically, when the competitor (e.g., *coat*) was new (e.g., *Put the comb above the square; now put the comb below the diamond.*), the hypothetical "new" accents H*L and L*HL triggered a higher proportion of fixations to the competitor picture than the other accent conditions. Interestingly, the effect of pitch accent type can also be reflected in how fast the participants could recognize the target word during the early presentation of the target word and shift their visual attention accordingly away from the "wrong" picture or launch more fixations to the "right" picture. Specifically, when the competitor was a given entity, the proportion of fixations to the competitor increased initially in most accent conditions but started to decrease substantially earlier in the H*L condition (a hypothetical "new" accent) than in the L*H (a hypothetical "given" accent) and deaccentuation conditions.

Our findings are in agreement with the hypothesis that H*L triggers the interpretation of newness, as may be derived from all theories of intonational meaning reviewed in section 2. Further, we have found that L*HL triggers the interpretation of newness, lending support to Brazil (1975) and Gussenhoven (1984, 2002). There are also indications that L*HL creates a stronger bias towards newness than H*L, as suggested by Brazil (1975) and Gussenhoven (1984, 2002). When the competitor was a given entity, the mean proportion of fixation was lowest in the L*HL condition at nearly every time point in the selected time window. When the competitor was a new entity, the proportion of fixation averaged over the selected window was the highest in the L*HL condition. In addition, we have found that L*H, like deaccentuation, triggers the interpretation of givenness, as suggested by P&H (1990), but contra Brazil (1975), Gussenhoven (1984, 2002) and Steedman (2000). Note also that L*H appears to create a bias towards a given entity without involving contradiction. This calls into question Steedman's (2000) claim that marked words in themes receive L*+H where contradiction is intended by the speaker.

At the level of conceptualizing intonational meaning, our results raise concerns on the strong compositionality of intonational meaning proposed by P&H (1990). The finding that H*L (H* L-L% in ToBI) and H*LH (H* L-H% in ToBI) do not have the same effect indicates that the meaning of the pitch ac-

cent changes when followed by different phrasal tones. Further, the finding that L*HL (L*+H H-L% in ToBI) functions like H*L (H* L-L% in ToBI) suggests that pitch accents with different starred tones do not necessarily have different meanings.

As for the H*LH contour, it has been claimed to signal givenness by Brazil (1975) and Gussenhoven (1984, 2002). However, our data show that H*LH did not seem to have consistent effects on the interpretation of information status. When the competitor was given and the target was new, it functioned like a newness accent; when the competitor was new and the target was given, it functioned like a givenness accent. These patterns imply that H*LH may have biased participants' interpretation to the target, independent of its information status. For the sake of textual coherence, we postpone the discussion on this bias of H*LH till Section 5.

4. Experiment 2 – Synthetic speech

Results from Experiment 1 have established that pitch accent type matters for the online processing of information status in natural speech. The question thus arises as to whether the effect of pitch accent type will be preserved in synthetic speech, of which the segmental quality is lower than the segmental quality of recorded natural speech. Studies examining speech comprehension in synthetic speech have reported mixed results, which can lead to conflicting predictions on the effect of pitch accent type in synthetic speech.

In a study on the processing differences in synthetic vs. natural speech, O'Bryan (2000) presented participants with spoken garden-path sentences in the two types of speech and asked them to judge for each sentence whether it was grammatical or not within 1.5 seconds. The spoken garden-path sentences differed in the position of disambiguation point in the sentence. The ones with a late disambiguation point (e.g., *John was sad that the car raced in the Indy 500 burned | up.*) were considered to be more difficult to be comprehended than the ones with an early disambiguation point (e.g., *The detective remembered that the defendant examined by | John laughed.*). O'Bryan found that there was a significant main effect of sentence type in natural speech with the difficult sentences triggering more grammaticality judgement errors than the easy ones. However, there was no such an effect in synthetic speech. O'Bryan suggested that this difference could be caused by the low intelligibility of synthetic speech, which either made the recovery from a garden path impossible in both sentence types or left little time for the participants to recover from a garden path such that the position of the disambiguation point in a sentence did not matter. It may thus be hypothesized that the high processing load required at the phoneme and word levels could hinder processing of suprasegmental in-

formation in the speech signal, and consequently, the effect of pitch accent type observed in natural speech would not be present in synthetic speech.

