Temporal distribution of interrogativity markers in Dutch: A perceptual study

Heuven, V.J. van; Haan, J.; Gussenhoven C., Warner N.

Temporal distribution of interrogativity markers in Dutch: A perceptual study

Vincent J. van Heuven and Judith Haan*

Phonetics Laboratory, Universiteit Leiden Centre for Linguistics (ULCL-HIL)
PO Box 9515, 2300 RA Leiden, The Netherlands

*Department of General Linguistics and Dialectology, Nijmegen University/
Center for Language Studies (CLS)
PO Box 9103, 6500 HD Nijmegen, The Netherlands

[chapter appeared in C. Gussenhoven, N. Warner (eds.) *Papers in Laboratory Phonology 7*, Phonology & Phonetics 4-1, Mouton de Gruyter, Berlin/New York, 2002, 61-86]

Abstract

A set of 128 different intonation patterns were synthesized on a single Dutch sentence with (potential) pitch accents on the subject (sentence-initial position) and on the object (sentence-medial position). The sizes of the accents were 0, 4, 8 and 12 semitones (st) in all 16 logical combinations. The accents were superposed on four different baseline slopes (B3, B1.5, 0 and +1.5 st/s), and utterances could or could not end in a terminal rise H% (0, 8 st). The 128 versions were presented to two groups of 20 native Dutch listeners. One group heard the 128 complete utterances and identified each version as either a statement or a question. The second group performed a gating task: they first heard (16) different versions that were truncated after the subject accent (gate 1), then (the same 16) with the truncation point delayed to the onset of the object accent (gate 2), next the (64) different versions with the truncation point at the end of the object accent (gate 3), and finally the (64) versions with truncation point at the onset of the (potential) terminal rise (gate 4). On hearing each gated stimulus listeners guessed whether they had heard the beginning of a statement or of a question. The results of both experiments were analysed in order to reveal the temporal development of cues that signal statement versus declarative question in Dutch. Sentence type is clearly signalled well before the terminal rise is heard. In terms of the experimental variables, the overall F0 trend (slope of baseline) is the most important early cue, followed by the size of the object accent. All early cues were overruled by the information in the terminal boundary. However, the percept of statement vs. question suffered significantly when early cues were contradictory to the absence versus presence of a terminal rise. The results bear on the issue whether the signalling of sentence type can still be modelled by a sequential, linear tone model or whether some global tone shape feature needs to be involved.

1. Introduction

The smooth exchange of information requires the exploitation of several distinct sentence types or speech acts. Among the most important of these are the expression of statement (presenting facts or beliefs) and the asking of questions (prompting the listener to take a turn and supply some information that is explicitly targeted by the speaker). The signalling of statement versus question seems universal (Sadock & Zwicky 1985: 195). In the large majority of the world's languages, the contrast is expressed by lexico-syntactic devices, as well as by intonational means. The statement type is generally considered the unmarked choice, requiring no special sign. Question is the marked sentence type, signalled lexically by a question word (a wh-word or question particle) or by dedicated syntactic means such as inversion of subject and finite verb. Whilst the lexico-syntactic question devices may vary considerably across languages, the apparent universal lies in the prosodic interrogativity marking: is has been claimed that questions are universally differentiated from statements by a higher beginning, a higher end and, more generally, by a higher overall pitch level throughout the utterance (Hermann 1942; Lindsey 1985). That is, apart from the canonical final rises, questions have been shown to display (some combination of) higher onset, raised register, raised peaks, and less downward or even upward overall trend. Obviously, such early cues have the advantage of drawing a listener's attention to the interrogative character of the utterance before it is finished, facilitating the subsequent response.

Although 'some high-pitched element' would appear to be a (near) universal characteristic of question intonation, languages may differ considerably in their choice of phonological tone configurations and phonetic implementation of the interrogative sentence type (Gussenhoven & Chen, 2000). Haan, van Heuven, Pacilly & van Bezooijen (1997) presented systematic analyses of the Dutch sentence melodies of statements and of three types of questions derived from these statements, i.e., wh-questions, yes/no-questions, and declarative questions (the latter differing from the corresponding statement in intonation only). The results revealed clearly different intonation profiles for each of these four sentence types. Moreover, although the three question types typically ended in a H% high boundary tone (the spoken equivalent of the question mark), all the sentence types were differentiated from the statement version at some earlier point in the time course of the utterance. Figure 1 shows averaged stylised pitch patterns for the statement and declarative question versions of the utterance *maRIna wil haar mandoLIne verkopen(?)* 'Marina wants her mandolin sell(?)', i.e. 'Marina wants to sell her mandolin(?)', spoken twice by five males and five females in three different positions in a short paragraph..[1]

CCCCCCCCC
figure 1 here
CCCCCCCCC

Figure 1. Stylised pitch contours (F_0 in ERB) for statements and declarative questions averaged over five male and five female speakers. Each data point nominally represents 100 measurements.Vertical lines represent the onset and nucleus segments of accented syllables. Low declination lines were drawn connecting the low onset pivot point to the last relevant low pivot. Note that the L% target extends clearly below the low declination line. The high declination lines connect the accent peaks on subject and object.

A number of differences are visible between the statement and question versions:

B The question ends with a high boundary tone H%; the statement with the low boundary tone

L%.

- The two accents in the statement are of roughly equal size; the accents are unequal in the question version:
  - The first pitch accent (i.e. on the subject) is smaller in the question than in the statement.
  - The second (final) pitch accent (i.e. on the object) is larger in the question than in the statement.
- The statement shows downtrend in the (imaginary) line connecting the low-pitched turning points in the contour, whilst a slight uptrend is shown in the question version.[2] This effect was noted earlier in a comparison of statements and echo questions in Danish (Thorsen, 1980).
- The accent peak on the object is reached slightly (ca. 40 ms) later (relative to either syllable or vowel onset) in the question version than in the statement. A rightward shift of the final accent peak in questions was reported for other languages as well, e.g. by Gosy & Terken (1994) for Hungarian, and by Makarova (1999) for Russian.[3]

Van Heuven & Haan (2000), using a gating task (see below), showed that Dutch listeners reliably detect the difference between these two prosodic versions as soon as the second pitch accent (on the object) is made audible. Clearly, then, the native listener need not await the end of the sentence (containing the terminal question marker H%) in order to know the sentence type intended by the speaker. However, it is unclear at this juncture what cue or combination of cues enables the listener to achieve this. Is it the large size of the object accent, or is it the difference between the small-sized subject accent and the large object accent? Or is the relevant cue not in the scaling of the accents but in the slope of the pitch trend in the low pivot points (i.e. presence versus absence of downtrend in the baseline)? Or are there still other cues in the signal that we did not measure?

This type of problem cannot be solved by analysing natural speech, since the various cues always co-occur. We need a way to disentangle them, and vary each cue orthogonally to the others. The only way to achieve this is to set up a listening experiment in which the necessary variations in pitch pattern are artificially generated. Fortunately, we have at our disposal signal processing techniques (such as PSOLA, cf. Moulines & Verhelst 1995) that allow us to create intonation patterns of our own choosing, and superpose these onto a pre-recorded human utterance, generally with little audible loss of sound quality or naturalness.

We created an orthogonal multidimensional stimulus space by introducing controlled variations along four dimensions (for details cf. ' 2.1):

1. Size of subject accent: varying from absent to extremely large.
2. Size of object accent: same range of variation as that of subject.
3. Slope of the baseline connecting low turning points of accents, from gentle downtrend to slight uptrend.
4. Presence versus absence of terminal rise H%.

Naturally, we expect the largest, even overriding, effect of H%: if the terminal rise is present, the overall percept will be that of a question, but the result may be less than convincing if the earlier prosody of the sentence does not set up the expectation of an interrogative, e.g. through clear downtrend or pitch accents of equal size. Specifically, we have reason to believe that the perception of interrogativity interacts with the way pitch accents are realised. Statements typically have pitch accents on two constituents, i.e. on the topic and on the comment. However,

when a speaker asks a question, there would appear to be a tendency to avoid accenting both topic and comment, presumably because there is a stronger need to mark the comment more clearly as new, and mark the topic as having continued relevance, not requiring a further attention marker. Consequently, when listeners hear (or expect to hear) a relatively high, single accent peak, they will be biased into assuming that the speaker intends a question.

Alternatively, our listeners' responses may be relatively independent of the specifics of the pitch contour and rather be governed by the overall shape of the sentence melody. Assuming, for instance, that listeners in most languages, including Dutch, will perceive interrogativity whenever the utterance has a concentration of high pitch towards the end of the utterance, similar results will be found for any combination of higher and later accent peaks, and a high terminal boundary tone. Under this view, the large object accent in the question version in figure 1 would be primarily motivated by the speaker's wish to elevate the average pitch towards the end of the utterance, rather than by considerations of focus.

