When and how do we hear whether a Dutch speech utterance is a statement or a declarative question?*

Vincent J. van Heuven and Judith Haan
Universiteit Leiden/HIL / Nijmegen University/CLS

1. Introduction

When interlocutors are engaged in a dialogue, 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 piece of information that is explicitly targeted by the speaker). The signalling of statement versus question seems universal; moreover, it is in the large majority of the world's languages expressed by lexical and or syntactic devices, as well as signalled by intonational means. Here the statement type is generally considered the unmarked choice, requiring no special signs. Question is the marked sentence type, signalled lexically by a choice of question words — in Germanic languages typically beginning with orthographic *w-* (*wat, wie, wanneer, waar*, etc.) or *wh-* (*what, who, when, where*, etc., hence the term *wh*-words) — or by dedicated syntactic means such as inversion of subject and finite. Whilst the lexico-syntactic question devices may vary considerably across languages, the one and only true universal lies in the prosodic interrogativity marking: it has been claimed that questions are universally differentiated from statements by the occurrence of a high-pitched element somewhere in the course of the question utterance (Hermann 1942). The high-pitched element is typically postulated at the end of the question utterance, a state of affairs that may correspond to the orthographic practice of writing the question mark at the end of the question. However, there are descriptions of non-Western languages where the high element is not sentence final (Lindsay 1985), and in fact, one may speculate that also for Western languages there may be high-pitched elements at some earlier point in the utterance (Haan, van Heuven, Pacilly and van Bezooijen 1997 and references therein).

Haan et al. (1997) presented the first systematic analyses of the Dutch sentence melodies of statements and of three types of questions derived from these statements, i.e., *wh*-questions (also called question word questions or information questions), yes/no-questions, and declarative questions (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 each other at some earlier point in the time course of the utterance. Figure 1 shows averaged stylized pitch patterns for the statement and declarative question versions of utterances *maRIna wil haar mandoLIne verkopen(?)* 'Marina wants her mandolin sell(?)' and *reNEE heeft nog VLEES over(?)* 'Renee has some/any meat left(?)', spoken two times by five males and five females in three different positions in a short paragraph (van Heuven, Haan and Pacilly 1998).

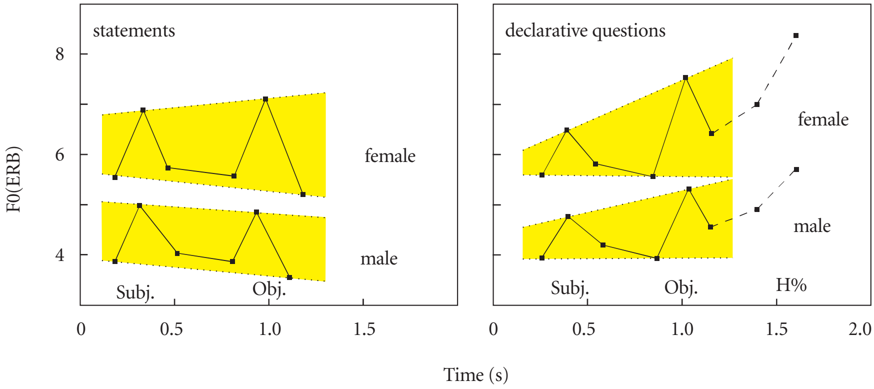

Figure 1. Stylised pitch contours (F_0 in ERB) for statements and declarative questions, drawn separately for male and female speakers. Each datapoint represents 100 measurements.

Five major differences are visible between the statement and question versions:

- The question ends with a high boundary tone H%; the statement with the low boundary tone L%.
- 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 two accents in the statement are of equal size; a quantum of pitch accent is taken away from the subject and transferred to the object in the question version.

- The statement shows downtrend in the (imaginary) line connecting the low-pitched turning points in the contour, whilst such downtrend is absent in the question version.

Van Heuven, Haan, Jansen and van der Torre (1997), using a gating task (see below), showed that Dutch listeners reliably hear 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 which cue or combination of cues enables the listener to perform this feat. Is it the unusually 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)? 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 cooccur. We need a way to disentangle the cues, 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 and Verhelst 1995) that allow us to create intonation patterns of our own choosing, and superpose these onto a prerecorded human utterance without any audible loss of sound quality or naturalness.

We will create a multidimensional stimulus space by introducing controlled variations along four dimensions:

1. Size of subject accent: varying from absent to extremely large.
2. Size of object accent: same as on subject.
3. Slope of imaginary trendline 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 distribution of accents. Typically, when a listener asks a question, (s)he targets a single referent; questions generally place only one constituent in focus, and assume all other referents as given. Consequently, when the listener hears (or expects) a single (contrastive) accent, (s)he will be biassed into assuming that the speaker intends a question.

