CLS 41: The Main Session (2008); 33-48

The role of pitch accent type in interpreting information status

Aoju Chen and Jan Peter de Ruiter
Max Planck Institute for Psycholinguistics - Nijmegen

Introduction

Discourse entities change their status from new information to given information as a discourse proceeds. Speakers can use intonation to signal information status by varying intonational prominence of the corresponding lexical entity. It is generally accepted (Culter, Dahan, van Donselaar 1997, references therein, Dahan, Tanenhaus, and Chambers 2002, Gussenhoven 2005a) that in languages such as English, the placement of pitch accent (i.e. absence and presence of pitch accent) is crucial for the marking of information status.[1] That is, new information tends to be accented but given information deaccented.[2] For example, Birch and Clifton (1995) examined the effect of accentuation in processing question-answer pairs (e.g., *Isn't Kerry good at math? / Yes, she teaches math.*) in American English. They found that answers in which the new information (i.e. *teaches*) was accented and the given information (i.e. *math*) was deaccented were judged to be prosodically more appropriate and triggered shorter Reaction Times than answers in which there was a mismatch between accent placement and information status.

In contrast, the role of the form of pitch accent, i.e. pitch accent type, is far from clear. In previous studies of accent placement and information status, the pitch accent(s) at issue often differ from study to study. For example, under the assumption that different types of pitch accent function in the same way, Nooteboom and Terken (1982) included three types of pitch accents in their Dutch stimuli, i.e. a rise early in the stressed syllable followed by a high plateau, a fall late in the stressed syllable preceded by a high plateau, and a rise-fall within one syllable. Birch and Clifton (1995) considered the L+H* accent in American English. In their study on the role of the interaction between accent placement and information status in reference resolution, Dahan, Tanenhaus and Chambers (2002) included both L+H* and H* in their stimuli in American English. Deaccentuation also appears to have different implementations in different studies. It was realised as absence of pitch movement in Nooteboom and Terken

[1] Pitch accent is defined as the pitch movement taking place on or starting from a sentence-accent bearing syllable. It can be described in terms of high (H) and low (L) tones. Pitch accent can be either monotonal (e.g. H*, L*) or bitonal (H*L, L*H). The star sign indicates that the tone is associated with the accented syllable. The other tonal movement of a contour is the boundary tone, associated with the two ends of an intonational phrase and demarcated with %. We adopt the ToDI notation (Gussenhoven 2005b) to describe pitch contours in our study but maintain the original notation when making reference to earlier studies, where ToBI was used (Beckman and Ayers 1994). See Gussenhoven (2005b) for a comparison between ToDI and ToBI.

[2] Note that this is the general tendency rather than the rule. See Gussenhoven (1983), Birch and Clifton (1995), Nooteboom and Terken (1982), and Terken and Hirschberg (1994) for discussions on factors that can be hold responsible for exceptions to this tendency.

(1982) and Birch and Clifton (1995), but as reduced pitch movement, transcribed as H+!H*, in Dahan et al. (2002). The differences in pitch accents under investigation and realisation of deaccentuation in these studies throw doubts on the unchallenged role of pitch accent placement in marking information status. There is some empirical evidence in recent studies suggesting that different types of pitch accents are used to convey different types of information status. Baumann and Hadelich (2003) investigated the appropriateness of H*, H+L* and deaccentuation (i.e. absence of pitch accent), in the marking of three types of information status in German: new (the lexical referent was introduced neither visually nor auditorily earlier), accessible (i.e. the lexical referent was earlier introduced only visually) and given (i.e. the lexical referent was earlier introduced both visually and auditorily). They found that both H* and H+L* were considered appropriate in marking new information, with H* being possibly more favoured. Furthermore, H+L* was preferred over H* for marking the accessible information. In addition, deaccentuation was most suitable to mark given information. Watson, Tanenhaus and Gunlogson (submitted) examined the role of H* and L+H* in on-line processing of information status and found that L+H* created a strong bias towards contrastive information (i.e. the lexical referent was mentioned early in the discourse but differed from a lexical entity in the immediately preceding sentence, which was also mentioned previously), whereas H* was compatible with new information (i.e. the lexical referent was not mentioned previously).