On the other hand, there is also evidence suggesting that the effect of pitch accent type would be similar to what is present in natural speech but would be delayed. Swift et al. (2002) examined the online processing of synthesized words in context in an eyetracking study, and found that it was not different from processing naturally spoken words. Listeners "entertain multiple lexical candidates on the fly depending on segmental overlap in the candidate set" (Swift et al. 2002: 1277) in both natural and synthetic speech. That is, fixations to the picture with a phonemically similar name (e.g., *beetle*) as the target picture (e.g., *beaker*) increased until the target word was recognized. However, the increase in fixations started earlier in natural speech than in synthetic speech. Swift et al. did not attribute the difference in the time course to the lower intelligibility of the synthetic speech. But it was likely that listeners needed more time to process phonemic information when presented with synthetic speech than in natural speech. As a result, they were later in directing their visual attention in synthetic speech.

We tested the two predictions in Experiment 2, in which we presented the stimuli used in Experiment 1 in diphone-synthetic speech to another group of participants and obtained eye fixation data following the same procedure.

4.1. Method

The eyetracking paradigm described in Section 3.1 was used in Experiment 2.

4.1.1. Experimental design. In Experiment 1, five accent conditions were included into the experimental design. When composing the stimuli in synthetic speech, we noted that the pitch accents L*HL and H*LH could not be successfully generated by our synthesizer when the sonorant material of the stressed syllable was sparse. We therefore decided not to examine the effect of L*HL and H*LH in synthetic speech. Combining the two types of information status of the target/competitor at the onset of the second instructions and the remaining three accent conditions gave us six experimental conditions, as illustrated in Table 3.

4.1.2. Materials. The picture stimuli used in Experiment 1 were used here. The spoken instructions were generated with the Festival diphone synthesizer, which can implement intonation choices via APML tags (De Carolis et al. 2004). Figure 4 shows the f_0 tracks for *now put the window below the circle* with the target word *window* realized with H*L L%, L*H H%, and deaccentuation with a low boundary tone. Assuming that the segmental quality of

Table 3. Illustration of the six experimental conditions in Experiment 2

First instruction	Second instruction	Information status
Put the *windmill* above the triangle (competitor)	Now put the *window* below the circle. (target) H*L L% %L*H H% deaccentuation L%	New target Given competitor
Put the *window* above the triangle (target)		Given target New competitor

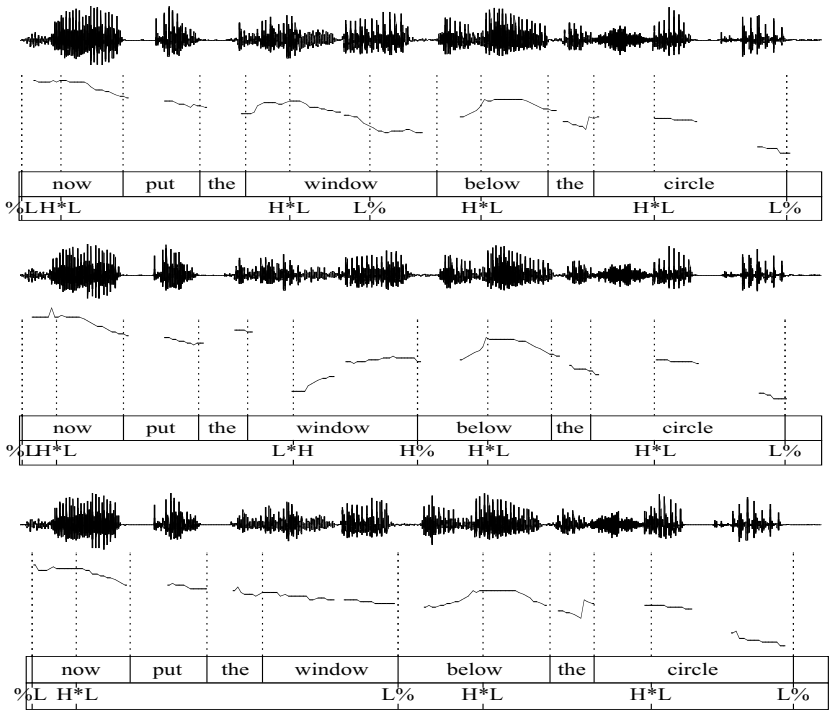

Figure 4. f_0 tracks for *now put the window below the circle* with *window* produced in three accent conditions

speech is positively correlated with intelligibility, we asked the participants of Experiment 2 to judge the intelligibility of the stimuli on a 7-point scale with 1 standing for "hardly intelligible" and 7 "very intelligible" at the end of the experiment. The mean intelligibility score of the stimuli is about 5.8. This indicates that the segmental quality of the spoken instructions is not optimal.

