ALIGNMENT OF "PHRASE ACCENT" LOWS IN DUTCH FALLING-RISING QUESTIONS: THEORETICAL AND METHODOLOGICAL IMPLICATIONS

Robin J. Lickley*, Astrid Schepman** and D. Robert Ladd***

*Queen Margaret University College, Edinburgh

**University of Abertay, Dundee

*** Edinburgh University, Edinburgh

Running Head: Dutch phrase accent lows

Acknowledgements

The research reported here was funded by the UK Economic and Research Council (ESRC) under grant no. R000-23-7447 to Edinburgh University, Principal Investigator D. R. Ladd. We are grateful to: Ellen Gurman Bard, for useful discussions about how to use the Map Task; Ineke Mennen for her help in devising sentences for the *woon je in* task; Andrew Ladd, for preparing new drawings for the Dutch maps; Carlos Gussenhoven, for providing access to the Nijmegen phonetics lab and its helpful staff, and for helping us to recruit speakers; Angela Vonk, for transcribing the Map Task corpus; Michael Bennett, Eddie Dubourg, Cedric Macmartin, and Stewart Smith, for much technical assistance in Edinburgh; the audience at the Colloquium of the British Association of Academic Phoneticians in Newcastle in March 2002, who made comments on a preliminary version of this paper; Francis Nolan and Vincent van Heuven, who read the paper as referees for *Language and Speech* and made many important suggestions for improvement; and above all our speakers, without whom there would have been nothing to study.

Address for correspondence:

Dr Robin J. Lickley, Speech and Hearing Sciences, Queen Margaret University College, Edinburgh, EH12 8TS, Scotland, UK. Email: rlickley@qmuc.ac.uk.

Abstract

In the first part of this study, we measured the alignment (relative to segmental landmarks) of the low F0 turning points between the accentual fall and the final boundary rise in short Dutch falling-rising questions of the form *Do you live in [place name]?* produced as read speech in a laboratory setting. We found that the alignment of these turning points is affected by the location of a post-accentual secondary stressed syllable if one is present. This is consistent with the findings and analyses of Grice et al. 2000 (*Phonology* 17: 143-185), suggesting that the low turning points are the phonetic reflex of a "phrase accent". In the second part of this study, we measured the low turning points in falling-rising questions produced in a task-oriented dialogue setting and found that their alignment is affected in the same way as in the read speech data. This suggests that read speech experiments are a valid means of investigating the phonetic details of intonation contours.

Keywords: Prosody, intonation, F0, alignment, Dutch

INTRODUCTION

This paper has a dual purpose. Our primary aim is to demonstrate, on the basis of experimental data, that the low fundamental frequency (F0) turning points in Dutch falling-rising question intonation are very precisely aligned with the segmental string, and that their alignment depends on the location of any post-nuclear secondary stresses. In so doing, we add to the growing body of data showing that the temporal alignment of turning points in F0 contours is highly lawful and apparently phonologically determined (e.g., Silverman & Pierrehumbert 1990; Prieto, van Santen & Hirschberg 1995; Arvaniti, Ladd & Mennen 1998; Xu 1998; Ladd, Mennen & Schepman 2000; D'Imperio 2000, 2001; Willis 2003). More specifically, we provide evidence for the claim (Grice, Ladd and Arvaniti 2000) that post-nuclear low F0 turning points in a falling-rising contour are the reflex of a "phrase accent" in the sense of Pierrehumbert 1980, and that it seeks to align with a secondary stress if one is available. In this aim, the paper is a mainstream laboratory phonology study, providing both basic data about the realisation of a specific intonation contour and empirical support for a particular analysis of the phonological structure of intonation.