Against this backdrop, the present study set out to pin down the role of four pitch accents, fall (H*L), rise-fall (L*HL), rise (L*H), fall-rise (H*LH), as well as deaccentuation, in interpreting new vs. given information in British English. In the simple discourse adopted for our investigation, 'given' is defined as 'previously mentioned' and 'new' is defined as 'not previously mentioned'. The pitch accents in question were claimed to convey information status in theories of English intonational meaning (Brazil 1975, Gussenhoven 1984, 2002, Pierrehumbert and Hirschberg 1990, and Steedman 2000). There is, however, no consensus on the postulated roles of these pitch accents. In what follows, we will give a brief overview of the postulated relations between pitch accent types and information status in these theories. Where possible, we give the ToDI label of the pitch accent in brackets.

Theoretical background

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 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 signalled by five nuclear tones.[3] Proclaiming tones are fall (H*L) and rise-fall (L*HL). Referring tones include fall-rise (H*LH) and rise (L*H). The neutral tone is the low-rise (a variant of 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, similar 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 the fall (H*L), selection by the fall-rise (H*L H%), and testing by rise (L*H). These tones were considered the basic nuclear tones of English. [4] 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 into the delayed fall (L*HL L%), the delayed fall-rise (L*HL H%), and the delayed rise. 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-routiness.

In line with Brazil (1975) and Gussenhoven (1984, 2002), Pierrehumbert and Hirschberg (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). The interpretations of six pitch accents, two phrase accents and two boundary tones are spelled out.[5] Here we mention briefly the postulated functions of H*, L*, L*+H and H*+L. 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 salient in the discourse. The L*+H accent signals a lack of speaker commitment to a scale that links the accented item to other items salient in the hearer's mutual beliefs. The H*+L accent appears to make a predication as the H* accent does but differs from H* in conveying that the hearer should locate an inference path supporting the predication.[6]

---

[3] In Brazil (1975), the nuclear tone refers to the last pitch accent in an intonational phrase.

[4] In Gussenhoven (1984, 2002), the nuclear tone refers to both the pitch accent and the boundary tone. The boundary tone can be identical to the preceding tone, as in the fall and the rise, or different from the preceding tone, as in the fall-rise.

[5] The phrase accent refers to the boundary tone of an intermediate phrase, which is smaller than an intonational phrase)

[6] H*L is said to be the same as H* in phonetic implementation when followed by the L phrase accent (Pierrehumbert and Hirschberg 1990).

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.

As may have become clear, these theories make different claims on the exact functions of pitch accents in marking information status. The only agreement may be found in claims about the fall (H*L in ToDI and H* in ToBI). It is generally accepted that the fall marks new information. As regards L*HL and H*LH, only Brazil (1975) and Gussenhoven (1984, 2002) treated these pitch accents. It was suggested in both theories that L*HL marks new information but H*LH marks given information. Most controversial is probably the rise (L*H in ToDI, L*+H and L+H* in ToBI), which can signal given information (Brazil 1975, Pierrehumbert and Hirschberg 1990), contrastive information (Steedman 2000), or the meaning 'testing' (Gussenhoven 1984, 2002).

Hypotheses

Two hypotheses on the role of H*L, L*HL, L*H, and H*LH in interpreting information status can be derived from the theories of intonational meaning reviewed above. The two hypotheses differ in the predicted role of L*H.

> **Hypothesis 1**: H*L and L*HL trigger the interpretation of newness; L*H and H*LH trigger the interpretation of givenness, like deaccentuation.

> **Hypothesis 2**: H*L and L*HL trigger the interpretation of newness and H*LH, like deaccentuation, triggers the interpretation of givenness, but L*H is compatible with neither givenness nor newness.

Method

We adopted the eye-tracking paradigm used in Dahan et al. (2002) to evaluate our hypotheses. Dahan et al. (2002) 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 and timing of fixations to the reference and its lexical competitor. The eye-tracking paradigm may thus offer a measure of the effect of pitch accent type on interpreting information status.

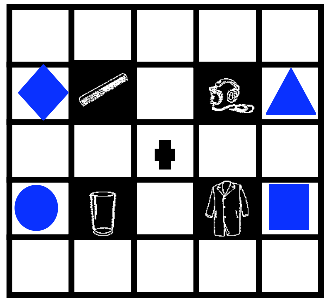


Figure 1. Example of a visual display. Geometric shapes were blue and in fixed positions.