4.1.3. Procedures. Twenty undergraduates and two postgraduates from the School of Psychology at the University of Birmingham who did not participant in Experiment 1 took part in the experiment. They all spoke Southern British English as their only native language. None of them reported to have hearing problems. They received either course credits or a small fee for their participation. The experiment took about 10 minutes.

4.1.4. Coding procedure. The incompletely sampled data from two subjects and data from one subject who launched few fixations before the end of the

target word were excluded from coding. Data from the other 19 subjects were coded following the same procedure as in Experiment 1 (see Section 3.1.4). For 18 of these subjects, data from the right eye were coded; for one of these subjects, data from the left eye were coded because of calibration problems with the right eye.

4.2. Results

The coded data from 19 subjects were further analyzed. The proportion of fixations to each location (i.e., target picture, competitor picture, distractor pictures, and elsewhere) was calculated in 33 ms time intervals for each condition and each participant, as in Experiment 1.

Figure 5 presents the proportions of fixations (averaged across 19 subjects) to the competitor picture for H*L, L*H and deaccentuation in 33 ms time intervals from 0 to 1023 ms after the onset of the target word. Because the effect of pitch accent was not very explicitly reflected in the patterns of fixations to the target picture in natural speech, we will only consider the patterns of fixations to the competitor picture. Assuming that the minimal latency to plan and launch a saccade in synthetic speech is at least as long as in natural speech, we expected to observe the effect of intonation over the time window from 300 ms to 900 ms, at which point proportions of fixations were nearly the same in the three accent conditions.

Figure 5 (top) presents the mean proportions of fixations to the competitor picture (e.g., *coat*) when it was a *given* entity at the onset of the second instructions (e.g., *Put the coat above the square; now put the comb below the diamond.*). As is apparent, the fixation patterns differed in different accent conditions. Although the proportion of fixations increased at about 430 ms in all the three accent conditions, it started to decrease at different time points. It was the earliest in the H*L condition (627 ms), followed by the L*H condition (726 ms), and the latest in the deaccentuation condition (792 ms). As in Experiment 1, this pattern can be interpreted to reflect how fast the participants shifted their visual attention away from the "wrong" picture to the target picture. Differences in the time point of visual attention shift are consistent with the hypothesis that H*L creates a bias towards a new entity and L*H, like deaccentuation, creates a bias towards a given entity, thus keeping the participants' attention longer to the "given" competitor.

Figure 5 (bottom) presents the mean proportions of fixations to the competitor picture (e.g., *coat*) when it was a *new* entity at the onset of the second instructions (e.g., *Put the comb above the square; now put the comb below the diamond.*). As can be seen, at about 300 ms, the proportion of fixations started to increase steadily in the H*L condition and reached its peak (0.23) at 528 ms

and started to drop at 594 ms. However, in the L*H condition and the deaccentuation condition, the proportion of fixations did not change much and started to drop already at 462 ms and 429 ms, respectively. The difference in the onset of the decrease in fixation proportion is in line with the hypothesis that H*L creates a bias towards a new entity while L*H and deaccentuation do not.

To evaluate the observations discussed above, we conducted two repeated measures ANOVAs on the mean proportions of fixations to the "given-competitor" and to the "new-competitor" with two variables at a significance level of 0.05: Time Interval (18 levels: 300 ms to 900 ms in 33 ms interval), and Accent Condition (3 levels: H*L, L*H, deaccentuation). The analyses revealed a main effect of Time Interval in both analyses ($F(17,306) = 4.415$, $p < .05$, partial $\eta^2 = .197$) when the competitor was "given", ($F(17,306) = 3.9$, $p < .05$, partial $\eta^2 = .178$) when the competitor was "new"). The two-way interaction of Accent Condition $\times$ Time Interval ($F(34,612) = 2.328$, $p < .05$, partial $\eta^2 = .115$) was significant when the competitor was a given entity, but not significant when the competitor was a new entity. The asymmetry in the results between the "given-competitor" condition and the "new-competitor" condition was also observed in the data from the natural speech and may again be accounted for by the fewer fixations launched in general when the competitor was new.