Our secondary aim is broader: to defend the validity of studying intonation using the methods of laboratory phonology. Doubts are sometimes raised about the "ecological validity" of conclusions based on the analysis of intonation in controlled sentences read aloud under laboratory conditions, and more generally about the intonation of "decontextualised" sentences. We argue that these concerns, though by no means entirely misplaced, are not sufficiently serious to discredit laboratory phonology methods. We base this conclusion on an analysis of naturally occurring questions from a corpus of Dutch task-oriented dialogues. We used the corpus in two ways: first, we compared the phonetic details of naturally occurring falling-rising questions with the read questions ("lab speech") on which our main phonological conclusion is based; and second, we analysed the syntactic, pragmatic and phonological conditions that favour the use of falling-rising (as opposed to falling or rising) contours on the questions in the corpus. Neither analysis gives us any reason to believe that the read speech differs from questions produced in the dialogues with respect to the variables of interest.

STUDY I – ALIGNMENT OF PHRASE ACCENT MINIMA

Background

Alignment of F0 and segmentals

Several studies since about 1990 have demonstrated that turning points in F0 contours (such as maxima and minima) are consistently aligned relative to identifiable "anchor points" in the segmental string (such as the edges of a stressed syllable). One specific line of research on this topic was initiated by Arvaniti et al. 1998, who studied the alignment of pre-nuclear accentual rises in Greek. They found that the beginning of such rises is aligned with the beginning of the accented syllable and the end of the rise is aligned early in the following unstressed vowel. This regularity holds true regardless of the actual segmental duration of the stressed syllable and any following consonants. That is, the F0 rise itself exhibits neither fixed duration nor fixed slope: its duration and slope are determined by the alignment and F0 level of the segmentally-anchored "tonal targets" that define the beginning and end of the rise. The regularities of segmental anchoring can be overridden in cases of "tonal crowding", when the accented syllable is close to a following accented syllable ("stress clash") or, for some speakers, when the accented syllable is immediately followed by a word boundary (cf. Silverman and Pierrehumbert 1990, Caspers and van Heuven 1993, Prieto et al. 1995, Schepman, Lickley and Ladd, in press). However, given sufficient space for full realisation, it appears that the most consistent phonetic regularity of intonational pitch movements is the "segmental anchoring" of the local maxima and minima that define them.

Arvaniti et al.'s finding of consistent segmental anchors for F0 turning points has since been extended to several other languages, including English (Ladd, Faulkner, Faulkner & Schepman 1999), Dutch (Ladd, Mennen & Schepman 2000), German (Atterer & Ladd 2004), and Japanese (Ishihara 2003). Similar results have been obtained independently for Chinese by Xu (e.g., 1998), for Spanish by, e.g., Willis (2003) and for Italian by, e.g., D'Imperio (2001). Many of these findings, however, relate to the alignment of F0 maxima or local peaks, and many of them are based on pre-nuclear rising pitch accents at or near the beginning of sentences read aloud. F0 peaks in this context have the methodological advantage of being easy to identify, but in order to establish the generality of the phenomenon of segmental anchoring of F0 turning points, we need eventually to investigate other turning points in other sentence contexts – to move beyond the narrow range of intonational phenomena for which segmental anchoring has been reported. With this in mind, this paper studies the alignment of post-nuclear low turning points in falling-rising questions in Dutch.

Dutch question intonation

Question intonation in Dutch has been extensively studied, most recently by Haan (2001). On the basis of Haan's investigations – which like ours are based on experimental materials read aloud under laboratory conditions – and from other sources (such as 't Hart, Collier & Cohen 1990), it seems clear that there are at least two distinct types of question intonation in Dutch. In traditional British terms (e.g., O'Connor and Arnold 1973) we might call these "fall-rise" and "rise". In the notation for Dutch intonation made familiar by 't Hart and his colleagues, the fall-rise has a Type A or 1A ("pointed hat") accent on the nuclear syllable, followed by a Type 2 boundary rise, whereas the rise has a Type 1 accent on the nuclear syllable, followed by a Type 2 boundary rise. It may be appropriate to subcategorise the "rise" further; in the Pierrehumbert-style ToDI notation proposed by Gussenhoven et al.