We ran two experiments. In the first experiment the subjects heard only parts of sentences, which were truncated at one of several carefully chosen points in time. This so-called gating technique has enjoyed increasing popularity in experimental research aiming to trace the ability of listeners to set up expectations of upcoming events, and to determine the nature of the acoustic properties in early portions of the speech utterance that enables the listener to generate such projections (Grosjean 1983). In our case, we applied this technique to see how well listeners differentiate between statement and declarative question at each of these points in the temporal development of the intonation pattern.

In the second experiment, the full set of variations was offered once more, to a fresh group of listeners, in order to determine the effects of the above four types of pitch variation when the utterance is not presented repeatedly in chunks of increasing length, but uninterruptedly as in normal speech situations.

2. Method

2.1 Stimulus material

The material for both experiments was constructed from a single question utterance *maRIna wil haar mandoLIne verkopen?* [ma'rina ʋɪl har mɑndo'linə vər'kopə] (cf. p. 1), spoken by a female speaker of Standard Dutch with H*L pitch accents (indicated by small caps in the orthographic representation above) on the subject and object of the sentence, and with a high boundary tone (H%). The recording was transferred from DAT to computer memory and downsampled to 16 kHz (16 bits amplitude resolution). Pitch extraction was performed by the autocorrelation method implemented in the Praat (Boersma & Weenink 1996) speech processing package. The pitch curve was interactively stylised with 9 pivot points interconnected with straight lines in a log-frequency (semitone) by linear time representation (see also figure 1), such that no audible difference existed between the original and the stylised pitch curves. By manipulating only the frequency values but leaving the time coordinates unaltered, 128 different pitch patterns were then generated according to the following schema:

1. *Size of subject accent*. The excursion size on the stressed syllable of the subject (*maRIna*) was varied in 4 steps: 0, 4, 8 and 12 semitones (st) above the baseline.
2. *Size of object accent*. The excursion on the object (*mandoLIne*) was also varied in four steps (same four steps as subject accent).
3. *Trend*. The slope of the baseline connecting the minima in the stylised pitch contour was varied in four steps: !3, !1.5, 0 and +1.5 semitones per second (st/s). Here, the sentence-initial pitch pivot point was kept constant (as did our speakers), yielding a natural sounding range of overall trend varying from rather steeply dropping, through level pitch, to slightly rising.
4. *Boundary tone*. An H% boundary rise of 8 st was either present or absent.

The set of stimuli is schematically represented in figure 2.

CCCCCCCCC
figure 2 here
CCCCCCCCC

Figure 2. Schematic representation of stimulus materials used in experiments 1 and 2. The 4 H *4 combinations of rise-fall accents were superposed on each of the 4 baselines.*

For experiment 1, the stimuli were organised into sets of 16 to 64 variations, depending on the truncation condition (or 'gate').

Gate 1. The truncation point was set at the end of the subject accent. This yielded 16 different stimuli: 4 subject accent sizes H 4 baseline slopes. These 16 stimuli were generated in random order, preceded by 4 practice items, yielding a set of 20 trials.

Gate 2. The truncation point was moved to the onset of the rise belonging to the second accent; the same 16 combinations of subject accent and baseline slope were used, in a different random order. This yielded the second set of 20 trials (including 4 practice trials).

Gate 3. The truncation point was set at the offset of the object accent; variation of subject accent, object accent and baseline yielded 64 combinations, in random order. The third set was preceded by 1 practice trial.

Gate 4. The truncation point was shifted to the onset of the terminal rise; this yielded same 64 combinations (in a different random order) as in gate 3 (plus 1 practice trial) but followed by a longer stretch of baseline.

The total of 170 stimuli (including practice trials) were played back in blocks of 10 trials, with 3-second intervals between trials (offset to onset) and a 5-second interval plus a beep between blocks. In the order of presentation shorter gates always preceded longer gates.

For experiment 2, the entire set of 128 complete utterances (4 subject accents H 4 object accents H 4 baseline slopes H 2 boundary tones) were played back in random order, preceded by two practice items, in blocks of 10 trials, with 5-second intervals between trials and 7-second intervals + beep between blocks.

2.2 Subjects and procedure

Two groups of 20 native Dutch listeners participated in the experiment. Both groups comprised volunteers, students and researchers at the Department of Linguistics and Phonetics at Leiden University. Group 1 listened to the 170 truncated (gated) stimuli played to them through high-quality loudspeakers (Quad ESL-63) in a quiet medium-sized seminar room. Subjects were instructed (in writing) to indicate for each trial on their answer sheets whether they thought they heard the beginning of a statement or of a (declarative) question, with binary forced choice. Experiment 1 lasted about 30 minutes.

The subjects in experiment 2 (group 2) performed a dual task. On hearing each stimulus they were to indicate on their answer sheets (i) whether they judged the stimulus to be a statement or a question, with binary forced choice, and (ii) how clearly they thought the stimulus represented a prototypical exemplar of a statement or question, depending on their response to (i).[4] Experiment 2 lasted about 20 minutes.

3. Results

3.1 Experiment 1: gated sentence fragments

Overall effect of gate length. We will first present the results for experiment 1 (gating). In order to obtain an overview of the most important results of this experiment, figure 3 shows percent question responses (and by implication percent statement responses, i.e., the complement to 100 percent) as a function of gate (truncation points 1 through 4) broken down further by slope of the baseline (but accumulated across all sizes of subject and object accents).

CCCCCCCCC
figure 3 here
CCCCCCCCC

Figure 3. Percent 'question' responses as a function of truncation point, broken down by slope of baseline.

During the first two gates the listeners do not differentiate statement from question; responses are random during these early gates. However, some differentiation is observed at the third truncation point, where the object accent is first made audible. Here the rising baseline clearly triggers more question responses than the level and falling baseline conditions. The same pattern of results, but more clearly differentiated, is obtained at gate 4, when the stimulus is truncated just before the onset of the final rise.

The data were submitted to a repeated measures analysis of variance (RM-ANOVA) with slope of baseline and gate number as fixed, within-subject factors. All p-values reported are Huynh-Feldt corrected. The effects of gate, $F(3,57) = 1.4$ (ins.), and of slope, $F(3,57 = 17.7$ ($p < .001$) are not relevant to the present issue. Importantly, however, the interaction between slope and gate reaches significance, $F(9,171) = 7.9$ ($p < .001$), which indicates that the differentiation between statement and question increases for longer gates. Specifically, there is no effect of slope during gates 1, $F(3,57) < 1$, and 2, $F(3,57) < 1$, but significant effects of slope of increasing

magnitude are found for gates 3, $F(3,57) = 27.9$ ($p < .001$) and 4, $F(3,57) = 38.7$ ($p < .001$).

These data show that the slope of the baseline used to generate our stimuli provides an important perceptual cue for the sentence type (statement versus question) before the end of the sentence is reached. Moreover, since the data were pooled across excursion sizes of subject and object accent, the cue provided by the baseline would appear to be independent of other interrogativity cues. These observations should not be interpreted as meaning that listeners project the sentence type on the perceptual cue afforded by the baseline as such. It may be the case, for instance, that listeners only attend to the relative pitch difference between the accent peaks in the stimuli. A more detailed analysis of the results may then reveal that more 'question' responses are given if the object has a higher accent peak than the subject. This condition will be met more often in stimulus types with a positive slope for the baseline. Let us therefore examine the results in more detail, for increasing gate length and for different sizes of subject and object accents.

Effects at gate 1. Figure 4 presents percent question responses at gate 1 (when just the portion of the sentence is heard up to and including the subject accent), broken down by size of the accent and by baseline slope. The results reveal no effect of baseline slope at this early point in the utterance, as is to be expected given the absence of any overall effect of baseline slope before truncation point 3 is reached. Moreover, there is no systematic effect of the excursion size of the subject accent either, with just one exception. Only when the subject accent is extremely large (12 st, which is twice the normal size of a Dutch pitch accent, cf. 't Hart, Collier & Cohen 1990) do we find a clear propensity on the part of our listeners to project an interrogative utterance. The overall effect of subject accent, $F(3,57) = 8.5$ ($p < .001$) is significant but only because the 12-st accent differs significantly from the other excursion sizes, which do not differ from each other (Bonferroni posthoc tests, $p < .05$). This finding is compatible with our suggestion that a single (and therefore contrastive) accent may bias the listener into believing that the utterance will develop into a question. Alternatively, the large accent elevated the mean pitch of the stretch of speech heard up to the truncation point. Since questions are (universally) cued by high pitch somewhere in the utterance, the unusually high mean pitch in the audible onset portion of the utterance may be responsible for the strength of the interrogative projection.