Experimental design

On experimental trials, two of the objects had names that were phonetically related, i.e. sharing the same stressed syllable (e.g. *candle* vs. *candy*) or the same onset-peak cluster (e.g. *comb* vs. *coat*). One served as the target and the other as the competitor. Each trial consisted of two consecutive instructions. The second instruction always mentioned the target. The first instruction mentioned either the target (e.g., *Put the comb below the triangle)*, marking the target in the second instruction as given information but the competitor as new information, or the competitor (e.g., *Put the coat below the triangle*), marking the target in the second instruction as new information but the competitor as given information. Because of the phonetic similarity between the target noun and the competitor noun, in the second instruction the target noun was temporarily ambiguous during the first syllable or the onset-peak cluster. 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 et al. 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 said with H*L, L*HL, L*H, H*LH and deaccentuation. 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.

Table 1. Illustration of the ten experimental conditions.

Information status	First instruction	Second instruction
New target (given competitor)	Put the coat above the triangle;	now put the comb below the diamond H*L L*HL L*H H*LH deaccentuation
Given target (new competitor)	Put the comb above the triangle;	

Materials

Twenty pairs of nouns that were phonetically similar were selected from the materials used in Dahan et al. (2002) and served as the materials for experimental trials. As pitch accents are realised differently in monosyllabic words than in disyllabic words, to minimize effects related to phonetic realisation of pitch accent we included twelve pairs of monosyllabic words and 8 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 and competitors were similar. 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 phonetically similar names were likely to be moved in either instruction. Twenty-four of the filler trials (Group 1) consisted of two phonetically related items and two phonetically unrelated items. On 12 of these filler trials (Group 1a), the object in the first instruction was one of the phonetically related items; the object in the second instruction was one of the phonetically unrelated items. On the other 12 of these filler trials (Group 1b), one of the phonetically unrelated item was mentioned in the first instruction and the other in the second instruction. The other 24 filler trials (Group 2) consists of only phonetically unrelated items. On 12 of these filler trials (Group 2a), the objects mentioned in the two instructions differed from each other. On the other 12 of these filler trials (Group 2b), the objects mentioned in the two instructions were identical but differed from the objects mentioned in Group 2a filler trials. In each subgroup of filler trials, six trials were assigned to the hypothetic newness accents and the other six were assigned to the hypothetical givenness accents (according to Hypothesis 1).

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

The 272 (20 experimental trials × 4 + 48 filler trials × 4) pictures were selected from Snodgrass and Vanderwart (1980) and the MPI picture database. All were black and white line drawings.

The spoken instructions were recorded by a male speaker in Southern Standard British English at 48 kHz sampling rate in the sound-proof studio at the Max Planck Institute for psycholinguistics. The speaker read the instructions from printed materials. The intonation for each instruction was transcribed in the ToDI notation. The speaker was familiar with the ToDI notation and with producing pitch contours on request. An example of the recording script for one experimental trial is shown in (2). A boundary tone was inserted after the pitch accent on the target word in the second instruction, because it was believed to facilitate the full realisation of the pitch accent.[7] Figure 2 shows the f_0 tracks for *now put the comb below the diamond* with the target word *comb* said with H*L, L*HL, L*H, H*LH and deaccentuation.

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

Additional acoustic analyses were conducted on the target word in the second instructions to establish the contrast between L*H and L*HL and between H*L and H*LH in the part of the tonal specifications they had in common. These analyses included mean f_0, maximal f_0 difference, and the highest f_0 of the similar part of the f_0 contour. As the available f_0 information was not always the same across pitch accent types, different measurements were obtained for H*L and H*LH than for L*H and L*HL. Table 2 presents the values averaged across target words for different pitch accents.[8] As can be seen, H*L and H*LH were acoustically distinguishable from each other; so were L*H and L*HL.

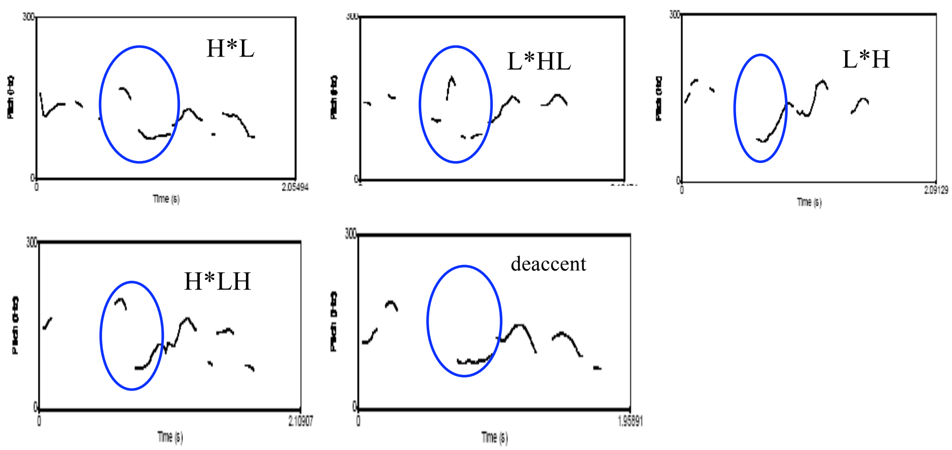


Figure 2. f_0 tracks for *now put the comb below the diamond* with *comb* said with H*L, L*HL, L*H, H*LH and deaccentuation.