(2003) and illustrated in Figure 1, (from Haan 2001), the fall-rise is H*+L H% while the rise is variously H* H%, L* H% and L*+H H%. We will not be concerned with the further analysis of rises in this paper, but see Gussenhoven and Rietveld (2000) and Haan (2001, esp. pp. 111-113 and 153-161) for data and discussion.

INSERT FIGURE 1 ABOUT HERE

We will refer to the contour under investigation throughout the paper as the "question fall-rise" or QFR. The QFR occurs in a variety of interrogative contexts in Dutch, and to the extent that cross-language comparison is possible, it appears to be similar or identical to falling-rising question contours in German and (especially British) English. (For English, it is important to distinguish the QFR from what we might call the "implicational fall-rise"; this is the distinction Halliday (1967) drew between his Tone 2 (question) and Tone 4 (implication).) Haan (2001: 113) reports that in Dutch the QFR "apparently … represents the obvious choice" for short question sentences read aloud, and that the various types of rise "can be seen as more idiosyncratic"; in her data the QFR occurred on more than two-thirds of the yes-no questions. The precise pragmatic effect of the QFR is difficult to pin down, but impressionistically it is common in all three languages with short yes-no questions at the opening of a transaction or the beginning of a new topic (e.g., *Could I speak with Mary?* or *Have you got aubergines today?*) (For some discussion, see Ladd 1996: 122f on English and Féry 1993:91 on German.) We are not concerned with the pragmatic function of the QFR in Study I, but with its phonological structure and phonetic realisation. However, the fact that it seems especially appropriate for short questions on new topics may explain why we were able to elicit it quite consistently in a sentence-by-sentence reading task without giving our speakers any intonational instructions other than the question mark at the end of the sentences to be read. We return to this point in Study II.

The status of post-nuclear low in the QFR

In an autosegmental description of intonational phonology, there are two basic possibilities for analysing the post-nuclear low of the QFR (i.e., the F0 minimum between the fall and the rise). One is to treat it as the reflex of a phrase accent, and the other is to take it as the endpoint of a falling nuclear accent. The first approach is based on the analysis of English presented by Pierrehumbert (1980), according to which the nuclear accent is high (e.g., H*) and is followed by a separate low phrase accent (L-) and a high boundary tone (H%). The second approach, which is more in keeping with older intonational descriptions based on pitch movements (e.g., ’t Hart and Collier 1975 for Dutch, O’Connor and Arnold 1973 for English, and Isačenko and Schädlich 1970 or Pheby 1975 for German), treats the nuclear accent as a fall (e.g., H*+L) followed by a high boundary tone (H%), and dispenses with the notion of the phrase accent altogether. The analysis without the phrase accent is adopted for Dutch by ToDI and by Haan; it is also the analysis proposed by Féry (1993) and Grabe (1998) for German, and by Ladd (1983) for English. The orthodox Pierrehumbert analysis is used the in the ToBI system for both English (Silverman et al. 1992; ToBI website) and German (GToBI website), and in much other recent work on English based on Beckman & Pierrehumbert 1986.

In a recent paper, Grice, Ladd and Arvaniti (2000) argue that Pierrehumbert’s original analysis making use of the phrase accent is justified on numerous grounds. Their reasoning is based primarily on the *rising-falling* question contour found in a number of languages in Southeastern Europe, including Greek, Romanian, and Hungarian (Grice et al. refer to this contour as the “Eastern European Question Tune” or EEQT). In the EEQT, the nuclear accented syllable is low in pitch, and the peak of the rise-fall occurs in different places depending on the segmental composition of the post-nuclear section of the utterance. If there is no lexically stressed syllable following the nucleus, the peak goes to a fixed location near the end of the utterance, as one might expect of a boundary-related tone. However, if there is a post-nuclear lexically stressed syllable, the peak is attracted to that syllable, creating an apparent post-nuclear accent[1].