CCCCCCCCC
figure 4 here
CCCCCCCCC

Figure 4. Percent question responses as a function of size of subject accent and slope of baseline, at truncation point 1.

Effects at gate 2. Figure 5 is organised in the same way as figure 4 but now the truncation point is delayed to the onset of the object accent. At gate 2 no information is given to the listener about the size (or even the existence) of the object accent.

CCCCCCCCC
figure 5 here
CCCCCCCCC

Figure 5. As figure 4, at truncation point 2.

Again, there is no systematic overall effect of baseline slope (which will not be manifest until gate 3, see figure 3). This time, however, the stimuli with a 0-st (i.e. no) accent on the subject are most strongly associated with interrogativity (supporting the expectation derived from the acoustic profile of the declarative question type in figure 1. The effect of subject accent is significant, $F(3,57) = 3.5$ ($p = .024$), but only because the 0-st accent differs from the other three conditions, which do not differ from each other. Note that baseline slope has a considerable effect (60% question responses for downtrend to 90% questions for uptrend, with intermediate scores for flatter slopes) if there is no accent on the subject. No effect of baseline slope is seen when there is an accent on the subject (whether small or large). It seems therefore that the information in the baseline can be picked up by the listener at a relatively early point in time, but only if its course is uninterrupted by (accent-lending) pitch obtrusions. This observation ties in with the result of Gussenhoven, Repp, Rietveld, Rump & Terken (1997) that only the declination portion before the (first) accent but not the post-accentual stretch of baseline provides the reference for the perceptual scaling of accentual prominence.

The stimuli with larger than 0-st subject accents are never associated with questions; at best, the 12-st subject accents are indifferent between statement and question, and the smaller subject accents are associated with statement; there is no effect of baseline slope in these stimuli. It is unclear to us why the addition of just a stretch of baseline after the subject accent caused the 12-st accents B which were strongly associated with interrogativity in gate 1 B to drop to indifferent between question and statement. This effect is commensurate neither with the contrastive accent hypothesis nor with the alternative elevated mean pitch explanation entertained above.

The next two figures present the effects of baseline slope and size of object accent on percent question responses, with the data accumulated across subject accents. Figure 6 presents the data at truncation point 3 (object accent audible); figure 7 shows the data at gate 4, including both the object accent and the stretch of baseline up to the point in time where the final rise would begin.

Effects at gate 3. Figure 6 shows that the size of the object accent has a powerful effect on the projection of interrogativity, $F(3,57) = 17.1$ ($p < .001$) by an RM-ANOVA with slope of basebase and size of object accent as fixed, within-subject factors. The interrogativity percept is strongest for the 12-st object accent, and gets progressively weaker (in fact crossing over to statements for the smaller object accents of 8 and 4 st. Bonferroni posthoc tests for contrast show that all accent sizes differ from each other, except 4 versus 8 st. This effect is in line with the expectation that would be derived from the acoustic profile of the declarative question type in figure 1, where the object accent was clearly larger than the subject accent. However, when there is no accent on the object at all (0 st), interrogativity perception is in between that found with 12 and 8 st. Possibly, when there is no accent on the object (0 st), any accent on the subject, but especially the larger types, may be interpreted as contrastive, and be associated with interrogativity. Possibly then, hearing a single accent, whether on the object or on the subject, biases the listener towards the question interpretation. Finally, there is an effect due to baseline slope, $F(3,57) = 27.9$ ($p < .001$),

which is to be expected given the earlier discussion of figure 3. Specifically, the rising baseline slope differs from the other three slopes, which do not differ from each other.

CCCCCCCCC
figure 6 here
CCCCCCCCC

Figure 6. Percent 'question' responses as a function of baseline slope and size of object accent, at truncation point 3.

Effects at gate 4. In figure 7, at truncation point 4, we find basically the same effect of size of object accent, $F(3,57) = 12.9$ ($p < .001$); here, however, the effect of baseline slope has become stronger than at gate 3, $F(57) = 38.7$ ($p < .001$). In fact, some 80% questions are projected regardless of size of object accent if the general pitch trend is upwards. All differences between baseline slopes are significant, except the contrast between flat (0 st/s) and slightly falling (B1.5 st/s).

CCCCCCCCC
figure 7 here
CCCCCCCCC

Figure 7. As figure 6, at truncation point 4.

Subject versus object. In figures 8 and 9 the effect of excursion size of object accent is pitted against that of the subject accent. The results were very similar for truncation points 3 and 4, so that we have accumulated the data across these two gates. However, given that baseline slope has a strong effect as of gate 3 (see figure 3), we present the results separately for global uptrend (figure 8) and for the flat and falling trends combined (figure 9).

CCCCCCCCC
figure 8 here
CCCCCCCCC

Figure 8. Percent 'question' responses broken down by size of object accent (across) and of subject accent (separate lines), at truncation points 3 and 4 combined. The baseline slope is upwards.

CCCCCCCCC
figure 9 here
CCCCCCCCC

Figure 9. As figure 8, but the baseline slope is either flat or falling.

When the baseline slope is upwards (figure 8), question perception is always above 60%, showing, once more, the strength of the baseline parameter. Generally, we observe no systematic overall effect of size of subject accent, $F(3,57) = 2.1$ (ins.). The size of the object accent,

however, matters, $F(3,37) = 5.7$ ($p = .002$): when there is no object accent, between 65 and 75% question responses are obtained, depending on the subject accent (slightly more question responses for 8 and 12-st subject accents); when the object accent is 12 st, about 90% question responses are obtained irrespective of subject size, and values of around 80% are found with object accents of 4 and 8 st, without any systematic interaction with subject accent size. Statistically, only the 12-st object accents differ significantly from the other three sizes, which do not differ among each other. Of course, larger object accents B on average B boost both mean and maximum pitch towards the final portion of the utterance, which would be in line the alternative account of the preponderance of question responses for larger object accents in terms of the overall shape hypothesis (p. 4).

When the global pitch trend is either flat or falling (figure 9), percent question responses is generally below 50 (i.e., statement rather than question). Only when the object accent is large (12 st) do we obtain more than 50% question projections, more strongly so if the subject accent is small (4 st) or absent (0 st). For smaller object accents the likelihood of statement responses increases steadily, with the exception, again, of 0-st object accents. When there is no accent on the object, a large subject accent tends to increase the chances of a question response (50% maximally, and excluding the accent-less utterances). Once more, this finding suggests that a single-accent sentence is conducive to projecting interrogativity. The effect of object accent size is significant, $F(3,57) = 18.1$ ($p < .001$); all accent sizes differ from each other, except the 0 vs. 4-st pair. Finally, and predictably from the results we presented above, there is no overall effect of subject accent, $F(3,57) = 2.0$ (ins.).

3.2 Experiment 2: complete utterances

Let us now turn to the results of experiment 2, where a fresh group of listeners decided whether they heard a question or a statement when the entire utterance was presented. Crucially, half of the complete utterances contained a terminal pitch rise, the prototypical interrogativity marker. We want to know whether hearing the final pitch rise, or its absence, obliterates all earlier projections of the sentence type as entertained by the listener during the earlier time-course of the utterance, or whether the final evaluation weighs the earlier cues along with the terminal boundary cue. The relevant results are presented in figure 10, where percent question responses is plotted as a function of global pitch trend, broken down by utterances ending in the terminal rise H% and those that do not.

CCCCCCCCC
figure 10 here
CCCCCCCCC

Figure 10. Percent 'question' responses as a function of global pitch trend, broken down by presence versus absence of H% (complete sentences).

The terminal pitch rise exerts an almost categorical effect. If H% is present, there is always a majority of question responses; whenever H% is absent, the perceived sentence type is statement, $F(1,19) = 235.3$ ($p < .001$). However, there is a quite noticeable effect of the slope of the

baseline, $F(3,57) = 28.4$ ($p < .001$). First, utterances without a terminal rise are less unanimously perceived as statements when the baseline is flat or upwards. Second, and more importantly, stimuli bounded by H% obtain (near) perfect question scores only if the baseline slope is upwards. Each time the baseline slope is decreased by a quantum of 1.5 st/s, the number of perceived questions drops by approximately 10%, leaving a mere 60% question responses for +H% utterances with the steepest downslope of baseline (B3 st/s).

Finally, we will briefly discuss some effects and interactions of the excursion size of subject and object accents. For sentences without the terminal rise, we do not obtain any effects of either subject or object accent when the baseline is flat or falling: all the stimuli are perceived as clear statements here. However, when the baseline slope is upwards, the size of the object accent matters: there are some 40% question responses when the object accent is large (12 st), but percentages drop for 8-st (20%) and 4-st (5%) object accents. Here, of course, baseline slope and size of the object accent conspire to elevate the (mean) pitch towards the end of the sentence, even if there is no high boundary tone, so that these observations do not necessarily support the idea that contrastive accent is associated with question responses. However, when there is no object accent (0 st), a (therefore single) accent on the subject generates roughly 25% question responses, suggesting that contrastive accentuation weakens the statement percept.