We will run two experiments. In the first experiment the subjects hear only part of the sentences. These will be truncated at one of several carefully chosen points in time, just to see how well the listener differentiates between statement and declarative question at each of these points in the temporal development of the intonation pattern. This is the so-called gating technique, an experimental paradigm that enjoys increasing popularity in 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 to listener to generate such projections (Grosjean 1983).

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

2. Method

Stimulus material. The entire material for both experiments was constructed from a single question utterance *maRIna wil haar mandoLIne verkopen?* [maːˈriˑnaː ʋɪl haːr mɑndoːˈliˑnə vərkoːpə] 'Marina wants her mandolin sell', i.e., 'Marina wants to sell her mandolin', 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 an H% final high boundary tone. 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 and Weenink 1996) speech processing package. The pitch curve was interactively stylized 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 stylized pitch curves. 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 *maRIna* was varied in 4 steps: 0, 4, 8 and 12 semitones (st) above the low reference line.
2. *Size of object accent.* The excursion on *mandoLIne* was likewise varied in four steps (same values as subject accent).
3. *Downtrend.* The slope of the imaginary low declination line, connecting the low pivot points in the stylized 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, yielding a natural sounding range of downtrend varying from rather steeply dropping, through level pitch, to slightly rising.

4. *Boundary tone.* An H% final boundary of 8 st was either present or absent.

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

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

Gate 2. Truncation point at the onset of the rise belonging to the second accent; same 16 combinations of subject accent and downtrend. This yielded the second block of 20 trials (including four practice trials).

Gate 3. Truncation point at the offset of the object accent; variation of subject accent, object accent and downtrend yielded 64 combinations. The third block was preceded by 1 practice trial.

Gate 4. Truncation point at the onset of the terminal rise; same 64 combinations as in gate 3 (plus 1 practice trial).

The total of 170 stimuli were played back in groups of 10 trials, with 3-second intervals between trials (offset to onset) and a 5-second interval plus a beep between groups.

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

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. The four blocks of stimuli were played to these listeners such that shorter gates preceded longer gates. 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 a 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 examplar of a statement or question, depending on their response to (i).[1] Experiment 2 lasted about 20 minutes.

3. Results

3.1 Experiment 1: gated sentence fragments

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 2 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 lower trendline (but accumulated across all sizes of subject and object accents).

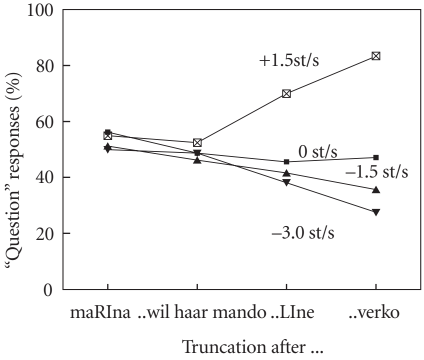

Figure 2. Percent "question" responses as a function of truncation point, broken down by global trend.

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 global trend clearly triggers more question responses than the level and falling trends. 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. These data show that global trend provides an important perceptual cue for the sentence type (statement versus question) well before the end of the sentence is reached. Since the effects of excursion size of subject and object accent cancel out in the present pooled data, there is an independent interrogative cue of global pitch trend, which is first picked up by the listener at the third truncation point.

Figure 3 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 the slope of the gobal pitch trend. The results reveal no effect of global trend at this early point in the utterance, as is to be expected given the absence of any overall effect of global trend 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

(12st, which is twice the normal size of a Dutch pitch accent, cf. ’t Hart, Collier and Cohen 1990) do we find a clear propensity on the part of our listeners to project an interrogative utterance. We assume that such an exceptionally large excursion size is interpreted as a contrastive accent on the subject, which — apparently — is conducive to projecting a question. Note that this finding runs counter to the accent scaling found in Haan et al. (1997, cf. Figure 1), i.e., small accent on the subject followed by a large accent on the object. Possibly (declarative) questions are characterized by a single (contrastive) accent that is typically on the object. We assume, then, that the specific constituent that is focused by that accent is not crucial to the projection of interrogativity; an unusually large accent on some earlier constituent has the same effect (but questions an earlier constituent). In the present stimulus sentence, a large accent on the subject would suggest ‘Did you really say that it is Marina, rather than someone else, who wants to sell her mandolin?’. In the alternative situation, with the single, large accent on the object, the intended focus of the question would seem ambiguous; it is either ‘Did you really say that Marina wants to sell her mandolin, rather than some other object?’ (narrow focus on object) or ‘Did you really say that Marina wants to sell her mandolin, rather than some other statement?’ (broad focus on entire predicate or sentence).