Grice et al. argue that the peak of the rise-fall is the reflex of a phrase accent in the phonology of the intonation contour; that is, the contour consists of a low nucleus followed by a high phrase accent followed by a final low boundary. Phrase accents, in their view, are essentially peripheral or boundary tones that have a "secondary association" (Pierrehumbert & Beckman 1988, Gussenhoven 2000) to a specific post-nuclear anchor point. The rise from low pitch that is often seen on the nuclear word of a question in Greek or Romanian is a transition from one phonological element (the low nuclear accent) to another (the high phrase accent).

This analysis, as Grice et al. note, is relevant to the analysis of falling and falling-rising nuclear contours in the West Germanic languages. That is, if the EEQT is best analysed as a sequence of low nucleus, high phrase accent, and final low boundary, then it is plausible to propose the conceptually identical Pierrehumbert/ToBI analysis for a German or English fall-rise – that is, a *high* (or rising) nucleus followed by a *low* phrase accent followed by a final *high* boundary. Grice et al. specifically discuss this analysis for the QFR in German and English, and present some evidence that the alignment of the putative phrase accent low of the QFR is influenced in the same way as the phrase accent high of the EEQT. However, the phonetic evidence they present is relatively informal, and many of those who are currently working on the intonational phonology of various European languages – in particular the creators of the ToDI system – remain sceptical about the phrase accent analysis for falling and falling-rising nuclear contours. The goal of this study, therefore, is to provide methodologically sound and statistically robust data showing that secondary association of the post-nuclear low in the Dutch QFR is as claimed by Grice et al. – and with that, perhaps, to persuade phrase-accent sceptics that the idea has merit.

Method

We designed a series of test sentences in which we manipulated the relevant linguistic variables (the nature of the post-nuclear syllables) and had native speakers of Dutch read the sentences aloud. We then measured the alignment of the F0 landmarks (nuclear peak and post-nuclear turning points) relative to potential segmental anchors (various segment boundaries) and then tested hypotheses about the factors that influence the alignment.

Speech materials

The test sentences were all questions of the form *Woon je in X?* [✂✌□⏋■ ℯɾ★ ✋■ ✂X] (‘Do you live in X?’), where X was one of a long list of real place names. The place names all had primary lexical stress on the first syllable, and the stressed syllable was followed by either two or three syllables. There were four prosodic patterns in the place names:

- **Sww:** Stressed syllable followed by two syllables with reduced (schwa) vowels, e.g., *Dommelen* [✂♎⚐❍★●★], *Steenderen* [✂⬧⧫♏⏋■♎★❒★][2].
- **Sws:** Stressed syllable followed by a syllable with a reduced vowel followed by a syllable with “secondary stress” (i.e., a full vowel), e.g., *Hengelo* [✂♒☜☠★∩●□⏋], *Oldenzaal* [✂⚐●♎★∩⌘♋⏋●]
- **Ssw:** Stressed syllable followed by a syllable with “secondary stress” (i.e., a full vowel), followed by a syllable with a reduced vowel, e.g., *Eindhoven* [✂☜✋■♎∩♒□⏋❖★], *Steenbergen* [✂⬧⧫♏⏋■∩♌☜❒⌧★].

- **Swsw:** Stressed syllable followed by a syllable with a reduced vowel followed by a syllable with "secondary stress" (i.e., a full vowel) followed by another syllable with a reduced vowel, e.g., *Hindeloopen* [✂♒✋■♎★∩●◻⌉◻★], *Zevenbergen* [✂⌘♏⌉❖★∩♌☜❒☒★].

There were roughly 20 place names for each of the four patterns: 18 for **Sww**, 26 for **Sws**, 19 for **Ssw** and 25 for **Swsw**, for a total of 88 test sentences. As far as possible we chose names with sonorant consonants, which have fewer distorting effects on F0, but we were constrained by wanting to use the names of real places, and many of the names contain obstruents as well. On the basis of the claims by Grice et al. and our own informal observations, we expected to find that the F0 minimum was aligned with the postnuclear full vowel in those cases that contained such a vowel (**Sws**, **Ssw**, and **Swsw**), but near the beginning of the final syllable in the cases with no postnuclear full vowel (**Sww**). A full list of place names is given in the Appendix.