For complete utterances ending in a terminal rise, we find no further effects of subject and object accents if the baseline slope is upward (ceiling effect). When the baseline is flat or falling, there is plenty of room for secondary effects of subject and object accent size. As was the case in the presentation of the sentence fragments in experiment 1, here too the size of the object accent has an important effect: the smaller the object accent, the less convincing the question percept. Again, the size of the subject accent has no overall effect, but does matter when the object is unaccented.

4. Conclusions

Summing up the effects observed in the gating experiment, we draw the following conclusions at the observational level:

- Statements are differentiated from questions before the end of the sentence; differentiation is found by the time the accent on the object is heard.
- The slope of the baseline used to construct our stimuli is one important parameter that distinguishes statement (falling slope) from question (rising slope).
- Utterances that have, or are expected to have, a single (large) accent are more readily interpreted as questions than utterances with two (smaller) accents.
- A (large) accent on the object is more compatible with interrogativity than a (large) accent on the subject.

Clearly, these effects allow Dutch listeners to generate expectations as to the speech act (making a statement, or asking a question) on-line as the utterance develops in time. Normally, the on-line expectations are confirmed by the presence or absence of the sentence-final rise. In our experiment 2, however, the stimulus manipulations were such that the presence or absence of the terminal rise could clash with the listener's expectation based on the earlier prosody of the utterance. In the case of such a clash, the effect of H% is strongly attenuated, impairing the communication of interrogativity to the listener.

These conclusions suggest that the marking of interrogativity in text-to-speech systems (reading machines) can be considerably improved by implementing the various cues that we discovered in our research. Not only will the final product sound more convincing, but also, and possibly more importantly, the listener is given the means to project the speech act as it develops in time.

5. General discussion

Our experiment employed a set of 128 intonationally different versions of one meaningful Dutch sentence. The set was created through systematic variation of four parameters, i.e., size of subject accent, size of object accent, slope of baseline, and presence vs. absence of H%. It is important to point out that we claim no theoretical status for these parameters; they were adopted as convenient descriptive devices to allow us to compactly characterize the perceptually relevant aspects of our speakers' statement and question utterances, and to construct a four-dimensional stimulus space that adequately sampled the range of variation found in the natural utterances. In particular, we do not claim that the slope of the baseline is an entity that is voluntarily manipulated by the human speaker in order to signal the contrast between statement and (declarative) question.

In the preceding sections we have developed two possible mechanisms that the speaker may use to signal the contrast between statement and (declarative) question, which we will refer to as the compositional approach versus the integral approach. The integral approach involves a general strategy on the part of the speaker, not just in Dutch, but probably in the majority of the world's languages, to encode questions by generating a sentence melody with a globally rising shape. From this point of view it would make sense if the speaker were to reduce the peaks of accents in the early portions of the utterance, whilst increasing the peaks of the final accents. The globally rising shape of the melody would be brought out even more clearly by superposing the small subject accent and large object accent on a rising baseline. The global rising shape would then culminate in the terminal high boundary tone. This effectively describes the overall intonational shape of the declarative question. The overall shape of statements, by contrast, would lack all these ingredients. This view, however, would require that the speaker employs a dedicated mechanism for encoding sentence types, that generates planned overall rising versus falling melodic patterns for questions and statements, respectively.

We contrast this view with what might be called a minimalist view on the encoding of sentence type that involves no special, dedicated global shapes. In this compositional view, the global rising shape of question melody would fall out as a by-product of two independently motivated tonal devices: (i) ending the question in a high boundary tone, and (ii) contrastive accentuation of the comment in question sentences, such that size of the object accent is increased while the size of other accents in the sentence is decreased. We will briefly discuss these two devices.

High boundary tone. In the first place, it is clear that our listeners overwhelmingly relied on the presence of a final rise on the last syllable. Speakers, also, use this rise as their foremost intonational device. Investigation of our corpus of 600 Dutch questions (cf. p. 1) showed that this rise occurred in more than 95% of the yes-no questions and declarative questions; even in

wh-questions the percentage was as high as 63%, despite frequent claims that this question type typically lacks final rises. As the final rise has traditionally been regarded as the hallmark of question intonation, this finding does not come as a surprise.

A potentially more interesting question is whether in actual speech production there is a (causal) relationship between the presence of a final rise and the slope of the baseline of the utterance. In the earlier production study, when wh-questions featured final rises, baseline slopes tended to be shallower than in wh-questions ending in low pitch. Consequently, it seems plausible that the final rise, as planned by the speaker, has the effect of gradually raising the pitch minima in earlier parts of the question. As the results of experiment 2 showed, a (gently) rising baseline is an important additional (and early) perceptual cue to the interrogative character of the utterance. Still, we would argue that the upward slope of the baseline that we found in our production study, falls out as a by-product of an attempt on the part of the speaker to economize on articulatory effort: given that questions typically involve a large final accent followed by a high boundary tone, the speaker elevates his pitch minima towards the end of the utterance. So, in our view, what the speaker aims for is not an upward slope of baseline, but ease of articulation in the production of high pitch targets towards the end of the utterance.

Contrastive focus. Second, interrogativity was also found to be signalled by a relatively salient accent, in particular on the object. This confirmed our earlier impression (inspired by the inequality of consecutive accent peaks in the production data) that questioning involves contrastivity. Increasing the (relative) salience of an accent peak is an intonational device for encoding contrastive focus (e.g. Eady, Cooper, Klouda & Lotts 1986, Bartels & Kingston 1994; Rump & Collier 1996). This increased salience conveys that the reason for bringing the item in focus is to establish it as the correct item from an (implicit) set of alternatives (cf. Chafe 1976).

Our suggestion that the raising of accent peaks in questions likewise reflects contrastivity is built on assumptions that would seem uncontroversial. First, utterances are to be divided into the universal pragmatic constituents topic (representing what the utterance is ‘about’) and comment (representing what is being said with respect to that topic). Asking a question implies that the questioner has reasons to suppose that, with respect to a particular topic, some comment can be made. Accordingly, a provisional comment is proferred by the speaker with a view to checking its correctness. Second, pragmatic constituents to which a speaker chooses to call a listener’s attention, are brought in intonational focus. However, speakers may have different reasons for highlighting a constituent, viz. (i) to (re)introduce it to the current discourse (i.e., the topic is focused), (ii) to add new information (i.e., the comment is focused), and (iii) to indicate that it is the correct item from some subset (i.e., the comment is contrastively focused). That is, given the various discoursal functions of focus different focus types have to be distinguished.

In our view, questions typically differ from statements in the distribution of focus type. Considering that the comment constituents in questions do not seek to add new information but merely check correctness, questions allow only the contrastive focus type (iii), whose prosodic correlate is a raised accent peak. Thus, in yes-no/declarative questions the speaker checks whether his comment is the correct member of an implicit set of other potentially appropriate comments respecting this particular topic. In wh-questions the comment corresponds to the wh-phrase, which explicitly indicates a limited range of possibilities only one of which is correct (for more details, see Haan 2001). Independent evidence for this approach is afforded by clefting questions; in cleft structures, the highlighted element is claimed to have the full implication of

contrastive focus (e.g. Quirk, Greenbaum, Leech & Svartvik 1987). For example, an adequate paraphrase of the declarative question *Marina wil haar mandoline verkopen?* would be: 'Is it selling her mandolin that Marina wants to do?', reflecting contrastive focus on the comment.

The experiment that was described in this paper was not constructed to allow us to choose between the integral versus the compositional view on question intonation. Nevertheless, there is some circumstantial evidence in the data that would support the compositional view. It was shown repeatedly that a single large (8 or 12-st) accent on the subject (i.e., not followed by an accent on the object) yields fewer statement judgments than smaller subject accents or subject accents followed by larger than 0-st object accents. Note that such contrastive subject accents – which occur early in the utterance – cause the global melodic shape to be falling, which should yield more statement judgments. Apparently, then, any contrastive (i.e., single) accent – whether on the subject or on the object – seems to be (weakly) associated with interrogativity.

Finally, we should realise that the integral and the compositional accounts of the observations are not mutually exclusive. It may be argued that the integral mechanism reflects a (near) universal tendency based, for instance, on the iconic use of high pitch whereby high and rising pitch (characteristic of young and/or small creatures) is used to express subservience and dependence (as required when asking a question) and low pitch (characteristic of strong and large creatures) corresponds with dominance (Ohala, 1984).