4.3. Discussion

Clearly, listeners make use of intonational cues, i.e., type of pitch accent as well as deaccentuation, in the interpretation of information status in synthetic speech, although the segmental quality in the synthetic speech is lower than in recorded natural speech. This is contra the prediction derived from O'Bryan (2000). The effect of pitch accent type is predominantly reflected in the time course of fixation proportions, in particular, the onset of a clear decrease in fixations to the competitor. When the competitor was a given entity, the decrease occurred substantially earlier in the H*L condition than in the L*H and deaccentuation conditions. When the competitor was a new entity, the decrease occurred earlier in the L*H and deaccentuation conditions than in the H*L condition. As in natural speech, the decrease in the deaccentuation condition was later than in the L*H condition when the competitor was given but earlier than in the L*H when the competitor was new. This suggests that deaccentuation creates a stronger bias towards a given entity than L*H in synthetic speech too.

Further, the effect of accent condition was observable from 627 ms onwards in the given-competitor condition in synthetic speech but from 430 onwards in the same condition in natural speech. This difference can be interpreted to mean a delay in the effect of pitch accent type, which can be related to extra

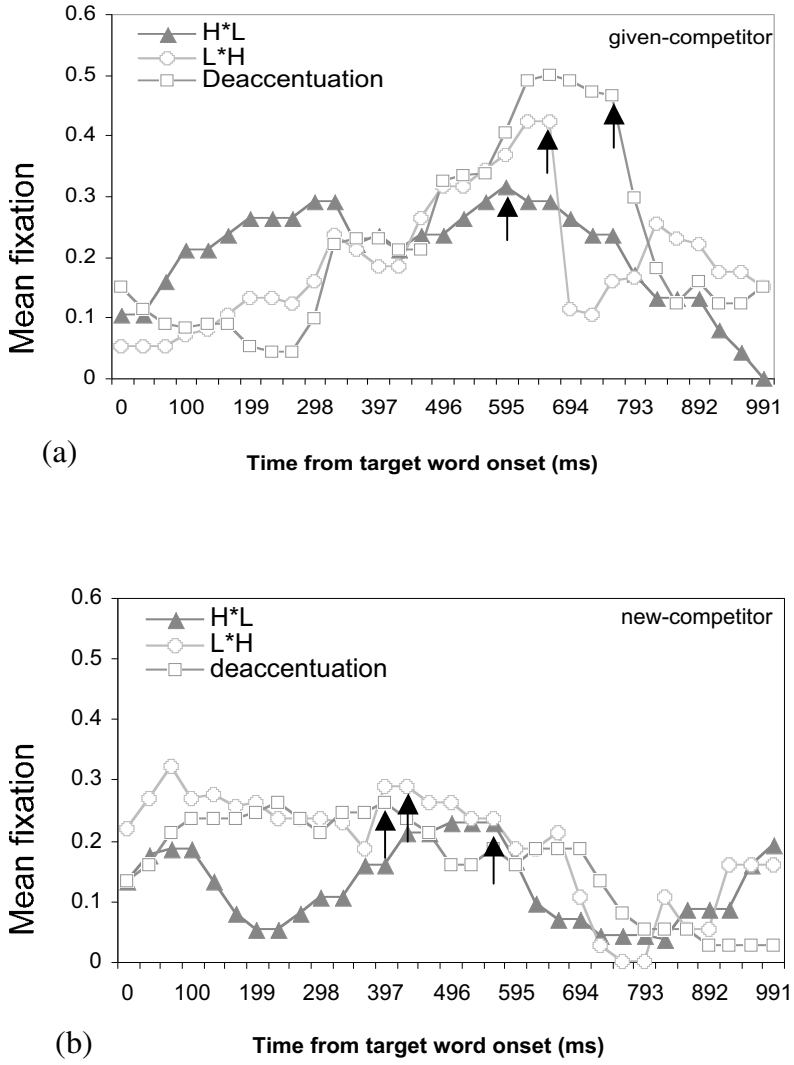

Figure 5. Fixation proportions (averaged across 19 participants) over time to the competitor picture from the onset of the target word in the second instruction as a function of accent condition when the competitor was given (top) and when the competitor was new (bottom). The arrows mark the time point of the decrease in fixation proportion in each accent condition.