Speakers

The speakers were seven students at the University of Nijmegen in their early to mid twenties. Four were female (SK, SH, ND and MF) and three male (AC, XC and JP). As would be expected from the Nijmegen student population, some of the speakers (especially XC, JP and AC) had identifiably southern Dutch accents, while the rest had general Dutch ("algemeen Nederlands") accents. The speakers were paid a small sum for their participation.

Recording procedures

The test sentences were printed on sheets of paper, ten sentences per A4 page. The test sentences were interspersed with materials for a separate experiment not reported here. The interspersed sentences were short declarative sentences. By including roughly equal numbers of declarative sentences and questions we hoped to prevent the speakers from settling into a monotonous rendition of a list of nearly identical questions.

Speakers were seated comfortably at a table in a sound-treated recording booth in the Nijmegen University Phonetics Laboratory. As suggested earlier, we gave the speakers no explicit instructions about intonation; they were simply asked to read the sentences naturally and to read each sentence individually rather than treating the whole series as a list.

Recordings were made on DAT tape using professional equipment. The digital recordings were transferred to a Sun workstation at the University of Edinburgh for acoustic analysis using ESPS Waves+ software. F0 was extracted, and acoustic measurements were made by the first author on the basis of simultaneous displays of waveform, F0 contour, and wide-band spectrogram (F0 was extracted with the *get_f0* program using its default settings, i.e. a 7.5 ms correlation window and a 10 ms frame shift). Sentences were analysed only if they exhibited the QFR contour, as judged by at least two of the three authors; in all but a handful of cases the decision was completely uncontroversial. One of the speakers (MF) used very few QFRs and her data were not analysed further. The other six speakers all used a substantial proportion of QFRs (between 67 and 86 of the 88 test sentences) and their data are reported here.

Analysis procedures

F0 target labels were all placed by hand on the basis of screen displays. Originally, we intended to label two targets in the extracted F0 contours: the H (F0 maximum on the nuclear accented syllable) and the L (the F0 minimum that is the focus of interest in this paper). However, it quickly became apparent that there was a complication in the definition of the "F0 minimum", which we need to discuss before proceeding. When the post-nuclear stretch is quite short (especially in the **Sww** sentences), it is often possible impressionistically to identify a single F0 minimum, before which the F0 is clearly falling and after which it is clearly rising. In these cases it would be sensible to take the alignment of the single F0 minimum as our measurement point, and this is what we had intended to do. However, when the post-nuclear stretch is somewhat longer (and especially in the **Ssw** and **Swsw** sentences), there are many cases where there is not a single obvious F0 minimum but a stretch of low F0, often slightly declining but approximately level, whose edges are roughly defined by "elbows" or "corners" in the pitch contour. Impressionistically, the F0 in these cases falls from the nuclear peak and then *levels out,* before rising again at the end of the phrase. (This is seen in the idealised contour shown for H* L H% in Figure 1 above.)

This is not *a priori* a conceptual problem for the phrase-accent analysis; something analogous is true of some varieties of the Eastern European Question Tune as well. Specifically, Grice et al (2000) note that in the Transylvanian varieties of both Hungarian and Romanian, the EEQT contour shows an F0 rise immediately following the nuclear F0 minimum, then levels out to form a plateau before falling again at the end of the phrase. The low-level stretch in the Dutch QFR is completely comparable to the *high*-level stretch in the EEQT. By analogy to the geographical metaphor of the F0 "plateau", we refer to the low-level stretch as the F0 "flood plain" throughout the rest of the paper.