Notes

1. The means underlying the F_0 pivot points in figure 1 were based on a selection of the production data such that data points that could not be realistically modelled in terms of a rise-fall accent configuration were excluded. For instance, missing data were entered for the third and fourth pivot points when the two rise-fall accents were linked into a flat hat contour, and for the sixth and seventh pivot points when the object accent was directly linked to H% without an intervening fall. Missing values were then replaced by the mean F_0 and relative timing values for the speaker concerned.

2. The baselines in figure 1 were drawn conservatively. The alternative would have been to draw linear regression lines through the six (in statements) or seven (in questions) non-terminal pivot points. The regression line for statements would have the same (downward) slope as in figure 1, but they would be considerably more steeply rising in the case of the declarative questions.

3. Moreover, Gussenhoven & Chen (2000) have shown that the rightward peak shift, even when it is not accompanied by a larger excursion size, provides a cue to the perception of interrogativity for native listeners of Dutch, Chinese and Hungarian, when judging stimuli in an unfamiliar language. However, the effect of peak shift was clearly smaller than that of excursion size (or peak height); even a 120-ms peak shift could not create a convincing cross-over from statement to question (43% question responses for 0ms, monotonically increasing to 57% question responses for 120 ms peak shift).

4. The correlation between percent perceived questions and quality of question prosody proved almost perfect. Therefore, we will only present results in terms of percent perceived questions.

Acknowledgements

This research was funded in part by the Netherlands Organisation for Research (NWO) under grant 200-50-073 (principal investigators: Renée van Bezooijen and Vincent van Heuven). The experiments were run by Rembrandt Hissink, Myrthe Lems and Els Nachtegaal as part of a research seminar in Experimental Phonetics at Universiteit Leiden supervised by the authors. We thank Bob Ladd for discussion and improvements.

References

Bartels, C. and J. Kingston (1994). 'Salient Pitch Cues in the Perception of Contrastive Focus'. In: P. Bosch and J. van der Sandt, eds. *Focus and Natural Language Processing*, IBM Working Papers of the Institute for Logic and Linguistics, 1-10.

Boersma, P. and D. Weenink (1996) 'Praat: a System for Doing Phonetics by Computer'. *Report of the Institute of Phonetic Sciences, University of Amsterdam* 132.

Chafe, W. (1976). 'Givenness, Contrastiveness, Definiteness, Subjects and Topics'. In: C. Li, ed. *Subject and Topic*. New York, Academic Press, 25-56.

Eady, S.J., W.E. Cooper, G.V. Klouda, P.R. Müller and D.W. Lotts (1986). 'Acoustical Characteristics of Sentential Focus: Narrow vs. Broad and Single vs. Dual Focus Environments'. *Language and Speech* 29, 233-251.

Gosy, M. and Terken, J. (1994). 'Question Marking in Hungarian: Timing and Height of Pitch Peaks'. *Journal of Phonetics* 22, 269-281.

Grosjean, F. (1983). 'How Long is the Sentence? Prediction and Prosody in the On-line Processing of Language'. *Linguistics* 21, 501-509.

Gussenhoven, C. and A. Chen (2000) 'Universal and Language-specific Effects in the Perception of Question Intonation'. *Proceedings of the Sixth International Conference on Spoken Language Processing*, Beijing, 2, 91-94

Gussenhoven, C., B.H. Repp, A. Rietveld, H.H. Rump and J. Terken (1997) 'The Perceptual Prominence of Fundamental Frequency Peaks'. *Journal of the Acoustical Society of America* 102, 3009-3022.

Haan, J. (2001) *Speaking of Questions*. Doctoral dissertation, Nijmegen University. LOT Dissertation Series (in preparation).

Haan, J., V.J. van Heuven, J.J.A. Pacilly and R. van Bezooijen (1997) 'On the Anatomy of Dutch Question Intonation'. In: H. de Hoop and J. Coerts, eds. *Linguistics in the Netherlands 1997*. John Benjamins, Amsterdam, 99-110.

Hart, J. 't, R. Collier and A. Cohen (1990) *A Perceptual Study of Intonation*. Cambridge University Press, Cambridge.

Hermann, E. (1942) *Probleme der Frage*. Nachrichten Akademie von Wissenschaft. Göttingen.

Heuven, V.J. van and J. Haan (2000) 'Phonetic Correlates of Statement versus Question Intonation in Dutch'. In: A. Botinis, ed., *Intonation: Analysis, Modelling and Technology*, Kluwer, Amsterdam, 119-143.

Lindsey, G. A. (1985) Intonation and Interrogation: Tonal Structure and the Expression of a Pragmatic Function in English and Other Languages. Ph.D. dissertation, University of

California, Los Angeles.

Makarova, V. (1999) 'Pitch Peak Alignment in Russian Declaratives, Interrogatives and Exclamations'. *Proceedings of the 16th International Congress of Phonetics Sciences*, San Francisco, 1173-1176.

Moulines, E. and W. Verhelst (1995) 'Time-domain and Frequency-domain Techniques for Prosodic Modification of Speech'. In: W.B. Kleijn and K.K. Paliwal, eds., *Speech Coding and Synthesis*. Elsevier Science, Amsterdam, 519-555.

Ohala, J.J. (1984) 'An Ethological Perspective on Common Cross-language Utilization of F0 in Voice'. *Phonetica* 41, 1-16.

Quirk, R., S. Greenbaum, G. Leech and J. Svartvik (1987) *A Grammar of Contemporary English*. Longman, Harlow (Essex, UK).

Rump, H.H. and R. Collier (1996) 'Focus Conditions and the Prominence of Pitch-Accented Syllables'. *Language and Speech* 39, 1-17.

Sadock, J. and A. Zwicky (1985) 'Speech Act Distinctions in Syntax'. In: T. Shopen, ed., *Language Typology and Syntactic Description, vol. 1*. Cambridge University Press, Cambridge, 155-196.

Thorsen, N. (1980) 'A Study of the Perception of Sentence Intonation – Evidence from Danish', *Journal of the Acoustical Society of America* 67, 1014-1030.

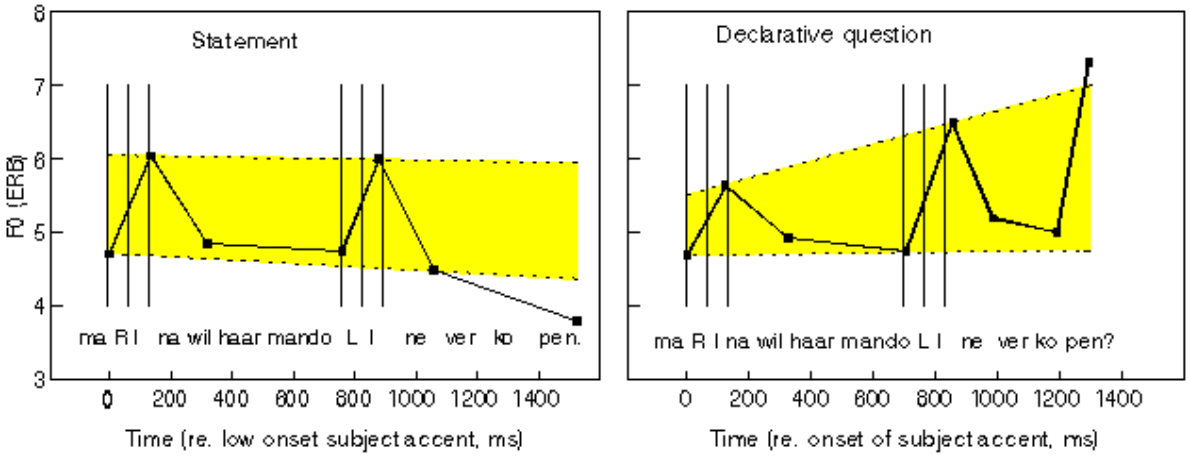

Figure 1. Stylised pitch contours (F_0 in ERB) for statements and declarative questions averaged over five male and five female speakers. Each data point nominally represents 100 measurements. Vertical lines represent the onset and nucleus segments of accented syllables. Low declination lines were drawn connecting the low onset pivot point to the last relevant low pivot. Note that the L% target extends clearly below the low declination line. The high declination lines connect the accent peaks on subject and object.