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

	H*L	H*LH	L*H	L*HL
Mean F0			87 Hz	116 Hz
Maximal F0 difference	84 Hz	101 Hz		
Highest F0	167 Hz	174 Hz		

[7] The pitch contours of the target word were labelled as H*L L%, L*HL L%, L*H H% and H*L H%. The equivalent labels in the ToBI notation are H* L-L%, L*+ H L-L%, L* H-H% and H* L-H% respectively.
[8] The mean f_0 was averaged across 19 words. The maximal f_0 difference were averaged across 10 words; the highest f_0 value was averaged across the other 9 words. For one word, the f_0 values required to compute these measurements were not available.

Procedures

Twenty-four undergraduates and two postgraduates from the School of Psychology at the University of Birmingham participated in the experiment. They received either course credits or a small fee for their participation.

Participants were tested individually. Prior to the experiment, participants were told about the experimental task by means of written instructions in English. An example of the visual display was also included in the written instructions. They 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 an instruction was initiated. The positions of the pictures were randomised across four fixed positions of the grid, while the geometric shapes appeared in fixed positions on every trial. As soon as the picture was moved, the second instruction was initiated. Once the participant completed the two instructions on a trial, 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 central fixation point appeared on the screen after every five trials, which allowed automatic drift correction in the calibration.

Participants were randomly assigned to one of the stimulus order/stimulus list combinations. In six cases, the eye movement data were not properly sampled due to technical failures or unforeseen problems. A second participant was then tested in the same condition. The total number of participants thus amounted to 26. The experiment took less than 10 minutes. At the end of the experiment, participants were asked to fill out a questionnaire on their language background.

Coding procedure

Data from six participants were excluded from coding for various reasons: incomplete data due to technical failure (4 participants), incorrect interpretation of instructions (1 participant), and difficulty in recognising the picture due to poor eyesight (1 participant, wearing neither contact lenses nor glasses). Data from the other 20 participants (i.e. one participant for each stimulus order/stimulus list combination) were coded in terms of 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 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 location 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 pertaining to that picture.