However, though it may not pose a conceptual problem, the existence of flood plains in these contours certainly does pose a methodological problem: in such cases it may not be very meaningful to identify a single F0 minimum, because that value would often reflect nothing more than random pitch fluctuations in an essentially level stretch of contour. Instead, it may be more appropriate to identify *two* low F0 turning points rather than just one – one at the beginning of the flood plain (i.e., the end of the accentual fall) and one at the end of the flood plain (i.e., the beginning of the utterance-final rise). But this poses a different methodological problem: unlike local maxima or minima (which can be identified more or less objectively), turning points are most easily identified by eye – hence less objectively – on a screen display of the contour. Moreover, it would obviously be inappropriate to adopt one measurement procedure for the cases with a single clear minimum and another for those with a clear flood plain, as that would introduce a further level of subjective judgement.

We dealt with these issues by doing the entire analysis both ways: once using the absolute F0 minimum throughout, and once marking the beginning and end of impressionistically observed flood plains. (Note that, even in cases where the single F0 minimum seemed meaningful, it was generally easy to identify points between which the slope of the F0 seems to level out at least briefly between the end of the fall and the beginning of the rise.) We refer to the turning points at the beginning and end of the flood plain as L1 and L2 respectively. We continue to refer to the single F0 minimum as L. Figure 2 shows the display of two typical utterances, together with our labels for the putative intonational targets. In general we identified H and L as the absolute F0 maximum and minimum in the relevant section of the contour, ignoring (as far as possible) any obstruent-induced perturbations and obviously spurious values due to irregular voicing. The turning points corresponding to L1 and L2 were identified on screen displays, by the first author in consultation with the third. In several cases either L1 or L2 coincided with L; in a few cases where it was really impossible to see a flood plain, both L1 and L2 were marked at the same location as L.

We did not perform any reliability check on the location of L1 and L2, for three reasons: first, because we found it was very easy to agree on most of the cases we looked at together; second, because Arvaniti (personal communication) has found that less impressionistic methods of locating turning points – such as modelling and line-fitting – yield the same conclusions as hand labelling; third and most importantly, because we could use the results of the single-minimum method as a check on the conclusions of the flood plain method, and vice versa. In what follows we report the results of both analyses together.

INSERT FIGURE 2 ABOUT HERE

Results

Results are presented graphically in Figure 3. The figure shows the mean alignment of the post-nuclear F0 turning points (L1 and L2) and of the absolute post-nuclear F0 minimum (L) relative to the syllable boundaries of the test words. It can be seen informally that the results for L conform roughly to the predictions based on Grice et al.: in the **Sww** case the L aligns near the beginning of the final syllable, whereas in the other three cases it aligns in the post-nuclear full vowel, i.e., at a point which is either before (**Ssw** and **Swsw**) or after (**Sws**) the beginning of the final syllable. It can also be seen that the alignment of L1 and L2 – the beginning and end of the flood plain – is affected in a similar way: L1 appears to be aligned at the beginning of the post-nuclear full-vowel syllable (except in **Sww**, where there is no such syllable), and L2 is aligned at the beginning of the final syllable except in the **Sws** cases, in which case it is aligned later. (The figure also shows the alignment of the F0 peak on the nuclear accent (H); it can be seen that there are at most small differences in H alignment between conditions.)

INSERT FIGURE 3 ABOUT HERE

Statistical analysis of the alignment of the low F0 points

In order to put these informal results on a more solid statistical basis, we report the results of a Multivariate Analysis of Variance (MANOVA) with six univariate analyses in which significant effects of our experimental manipulation (henceforth STRESSPATTERN) on the alignment of L (or of L1 or L2) are found. The univariate analyses were two-way ANOVAs with STRESSPATTERN and SPEAKER as fixed between-items factors. Although SPEAKER is traditionally a sampling or random factor, in small N experiments such as are typical in phonetics, there are often (as in our case) too few speakers to treat it as a sampling factor. Our decision to use SPEAKER as a between-items factor, despite the fact that it was a within-items factor in the design, was based on the pattern of data loss. We have used and justified both these approaches elsewhere (e.g., Ladd and Schepman, 2003).