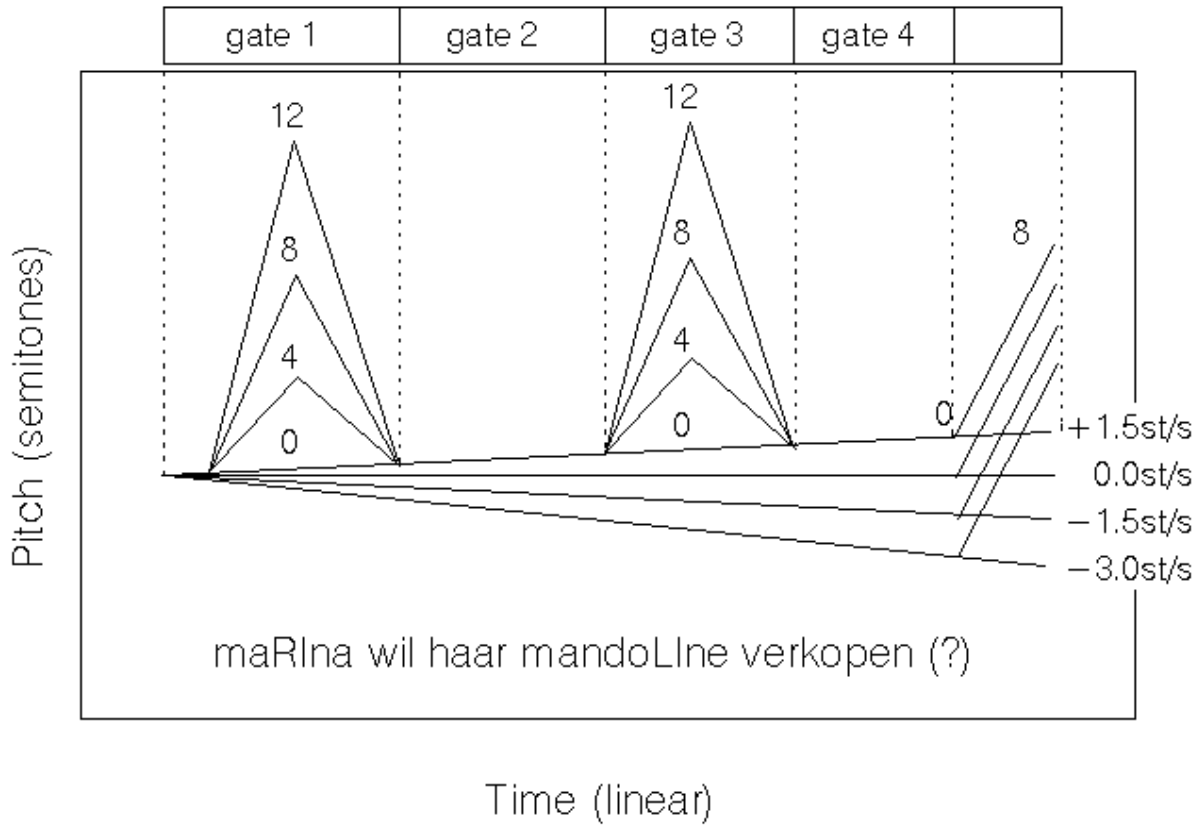

Figure 2. Schematic representation of stimulus materials used in experiments 1 and 2. The 4 H *4 combinations of rise-fall accents were superposed on each of the 4 baselines.*

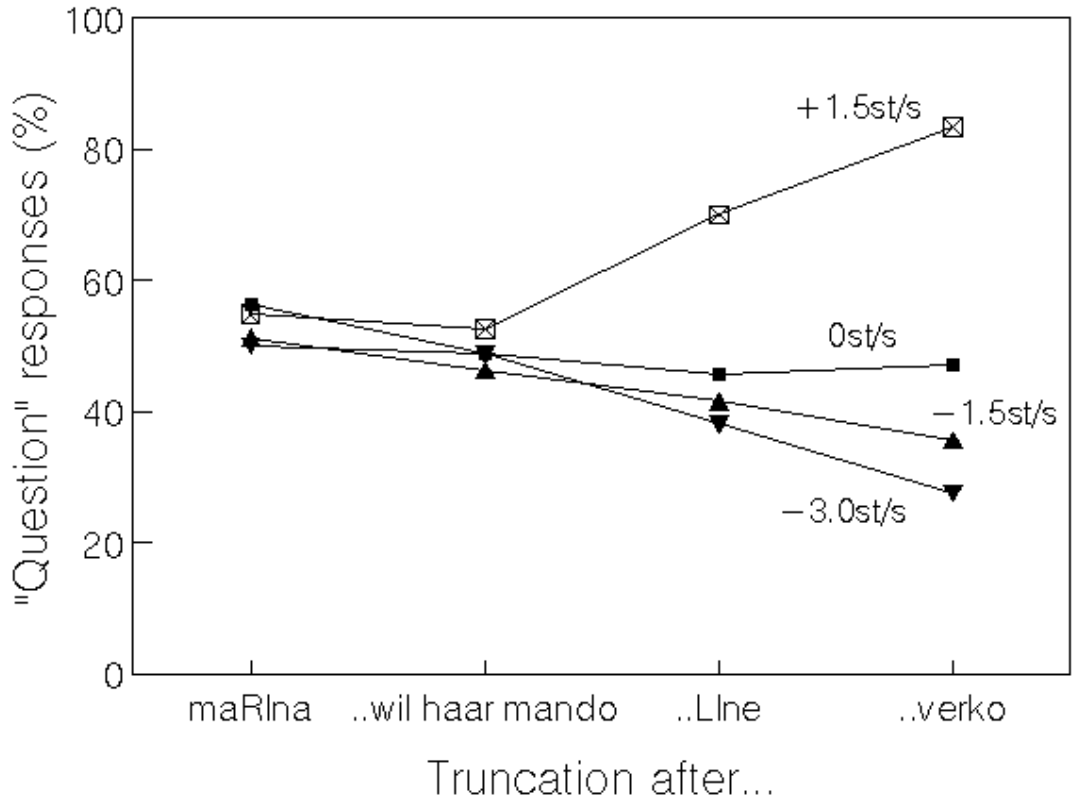

Figure 3. Percent 'question' responses as a function of truncation point, broken down by baseline slope.

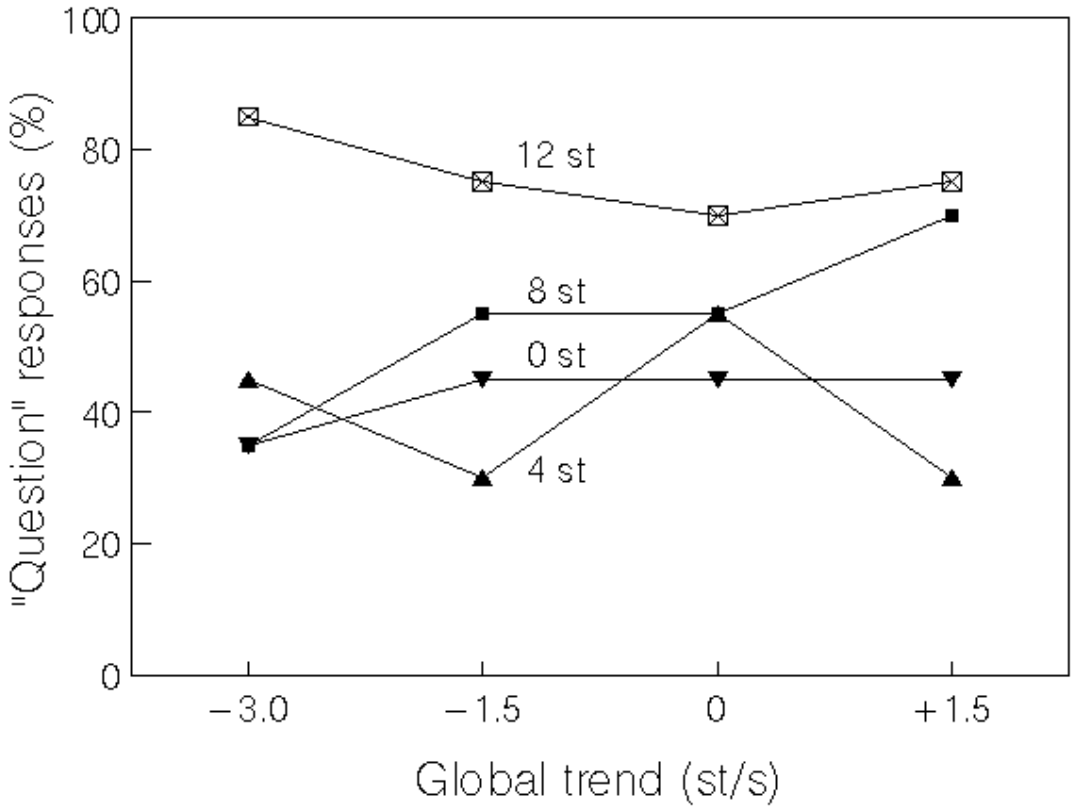

Figure 4. Percent question responses as a function of size of subject accent and global pitch trend, at truncation point 1.