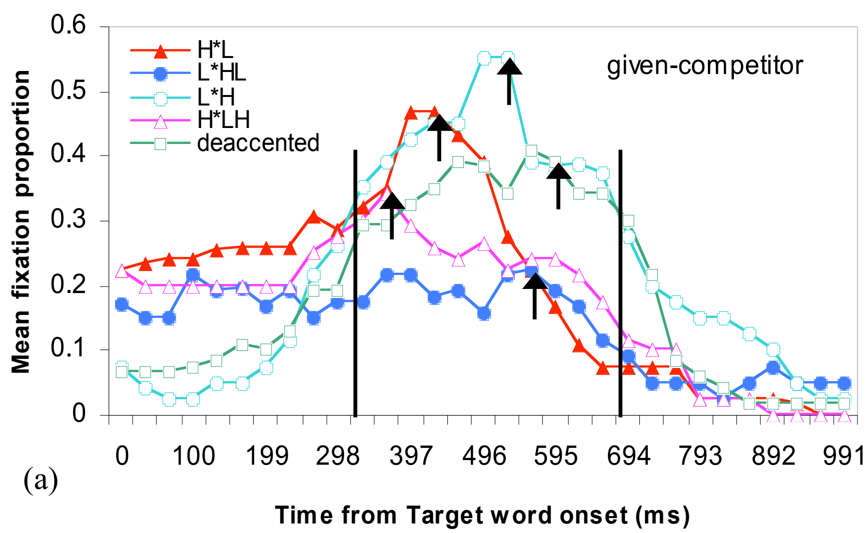
Results

The coded data from two participants were excluded from further analysis because few fixations were launched in the time window between the onset of the target word and the confident identification 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 (see Dahan et al. 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 (see Salverda et al. 2003). Figures 3 and 4 present the proportions of fixations (averaged across participants) to the competitor picture and the target picture respectively for H*L, L*HL, L*H, H*LH and deaccentuation in 33 ms time intervals from 0 to 1023 ms after the onset of the target word in the second instructions. As the minimal latency to plan and launch a saccade is 200-300 ms in tasks like visual search (Allopenna, Magnuson, and Tanenhaus 1998, Dahan et al. 2002), fixations realised in the first 300 ms of the target word were likely to be related to speech input preceding the target word. Because the part of the target word that overlapped with the competitor was about 400 ms long on average, the effects of accent conditions were expected to be strongest in the time window from 300 ms to 700 ms. We will therefore consider in detail the fixations on the competitor and the target realised in the time window from 300 ms to 700 ms after the target word onset in what follows.

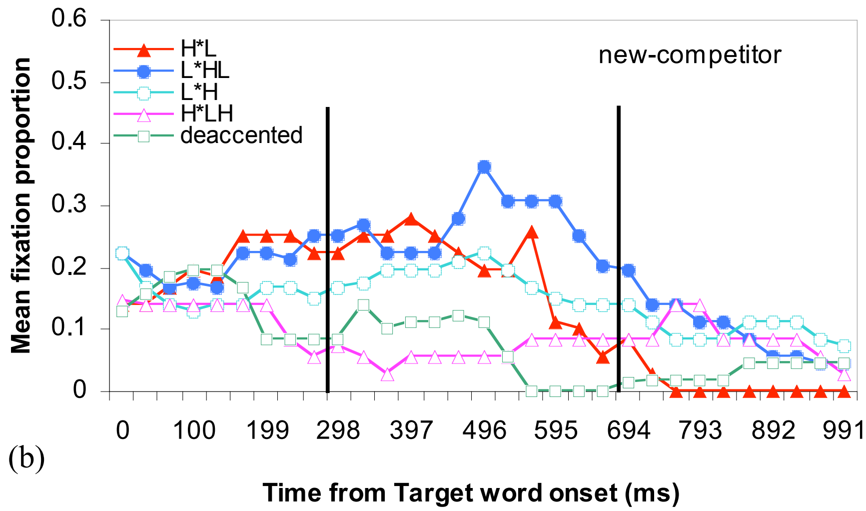
Fixation proportions to the Competitor

Figure 3a presents the mean proportions of fixations to the competitor when it was mentioned in the first instruction. Here the competitor was a *given* entity. The proportion of fixations to the *given* competitor started to increase steadily at 300 ms or earlier in all conditions except for the L*HL condition. In this condition, the fixation proportion remained low throughout the time window from 300 ms to 500 ms and began to decrease at 562 ms. These patterns are consistent with the hypothesis that L*HL creates a bias towards a new entity and therefore triggers fewer looks to a given entity. Moreover, the proportion of fixations started to decrease observably earlier in the H*L condition (at 430 ms) than in the L*H (at 529 ms) and deaccentuation conditions (at 595 ms). This pattern can be interpreted to reflect a shift of visual attention from the competitor picture to the target picture in the process of identifying the target in the second instructions. That is, as the target word was unfolding itself, participants realised that they were looking at the 'wrong' picture and began to shift their visual attention away from it. But they differed in the how fast they came to this realisation in different accent conditions, i.e. fastest in the H*L condition and slowest in the deaccentuation condition. As a consequence, there were more looks to the *given* competitor in the L*H condition (0.46) and the deaccentuation condition (0.33) than in the H*L condition (0.29) in the time window from 300 ms to 700 ms. These patterns are in line with the hypothesis that L*H, like deaccentuation, creates a bias towards a given entity but H*L creates a bias towards a new entity. The effect of H*LH is, however, unexpected. Although the proportion of fixations started to increase about 300 ms after the target word onset, it did not go higher than 0.34, which was only somewhat higher than in the L*HL condition. The proportion of fixations also started to decrease at a very early time point (343 ms). These patterns suggest that H*LH did not create a bias towards a given entity but functioned like a newness accent instead.

Figure 3b presents the mean proportions of fixations to the competitor when it was neither mentioned in the first instruction nor in the second instruction. In this condition, the competitor was a *new* entity. The proportions of fixations were relatively low in general (< 0.32). In the time window from 300 ms to 700 ms, H*L and L*HL triggered a relatively higher proportions of fixations to the *new* competitor than the other three accent conditions. The mean fixation proportion for each accent condition in this time window was: H*L – 0.20, L*HL – 0.27, L*H – 0.17, H*LH – 0.06, deaccentuation – 0.07. This observation is compatible with the hypothesis that H*L and L*HL mark newness but L*H and H*LH, like deaccentuation, mark givenness. Note that here H*LH seemed to function like a givenness accent.