In these three pairs of ANOVAs, we consider the alignment of a low F0 point relative to three different possible reference points, each of which allows us to reject a plausible null hypothesis about how alignment works. In the first pair of analyses, the alignment of L (or L1) is expressed relative to the preceding H (the accent peak), thereby testing the null hypothesis that the duration of the accentual F0 fall is invariant. In the second analyses, alignment of L (or L2) is expressed relative to the beginning of the final vowel, thereby testing the null hypothesis that the final F0 rise is anchored to the final syllable or final rhyme. In the third analyses, alignment of L (or L2) is expressed relative to the end of the utterance, thus testing the null hypothesis that the duration of the final F0 rise is invariant. In all three cases we are able to reject the null hypothesis, and our results are consistent with the predictions based on Grice et al. (2000).

Within the MANOVA, the analyses were two-way (4x6) ANOVAs with STRESSPATTERN and SPEAKER as between-items factors. In the following sections, however, we report only the main effects of STRESSPATTERN. There were both main effects for SPEAKER and various interactions between SPEAKER and STRESSPATTERN, but the effects and interactions were small, and the overall pattern of alignment was substantially the same for all speakers. As we are not concerned with individual differences in this paper, we felt that it made sense to streamline the reporting of results by dealing only with the effects of STRESSPATTERN.[4]

L (or L1) alignment relative to nuclear H peak

In this analysis we studied the alignment of L (or L1) relative to the preceding accentual H. If the accentual fall is conceived of as a phonological unit H+L (as it is in the ToDI analysis, for example), it is reasonable to think that this unit might have a constant duration, i.e., that its duration is governed by the length of time it takes F0 to drop by an appropriate interval. Such invariant duration for accentual falls was in fact reported by Caspers and van Heuven 1993. This is the null hypothesis in this analysis; the experimental hypothesis is that STRESSPATTERN, by affecting the alignment of the L (or L1), will have an effect on the duration of the fall.

Results are presented in Table 1.

INSERT TABLE 1 ABOUT HERE

With both measures of the alignment of the F0 minimum, the ANOVA shows a highly significant effect of STRESSPATTERN (for H-to-L, $F(3,454) = 84.8$, $p < .001$; for H-to-L1, $F(3,454) = 16.798$, $p < .001$). For H-to-L, post hoc Bonferroni t-tests showed that all the four conditions were significantly different from all the others except for the pair **Sws-Ssw**; for H-to-L1, post hoc Bonferroni t-tests show that **Swsw** was significantly different from the other three cases. We may thus reject the null hypothesis of a fixed fall time, and accept the hypothesis that STRESSPATTERN does have an effect on the alignment of the F0 minimum.

L (or L2) alignment relative to final vowel

In this analysis we studied the alignment of the F0 minimum relative to the onset of the final vowel. If the final rise from the F0 minimum or the end of the flood plain to the end of the utterance is taken as a property of the sonorant part of the final syllable, it is a reasonable hypothesis that the final rise might begin at the beginning of the final syllable rhyme, i.e., the onset of the final vowel. This is the null hypothesis in this second analysis; the experimental hypothesis is that STRESSPATTERN will have an effect on the alignment of L (or L2) measured relative to the beginning of the final vowel.

Results are shown in Table 2 (a negative number means that the F0 minimum was aligned before the onset of the final vowel):

INSERT TABLE 2 ABOUT HERE

These results are what we would expect from the phrase accent hypothesis: in the **Sww** case, L and (to a lesser extent) L2 are aligned with the beginning of the final *syllable*, and hence a short distance before the beginning of the final vowel; in the **Sws** case, L and (to an even greater extent) L2 are aligned *during* the final vowel, and therefore after its onset; in **Ssw** and **Swsw** the L is aligned during the penultimate vowel, and L2 is aligned at the beginning of the final syllable, both a considerable distance before the onset of the final vowel.

The ANOVA showed a highly significant effect of STRESSPATTERN (for L-to-vowel onset, $F(3,454) = 345.82$, $p < .001$; for L2-to-vowel onset, $F(3,454) = 259.74$ $p < .001$). For both definitions of the relevant F0 minimum, Bonferroni t-tests showed that all the conditions were significantly different from the others, except for the pair **Swsw-Ssw**. We may thus reject the null hypothesis of a fixed alignment at the final vowel onset, and once again accept the hypothesis that STRESSPATTERN does have an effect on the alignment of the F0 minimum.