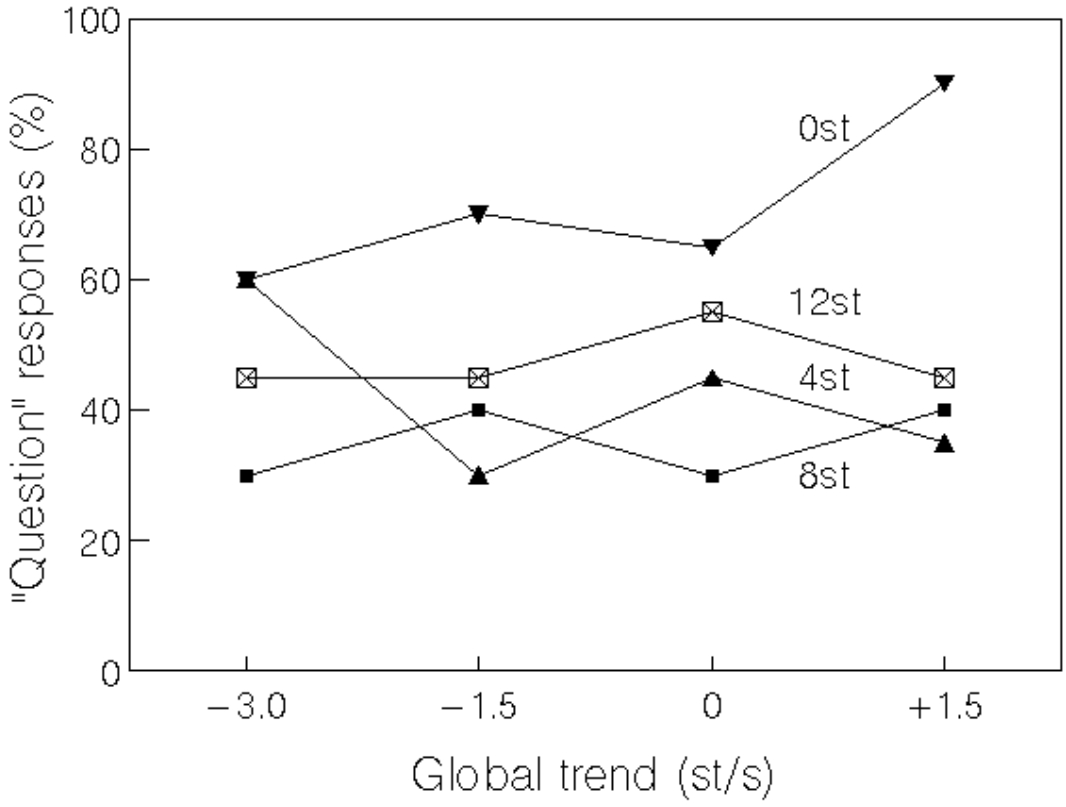

Figure 5. As figure 4, at truncation point 2.

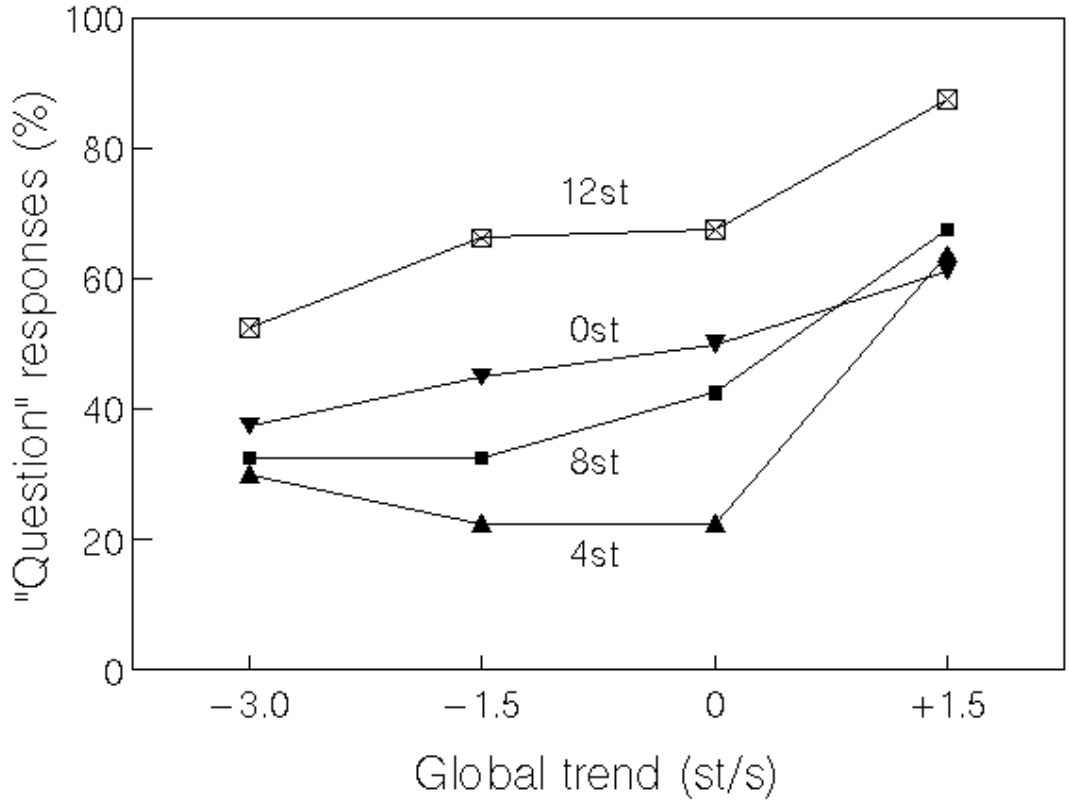

Figure 6. Percent 'question' responses as a function of baseline slope and size of object accent, at truncation point 3.

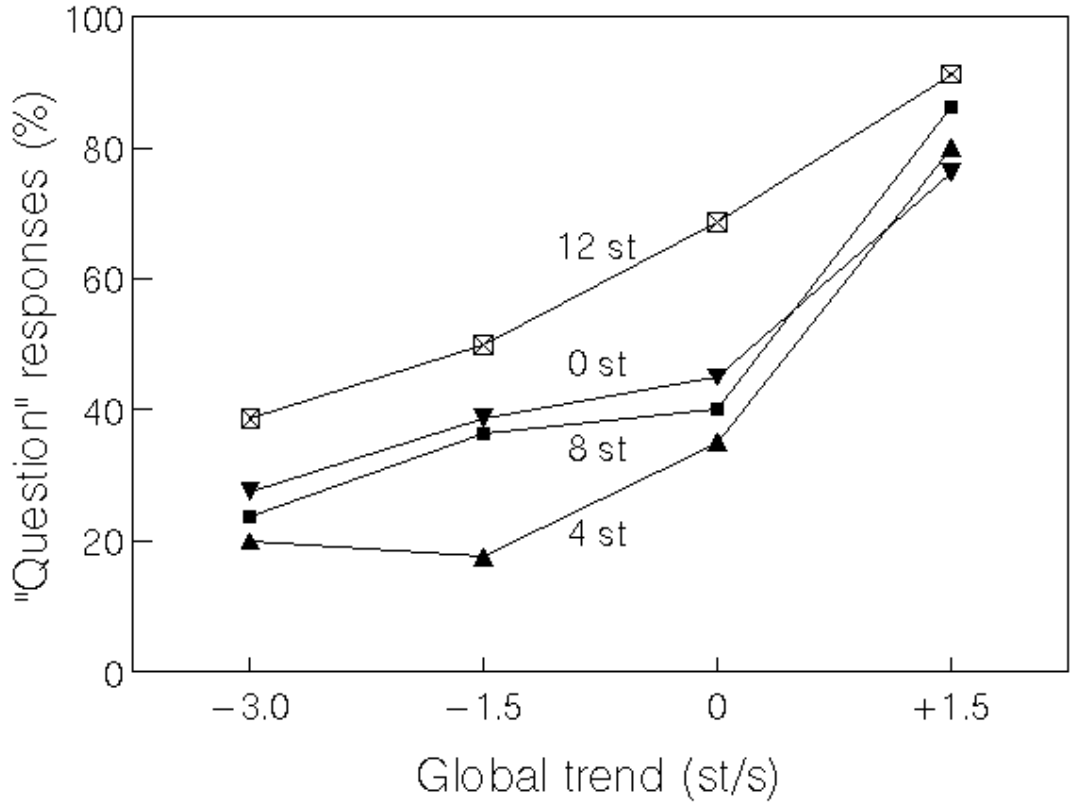

Figure 7. As figure 6, at truncation point 4.