(a)



(b)

Figure 3. Fixation proportions (averaged across participants) over time to the competitor picture from the onset of the target word in the second instructions as a function of pitch accent condition when the competitor was new (Figure 3a) and when the competitor was given (Figure 3b). The two vertical lines mark the time window from 300 ms to 700 ms; the arrows (Figure 3a) mark the time point when the decrease in fixation proportion occurred in each accent condition.

To evaluate the fixation patterns statistically, we conducted an ANOVA with two variables: Information Status (given, new) and Accent Condition (H*L, L*HL, L*H, H*LH, deaccentuation) for the mean fixation proportions over the time window from 300 ms to 700 ms at a significance level of 0.05. The analysis revealed a main effect of Information Status ($F(1, 17) = 15.35$, $p < 0.05$) and a significant interaction between Accent Condition and Information Status ($F(4, 68) = 3.12$, $p < 0.05$). The effect of Information Status is such that the fixation proportions to the competitor was higher when it was a given entity than when it was a new entity. This indicates that there was a general bias towards a given entity and the effects of accent conditions were reflected in their modifications to this bias as the target word unfolded itself. The statistical significance of the two-way interaction between Accent Condition and Information Status confirms that accent condition had effects on the interpretation of information status with H*L and L*HL creating a bias for a new entity but L*H and deaccentuation for a given entity. The effects of H*LH were compatible neither with triggering the interpretation of givenness nor with triggering the interpretation of newness.

Fixation proportions to the Target

Figure 4a presents the mean proportions of fixations to the target when it was not mentioned in the first instruction. In this condition, the target was a *new* entity. The proportion of fixations to the *new* target started to increase at different time points across accent conditions. The increase was noticeably earlier in the H*L (at 430 ms) and L*HL (at 463 ms) conditions than in the L*H (at 529 ms) and deaccentuation (at 562 ms) conditions. This pattern is consistent with the hypothesis that H*L and L*HL create a bias towards a new entity whereas L*H and deaccentuation create a bias towards a given entity. Note that when the fixation proportions to the *given* competitor started to decrease (Figure 3a), the fixation proportions to the *new* target started to increase. This indicates that the use of newness accents facilitated the recognition of the target, which was a new entity, but the use of givenness accents delayed it by keeping the participants looking at the *given* competitor for a longer time. The effect of H*LH was again unexpected. It seemed to function like a newness accent: the proportion of fixations started to increase rather early (at 397 ms) and were higher than in other conditions until 826 ms.

Figure 4b presents the mean proportions of fixations to the target when it was mentioned in the first instruction. In this condition, the target was a *given* entity. The proportions of fixations started to increase at different time points across accent conditions: H*L – 496 ms, L*HL – 463 ms, L*H – 397 ms, H*LH – 430 ms, deaccentuation – 166 ms. The mean fixation proportion for each accent condition in the time window from 300 ms to 700 ms were: H*L – 0.39, L*HL – 0.34, L*H – 0.38, H*LH – 0.431, deaccentuation – 0.58. These observations are largely compatible with the hypothesis that H*L and L*HL create a bias towards a new entity but L*H and H*LH, like deaccentuation, create a bias towards a given entity. Note that here H*LH seemed to function like a givenness accent.

To evaluate the fixation patterns statistically, we conducted an ANOVA with two variables: Information Status (given, new) and Accent Condition (H*L, L*HL, L*H, H*LH, deaccentuation) for the mean fixation proportions over the time window from 300 ms to 700 ms at a significance level of 0.05. The analysis revealed a main effect of Information Status ($F(1, 17) = 15.35$, $p < 0.05$). The effect of Information Status is such that the fixation proportions to the target was higher when it was a given entity than when it was a new entity. This confirms the general bias towards a given entity observed in the data obtained for the competitor. The interaction between Information status and Accent Condition did not reach significance, even though the general patterns of fixations confirmed the findings from data obtained for the competitor. This asymmetry in results may be related to the fact that the part of the target word overlapping with its competitor contained cues (i.e. duration of the overlapping segments) that biased its lexical interpretation towards the target word (see Session Discussion and Conclusions for more discussion). Consequently, the effect of the interaction between information status and accent condition became less strong on fixation proportions to the target.