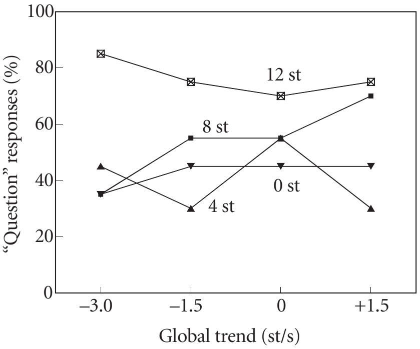

Figure 3. Percent question responses as a function of size of subject accent and global pitch trend, at truncation point 1.

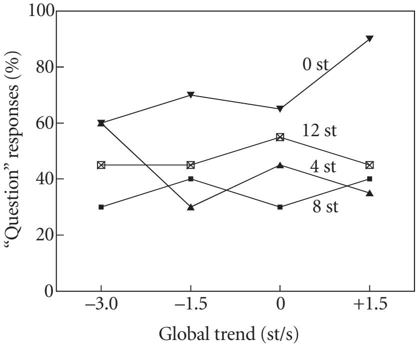

Figure 4. As Figure 3, at truncation point 2.

Figure 4 is organised in the same way as Figure 3 but now the truncation point is delayed to the onset of the object accent; as a result, the effect of the global trend might be stronger in Figure 4 than in Figure 3, whilst the effect of the size of subject accent should remain the same. Again, there is no systematic overall effect of global trend (which will not be manifest until gate 3 is made audible, see Figure 2). This time, however, the stimuli with 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). We assume that our listeners know that sooner or later the utterance must contain an accent; when no accent is heard before the truncation point, there must be a (relatively) strong accent on the object — even if it has not yet been made audible. Therefore, the 0-st subject accent heralds a strong accent on the object, and hence triggers the question response. A 12-st accent on the subject is indifferent between statement and question, while intermediate 4 and 8-st accents on the subject project statement rather than question. Finally, note that global trend has a considerable effect (60% question responses for downtrend to 90% questions for uptrend, and intermediate values for flatter slopes) if there is no accent on the subject. It seems, therefore, that the information in the global trend can be picked up by the listener at a relatively early point in time, but only if the trend is not interrupted by (accent-lending) pitch obtrusions.

The next two figures present the effects of global trend and size of object accent on percent question responses, with the data accumulated across subject accents. Figure 5 presents the data at truncation point 3 (object accent first audible); Figure 6 shows the data at gate 4, including the stretch of global trend line up to the point in the time where the final rise would begin.

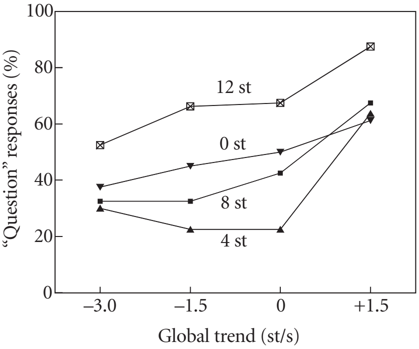

Figure 5. Percent "question" responses as a function of global trend and size of object accent, at truncation point 3.

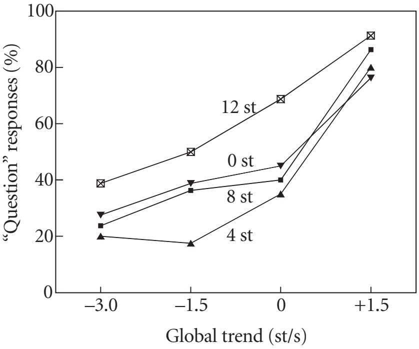

Figure 6. As Figure 5, at truncation point 4.

Very clearly, the size of the object accent has a strong effect on the projection of interrogativity. The interrogativity percept is strongest for the 12-st object accent, and grows progressively weaker (in fact crossing over to statement for the smaller object accents of 8 and 4 st. This effect is fully in line with the expectation that would be derived from the acoustic profile of the declarative question type in Figure 1. 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. We interpret this as yet another indication that hearing a single accent, whether on the object or on the subject, generates the expectation of a question. Finally, there is a small effect due to global trend, which is only to be expected given the earlier discussion of Figure 2. At truncation point 4 (Figure 6) we find basically the same effect of size of object accent; here, however, the effect of global trend has become much stronger than at gate 3. In fact, some 80% questions are projected regardless of size of object accent if the general pitch trend is upwards.

In Figures 7 and 8 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 global trend has a strong effect as of gate 3 (see Figure 2), we present the results separately for global uptrend (Figure 7) and for the flat and falling trends combined (Figure 8).

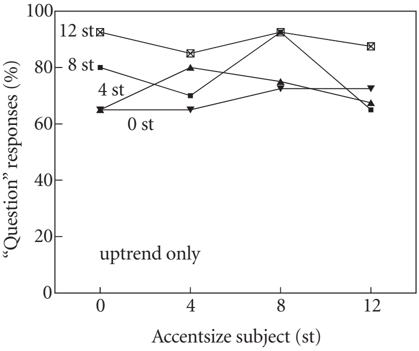

Figure 7. Percent "question" responses broken down by size of object accent (across) and of subject accent (separate lines), at truncation points 3 and 4 combined. The global trend is upwards.