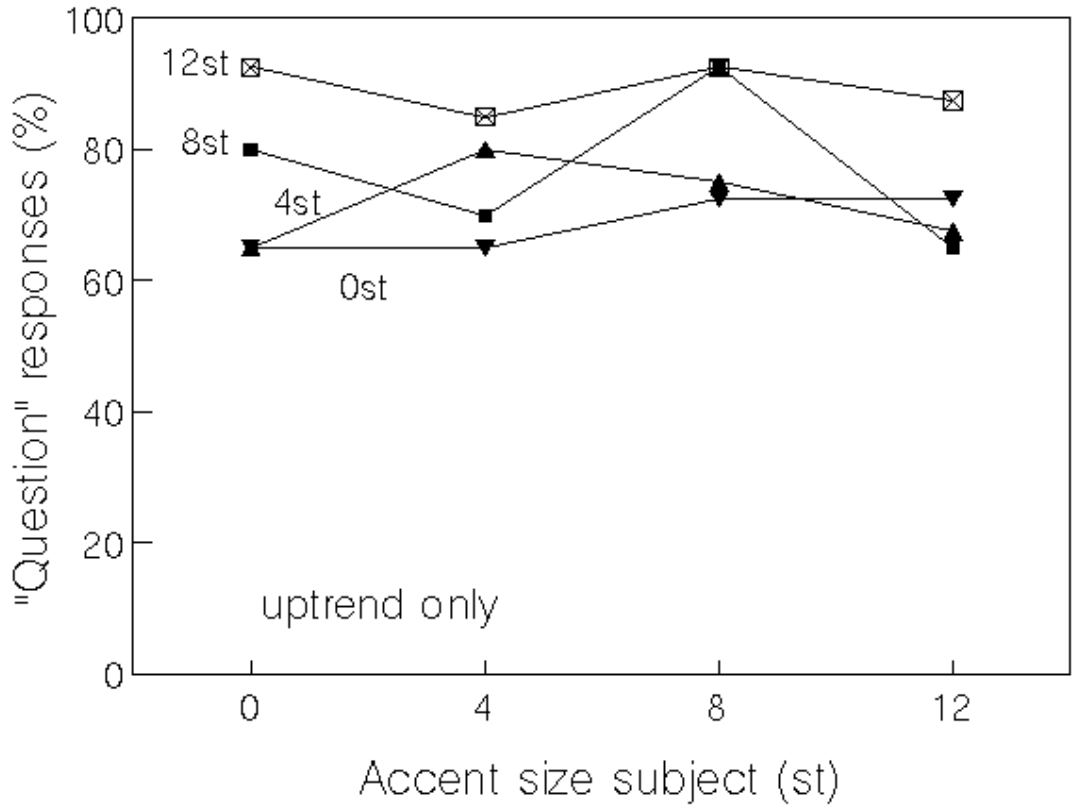

Figure 8. Percent 'question' responses broken down by size of object accent (across) and of subject accent (separate lines), at truncation points 3 and 4 combined. The baseline slope is upwards.

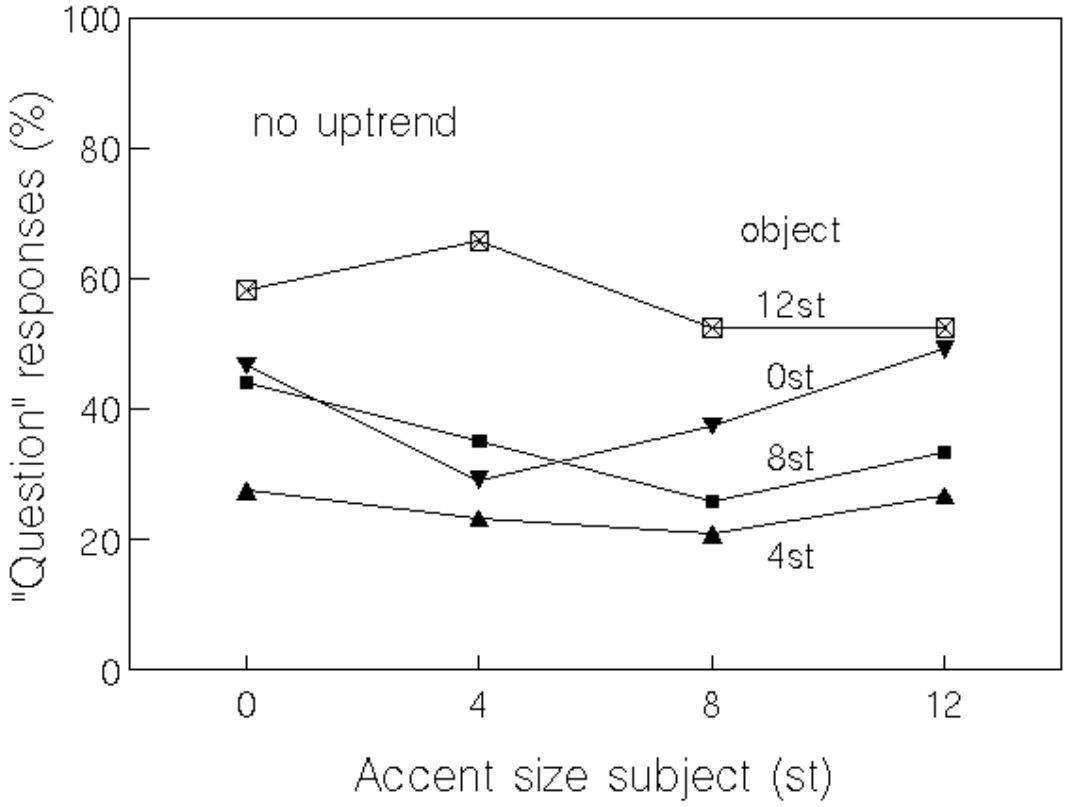

Figure 9. As figure 8, but baseline slope is either flat or falling.

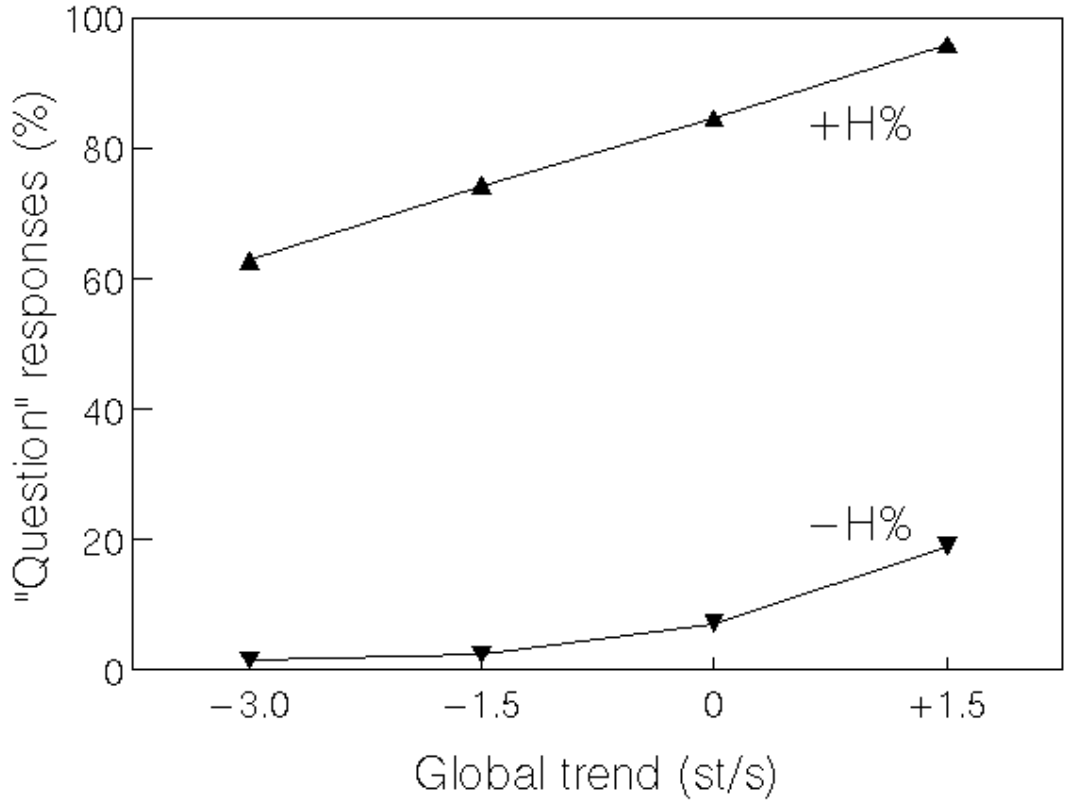

Figure 10. Percent 'question' responses as a function of global pitch trend, broken down by presence versus absence of H% (complete sentences).