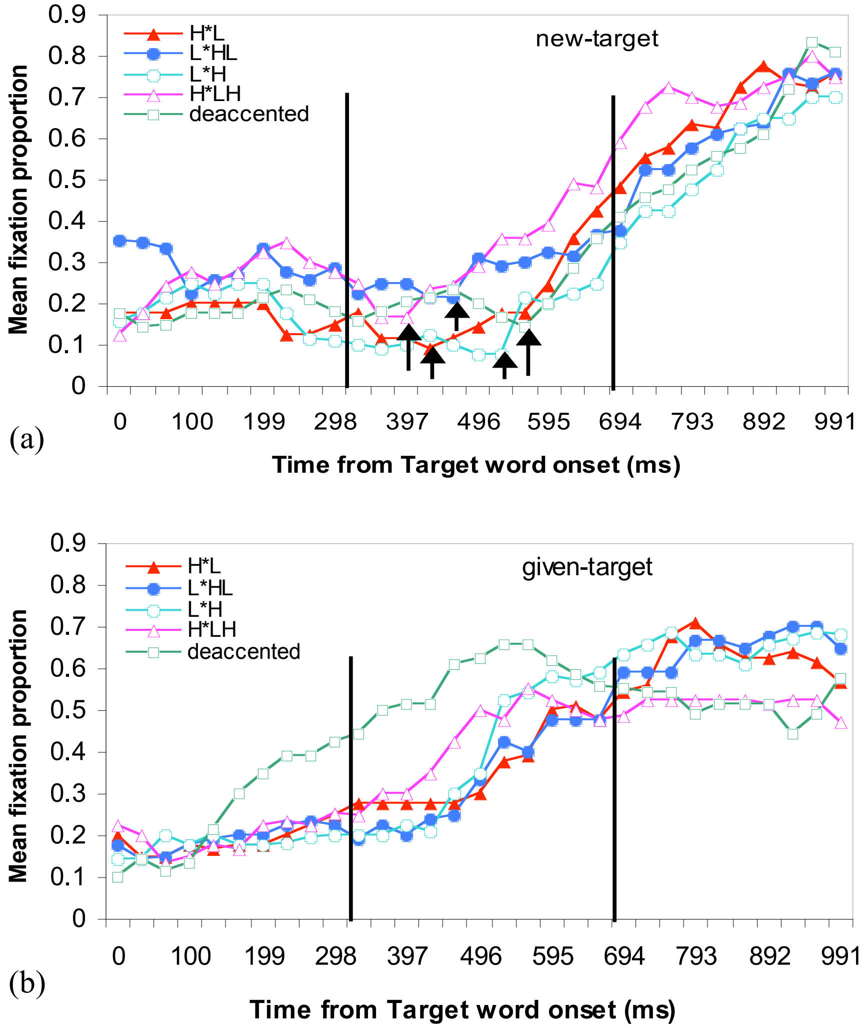


Figure 4. Fixation proportions (averaged across participants) over time to the target pictures from the onset of the target word in the second instructions as a function of accent condition when the target was new (Figure 4a) and when the target was given (Figure 4b). The two vertical lines mark the time window from 300 ms to 700 ms; the arrows (Figure 4a) mark the time point when the increase in fixation proportion occurred in each accent condition.

Discussion and Conclusions

In this investigation, we examined the role of pitch accents H*L, L*HL, L*H and H*LH in interpreting information status in British English by the eyetracking paradigm (Dahan et al. 2002). Results clearly show that pitch accent type can and does matter when interpreting information status. The effects can be reflected in the mean proportions of fixations to the competitor in a selected time window. For example, when the competitor was given, L*H and deaccentuation triggered more looks to it than H*L and L*HL; when the competitor was new, H*L and L*HL triggered more looks to it than L*H and deaccentuation. These patterns are also present in proportions of fixations to the target but to a lesser extent. Interestingly, the effects of pitch accent types are also reflected in how fast the participants could adjust their decision as to which picture to move before the name of the picture was fully revealed. For example, when the competitor was a given entity, the proportion of fixations to the competitor increased initially in most accent conditions in the first as a result of subjects' bias towards a given entity, but started to decrease substantially earlier in the H*L condition than in the L*H and deaccentuation conditions.