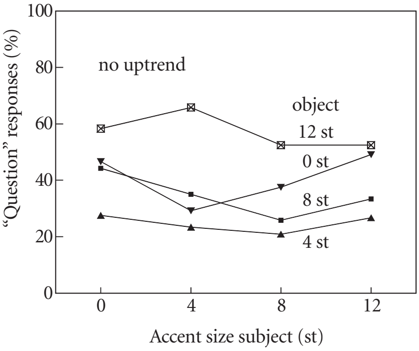

Figure 8. As Figure 7, but global trend is either flat or falling.

When the global pitch trend is upwards (Figure 7), question perception is always above 60%, showing, once more, the strength of the global trend parameter. Generally, we observe no systematic overall effect of size of subject accent. The size of the object accent, however, matters: 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 percentages around 80 are found with object accents of 4 and 8 st, without any systematic interaction with subject accent size.

When the global pitch trend is either flat or falling (Figure 8), 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 increases the chances of a question response (50% maximally). Once more, we take this as an indication that a single-accent sentence is conducive to projecting interrogativity. Finally, and predictably, there is no overall effect of subject accent.

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 would like 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 9, 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.

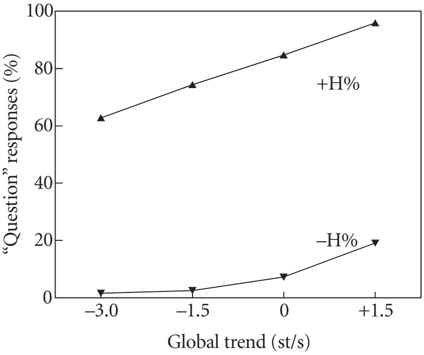

Figure 9. Percent "question" responses as a function of global pitch trend, broken down by presence versus absence of H% (complete sentences).

The terminal pitch rise exerts an almost categorical effect. If H% is present, there is always a majority of question responses; when H% is absent, the perceived sentence type is statement. However, there is a quite noticeable effect of the global pitch trend. First, utterances without a terminal rise are less unanimously perceived as statements when the global trend is flat or upwards. Second, and more importantly, utterances bounded by H% are perceived in close to 100% as questions only if the global pitch trend is upwards. Each time the global trend is decreased by a quantum of 1.5 st/s, the number of perceived questions drops by 10%, leaving a mere 60% question responses for +H% utterances with the steepest global downtrend (−3 st/s).

For lack of space we will not present any figures illustrating the effects and interactions of the excursion sizes of subject and object accents. Instead, we will just offer some verbal descriptions of the main findings here.

For sentences without the terminal rise, we do not obtain any effects of either subject or object accent when the global pitch trend is flat or falling: all stimuli are perceived as clear statements here. However, when the global trend is upwards, the size of the object accent matters: there are some 40% question responses when the object accent is large (12 st), and even fewer for 8-st (20%) and 4-st (5%) object accents. When there is no object accent (0 st), a (therefore single) accent on the subject generates roughly 25% question responses, indicating once more that contrastive accentuation suggests interrogativity.

For complete utterances ending in a terminal rise, we find no further effects of subject and object accents if the global pitch trend is upward (ceiling effect). When the global trend 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. 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:

- Statements are differentiated from questions before the end of the sentence; differentiation is found by the time the accent on the object is heard.
- Global pitch trend is one important cue that distinguishes statement (falling trend) from question (rising trend).
- Utterances that have, or are expected to have, a single (large) accent are more readily interpreted as questions than utterances with two equal (and 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 on-line generate expectations as to the speech act (making a statement, or asking a question) 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, i.e., the communication of interrogativity from speaker to listener suffers severely.

These conclusions suggest, finally, 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.

Notes

* This research was funded in part by the Netherlands Organisation for Research (NWO) under grant 200–50–073. The experiments were run by Rembrandt Hissink, Myrthe Lems and Els Nachtegaal as part of a research seminar at Universiteit Leiden supervised by the authors.

[1] 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.

References

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.

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

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, J. Haan, E. Janse and E.J. van der Torre (1997) "Perceptual Identification of Sentence Type and the Time-distribution of Prosodic Interrogativity Markers in Dutch". *Proceedings of an ESCA Workshop on Intonation*, Athens, 317–320.

Heuven, V.J. van, J. Haan and J.J.A. Pacilly (1998) "Global and Local Characteristics of Dutch Questions in Play-acted and Spontaneous Speech". *Proceedings of an ESCA Workshop on Sound Patterns of Spontaneous Speech*, La Baume-les-Aix, 139–142.

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.

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.