processing load at the segmental level. Intriguingly, the delay is not present in the new-competitor condition in synthetic speech. If extra processing load at the segmental level costs more time and therefore postpones the processing of suprasegmental information, the delay should occur in the new-competitor conditions too. Earlier we noted a bias towards given entities in the participants. When the competitor was given, the proportion of fixation increased initially in all accent conditions in both natural and synthetic speech. The duration of the bias towards a given entity may have been lengthened by the segmental quality and thus led to the delay of intonational effect. In the new-competitor conditions, participants did not seem to have a bias towards the competitor. Consequently, the segmental quality may not have affected the processing of suprasegmental quality in any significant way.

It can thus be concluded that pitch accent type has the same effect on online processing of information status in synthetic speech as in natural speech, but there is a delay in the effect in synthetic speech, as predicted on the basis of findings from Swift et al. (2002). The delay is, however, not caused by the difference in the segmental quality between synthetic speech and natural speech alone but possibly in combination with a strong bias towards a given entity in the listeners.

5. General discussion

In this investigation, we examined the role of nuclear H*L, L*HL, L*H and H*LH in online processing of information status in British English via the eyetracking paradigm (Dahan, Tanenhaus and Chambers 2002). Results clearly show that type of pitch accent matters in both natural speech (Experiment 1) and synthetic speech (Experiment 2), and in the same way. The effect can be reflected in the mean proportion of fixations to the competitor in a selected time window. Unexpectedly, the effect of pitch accent type can also be reflected in how early participants recognize the target word and begin to direct their visual attention away from the "wrong" pictures to the target picture.

Findings in natural and synthetic speech are in agreement with the hypothesis that H*L creates a bias towards newness, as may be derived from theories of intonational meaning. Furthermore, data obtained in natural speech show that L*HL , like H*L, creates a bias towards givenness but the bias appears to be stronger in the L*HL condition, lending support to Brazil (1975) and Gussenhoven (1984, 2002), but contra P&H (1990). We have also found in both natural and synthetic speech that L*H, like deaccentuation, creates a bias towards givenness, though the bias is stronger in the deaccentuation condition. This result confirms P&H (1990) analysis but argues against Brazil's (1975) and Gussenhoven's (1984, 2002) analyses. The finding that L*H appears to create

a bias towards a given entity without involving contradiction casts doubts on Steedman's analysis (2000).

Our findings extend previous results on the effect of intonation on the processing of information status from accent placement to accent type, together with two recent studies that were brought to our attention after the completion of our experiments. In these two similarly designed and independently conducted studies, evidence has been found for the different roles of H* and L+H* in the signaling of information status in American English. Watson, Tanenhaus and Gunlogson (2004) found that L+H* created a strong bias towards contrastive information but not towards new information, whereas H* created a bias towards both contrastive information and new information. Watson, Tanenhaus and Gunlogson's definition of contrastive and new information can be illustrated by the second mention of *camel* and the first mention of *candle* in a three-step instruction respectively: *Click on the camel and the dog; move the dog to the right of the square, now, move the camel/candle below the triangle*. Using a similar experimental paradigm, Ito and Speer (2006) found that the placement of L+H* on the color adjective of a noun phrase (e.g., *green drum*) created a stronger bias towards contrastive information (e.g., the referent *drum* in the two-step instructions: *Hang the green drum; next hang the orange drum*.) than H*.