Our findings are in agreement with the hypothesis that H*L triggers the interpretation of newness, as claimed in theories of intonational meaning. Furthermore, we have found that L*HL triggers the interpretation of givenness, lending support to Gussenhoven's (1984, 2002) proposal that the delayed fall signals Addition. We have also found that L*H, like deaccentuation, triggers the interpretation of givenness. This result confirms Brazil's (1975) proposal that rise marks Referring and arguably Pierrehumbert and Hirschberg (1990)'s analysis but argues against Gussenhoven's (1984, 2002) proposal that rise signals Testing. Note also that L*H appears to create a bias towards a given entity without involving contradiction. This calls into question Steedman's claim that marked words in themes receives L*+H where contradiction is intended by the speaker.

As regards H*LH, it was claimed to signal givenness in Brazil (1975) and Gussenhoven (1984, 2002). However, out 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., /kɑ:t/ vs. /kɑ:d/). 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 than in target. Acoustic analyses indicated that on average the sequence was 57 ms longer in the competitor than in the target when said 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 said 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).[9] In the other nine target-competitor pairs, the sequence was supposed to be longer in target than in competitor. Acoustic analyses indicated that on average the sequence was 40 ms longer in the target than in the competitor when said 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), 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 sequence, and information status in lexical interpretation.

Acknowledgements

This research was supported by the COMIC project (IST-2001-32311). We thank Dephine Dahan, Els den Os, Barbara Schmiedtová, and Michael White for helpful comments on the experimental design, Carlos Gussenhoven for recording the stimuli, Doug Davidson, Herbert Baumann, John Nagengast, Leah Roberts, Keren Shatzman, Johan Weustink, and Anna Zumach for their help with setting up the experiment, Antje Meyer for making it possible to conduct the experiment at the University of Birmingham, Linda Mortensen for her support during the testing and useful comments on an earlier version of the text, and Yang Luo for his help with automating the data processing procedure.

References

Allopenna, Paul D., Magnuson, James S., & Tanenhaus, Michael K. 1998. Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *Journal of Memory and Language* 38: 419–439.

Baumann, Stefan & Hadelich, Kerstin. 2003. Accent type and givenness: an experiment with auditory and visual priming. *Proceedings of the 15th ICPhS*, 1811-1814.

Beckman, Mary. E. & Ayers, Gayle M. 1994. *Guidelines for ToBI transcription* (version 2.0).

Birch, Stacy, & Clifton, Charles Jr. 1995. Focus, accent, and argument structure: effects on language comprehension. *Language and Speech* 38: 365–391.

Brazil, David. 1975. *Discourse intonation*, vol. 1. Birmingham University.

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

---

[9] The durational difference was however best maintained in the condition Deaccentuation when the sequence in the target was supposed to be shorter than in the competitor, because vowels become shorter when deaccented.

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

Davis, Matthew H, Marslen-Wilson, William D, & Gaskell M Gareth. 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.

Gussenhoven, Carlos. 1983. Testing the reality of focus domains. *Language and Speech* 26: 61-80.

Gussenhoven, Carlos. 1984. *On the grammar and semantics of sentence accents*. Dordrecht: Foris.

Gussenhoven, Carlos. 2002. Intonation and interpretation: Phonetics and Phonology. *Proceedings of the First International Conference on Speech Prosody*, 47-57.

Gussenhoven, Carlos. 2005a. Types of Focus in English. In Daniel Bäuring, Matthew Gordon & Chungmin Lee (eds.) *Topic and Focus: Intonation and Meaning. Theoretical and Crosslinguistic Perspectives*. Dordrecht: Kluwer.

Gussenhoven, Carlos. 2005b. Transcription of Dutch Intonation. In Jun, Sun-Ah (ed.), *Prosodic Typology and Transcription: A Unified Approach. Oxford*. Oxford University Press.

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

Pierrehumbert, Janet. B. & Hirschberg, Julia. 1990. The meaning of Intonational Contours in the Interpretation of Discourse. In R. R. Cohen, J. Morgen, & M. E. Pollack (Eds.), *Intentions in Communication* (pp. 271-311). MA: MIT press.

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

Snodgrass, Joan G, & Vanderwart, Mary. 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 struction and the syntax-phonology interface. *Linguistic inquiry* 31 (4): 649-689.

Terken, Jacques, & Hirschberg, Julia. 1994. Deaccentuation of words representing 'given' information: effects of persistence of grammatical function and surface position. *Language and Speech* 37: 125–145.

Watson, Duane, Tanenhaus, Michael K, & Gunlogson, Christine A. (under review) Interpreting pitch accents in on-line comprehension. Retrieved 25 November, 2004 from http://www.bcs.rochester.edu/people/dwatson/#publications