The pitch accent H*LH, has been claimed to signal givenness by Brazil (1975) and Gussenhoven (1984, 2002). However, our data show that H*LH did not seem to have consistent effects on the interpretation of information status. When the competitor was given and the target was new, it functioned like a newness accent; when the competitor was new and the target was given, it functioned like a givenness accent. These patterns imply that H*LH biased participants' interpretation to the target independent of its information status. This effect of H*LH may be explained in the light of the effect that the duration of a phonemically identical sequence has on its lexical interpretation. In a recent study on lexical garden-path in spoken word recognition, Davis, Marslen-Wilson, and Gaskell (2002) found that there was more activation for the shorter word (e.g., *cap*) when the sequence (e.g., /kæp/) came from a shorter word than when it came from a longer word (e.g., *captain*), and there was more activation for the longer word when the sequence (e.g., /kæp/) came from a longer word than when it came from a shorter word. These results were accounted for by referring to the durational difference of the phonemically identical sequence in the shorter and longer words, i.e., the sequence was longer (291 ms) in shorter words but shorter (243 ms) in longer words. In our experiment, the two words in 15 target-competitor pairs differed either in the syllable structure (e.g., CV vs. CVC; CVC vs. CVCC; CVCVC vs. CVCV) or in the voicing status of the coda (e.g., *card* /kɑːd/ vs. *cart* /kɑːt/). These differences led to durational differences in the phonemically identical sequences. In six target-competitor

pairs, the sequence was supposed to be longer in competitor (e.g., *plane*) than in target (e.g., *plate*). Acoustic analyses indicated that on average the sequence was 57 ms longer in the competitor than in the target when produced with H*L in the first instruction. In the second instruction, the durational difference of the sequence (= sequence duration of the target in an accent condition in the second instruction – sequence duration of the competitor produced with H*L in the first instruction) was relatively better maintained in the accent condition H*LH (31 ms) than in the accent conditions L*H (29 ms), L*HL (25 ms) and H*L (17 ms).[3] In the other nine target-competitor pairs, the sequence was supposed to be longer in target (e.g., *comb*) than in competitor (e.g., *coat*). Acoustic analyses indicated that on average the sequence was 40 ms longer in the target than in the competitor when produced with H*L in the first instruction. In the second instruction, the durational difference of the sequence appeared to be substantially enhanced in the accent condition H*LH. That is, the mean durational difference was the largest in the accent condition H*LH (124 ms), followed by H*L (97 ms), L*HL (93 ms), L*H (66 ms), and deaccentuation (31 ms). It may thus be suggested that the enhancement of the durational difference in the phonemically identical sequence in the accent condition H*LH produced facilitation to the recognition of the target word, which appeared to overrule potential effects of the interaction between accent condition and information status. Our data obtained from H*LH thus suggest an interesting topic for research on spoken word recognition, i.e., the interplay between pitch accent type, the duration of phonemically identical sequences, and information status in lexical interpretation.

The finding that pitch accent type has the same effect in synthetic speech as in natural speech, though with some delay, has useful implications. First, methodologically, it shows the potential of using stimuli in synthetic speech to investigate the effect of intonation on speech processing. Second, unit selection synthesis, contrary to diphone synthesis, has made it possible to generate speech with high quality. However, it allows little control over intonation and is largely dependent on the intonation of the selected units. Consequently, it is prone to produce contextually inappropriate intonation. It has been shown that listeners judged synthetic question-answer pairs to be intonationally more appropriate when the prosodic encoding of information structure was implemented, than when it was not (Baker, Clark, and White 2004). Our results imply that implementing prosodic encoding of information structure in synthetic

3. The durational difference was, however, best maintained in the deaccentuation condition because vowels become shorter when deaccented.

speech can also facilitate the processing of information. It is therefore important to integrate intonational signaling of information status into unit selection synthesis.

Max Planck Institute for Psycholinguistics (Chen, de Ruiter)
Radboud University Nijmegen (den Os)

References

Baker, Rachel, Robert Clark and Michael White (2004). Synthesizing contextually appropriate intonation in limited domains. In *Proceedings of the 5th ISCA Speech Synthesis Workshop*, Alan W. Black and Kevin A. Lenzo (eds.), 91–96. Pittsburgh: Carnegie Mellon University.

Baumann, Stefan and Martine Grice (2006). The intonation of accessibility. *Journal of Pragmatics* 38: 1636–1657.

Baumann, Stefan and Kerstin Hadelich (2003). Accent type and givenness: an experiment with auditory and visual priming. *Proceedings of the 15th International Congress of Phonetic Sciences*, Maria-Josep Solé, Daniel Recasens and Joaquín Romero (eds.), 1811–1814. Barcelona: Causal Productions.

Beckman, Mary E. and Gayle M. Ayers (1994). *Guidelines for ToBI Transcription* (version 2.0). Retrieved from http://www.ling.ohio-state.edu/~tobi/ame_tobi/.

Birch, Stacy and Charles JR. Clifton (1995). Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech* 38 (4): 365–391.

Brazil, David (1975). *Discourse Intonation*, vol. 1. Birmingham: Birmingham University.

Cutler, Anne, Delphine Dahan and Wilma van Donselaar (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech* 40: 141–201.

Dahan, Delphine, Michael K. Tanenhaus and Craig G. Chambers (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language* 47: 292–314.

Davis, Matthew H, William D. Marslen-Wilson and M. Gareth Gaskell (2002). Leading up the lexical garden-path: segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 28: 218–244.

De Carolis, Berardina, Catherine Pelachaud, Isabella Poggi and Mark Steedman (2004). Apml, a mark-up language for believable behavior generation. In *Life-like Characters, Tools, Affective Functions and Applications*, Helmut Prendinger (ed.), 65–85. Berlin: Springer.

Francis, W. Nelson and Henry Kučera (1982). *Frequency Analysis of English Usage. Lexicon and Grammar.* Boston: Houghton Mifflin Company.

Grabe, Esther (2004). Intonational variation in urban dialects of English spoken in the British Isles. In *Regional Variation in Intonation*, Peter Gilles and Jörg Peters (eds.), 9–31, Linguistische Arbeiten. Tübingen: Niemeyer.

Gussenhoven, Carlos (1984). *On the Grammar and Semantics of Sentence Accents*. Dordrecht: Foris.

— (2002). Intonation and interpretation: Phonetics and Phonology. In *Proceedings of the First International Conference on Speech Prosody*, Bernard Bel and Isabelle Marlien (eds.), 47–57. Aix-en-Provence: LPL – CNRS/SproSig.

— (2005). Transcription of Dutch intonation. In *Prosodic Typology and Transcription: A Unified Approach*, Sun-Ah Jun (ed.), 118–171. Oxford: Oxford University Press.

— (2006). Types of focus in English. In *Topic and Focus: Cross-linguistic Perspectives on Meaning and Intonation* (Studies in Linguistics and Philosophy, vol. 82), Chungmin Lee, Matthew Gordon, Daniel Büring (eds.), 83–100. Dordrecht: Kluwer.

Hallett, Peter E. (1986). Eye movements. In *Handbook of Perception and Human Performance,* Kenneth R. Boff, Lloyd Kaufman, and James P. Thomas (eds.), Chapter 10, 25–28. New York: Wiley.

Ito, Kiwako and Shari R. Speer (2006). Immediate effects of intonational prominence in a visual search task. In *Proceedings of the 3rd International Conference on Speech Prosody*, Rüdiger Hoffmann & Hansjörg Mixdorff (eds.). Dresden: TUDpress [CD-ROM].

Nooteboom, Sieb G. and Jacques M. B. Terken (1982). What makes speakers omit pitch accents? An experiment. *Phonetica* 39: 317–336.

O'Bryan, Erin (2000). Processing differences in synthetic versus natural speech. *MIT Working Papers in Linguistics* 38: 169–177.

Pierrehumbert, Janet B. and Julia Hirschberg (1990). The meaning of intonational contours in the interpretation of discourse. In *Intentions in Communication,* Philip R. Cohen, Jerry Morgan, and Martha E. Pollack (eds.), 271–311. MA: MIT Press.

Salverda, Anne P., Delphine Dahan and James M. McQueen (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition* 90: 51–89.

Snodgrass, Joan G., and Mary Vanderwart (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory* 6: 174–215.

Steedman, Mark (2000). Information structure and the syntax–phonology interface. *Linguistic Inquiry* 31 (4): 649–689.

Swift, Mary D., Ellen Campana, James F. Allen and Michael K. Tanenhaus (2002). Incremental referential domain circumscription during processing of natural and synthesized speech. Paper presented at the The 24th Annual Cognitive Science Society, George Mason University, USA.

Viviani, Paolo (1990). Eye movements in visual search: Cognitive, perceptual, and motor control aspects. In *Eye Movements and their Role in Visual and Cognitive Processes*, Eileen Kowler (ed.), 353–393. Amsterdam: Elsevier.

Watson, Duane, Michael K. Tanenhaus and Christine A. Gunlogson (2004). Processing pitch accents: Interpreting H* and L+H*. Paper presented at the *17th Annual CUNY Conference on Human Sentence Processing*, University of Maryland, Washington DC.