L (or L2) alignment relative to the end of the utterance

In this analysis we studied the alignment of the F0 minimum relative to the end of the utterance. If the utterance-final rise is taken as a unit, it is a reasonable hypothesis that it might be of roughly constant duration, governed by the time it takes F0 to rise by the appropriate amount. This is the null hypothesis in this third analysis; the experimental hypothesis is that STRESSPATTERN, by affecting the alignment of the L (or L2), will have an effect on the duration of the final rise. Note, however, that we defined the duration of the final rise in terms of the end of the acoustic signal for the utterance, rather than on the last or the highest F0 point, since F0 extraction at low energy levels is not very reliable.

Results of these analyses are shown in Table 3. The exclusively negative numbers indicate simply that the F0 minimum was aligned before the end of the utterance.

INSERT TABLE 3 ABOUT HERE

The ANOVA showed a highly significant effect of STRESSPATTERN (for L-to-end, $F(3,454) = 52.59$, $p < .001$; for L2-to-end, $F(3,454) = 16.80$, $p < .01$). For L, the post hoc test showed that only the pattern for **Sww** was different from the three patterns involving a post-nuclear full vowel; for L2, the post hoc test showed that all the conditions were significantly different from the others, except for the pair **Swsw-Ssw.** We may thus reject the null hypothesis of fixed alignment relative to the end of the utterance, and once again accept the hypothesis that STRESSPATTERN affects the alignment of the minimum.

Discussion

The findings just reported show clearly that the alignment of post-nuclear F0 minima in Dutch falling-rising question contours is conditioned by the phonological structure of the post-nuclear material, specifically the location of any post-nuclear secondary stressed syllable. This is true regardless of the precise operational definition of the "post-nuclear F0 minimum", though of course the definition affects the details. For two different definitions of the relevant minima, we were able to reject both null hypotheses based on the assumption that the duration of the fall or the rise is invariant, and a null hypothesis that the duration of the rise is linked to the duration of the sonorant part of the final syllable.

The fact that the alignment of the post-nuclear minima depends on the phonological structure of the post-nuclear material is, in general terms, yet another demonstration that the alignment of F0 targets with the segmental string is highly lawful. As such it is consistent with what has been found about F0/segmental alignment in the studies discussed in the introduction. More specifically, our findings are consistent with the idea, proposed by Grice et al. 2000, that the low F0 between the fall and rise – whether it is a single minimum or what we have called a flood plain – is the reflex of a "phrase accent" that seeks a secondary association with a secondary stressed syllable if one is present following the primary stress. The attraction of the F0 minimum to the secondary stressed syllable is seen most clearly when we compare the **Sws** cases (where the post-nuclear secondary stressed syllable is utterance-final) to the **Ssw** and **Swsw** cases (where it is utterance-penultimate). In the latter cases, there is a F0 flood plain that spans most of the secondary-stressed syllable, while the final rise roughly spans the final (unstressed) syllable; in the former cases there is at most a short flood plain, which extends into the secondary-stressed syllable, and the final rise begins some distance into the final (secondary-stressed) vowel.

At first glance it might appear simpler to account for the regularities just summarised with reference to the location of the accentual fall and the final rise. That is, we might hypothesise that the accentual fall and the final rise are of invariant duration, and that the location and duration of the flood plain depends simply on how much time there is between the end of the fall and the beginning of the rise. But as we saw, there is no invariance about the duration of either fall or rise. Instead, it appears that the end of the fall and the beginning of the rise are "trying" to align with specific points in structure: not only does this determine that duration of the flood plain, but it causes the duration of the fall and the rise to adjust as well. This is especially noticeable in the **Swsw** cases, where the duration of the fall increases to accommodate the extra distance to the beginning of the post-nuclear secondary-stressed syllable.

To be sure, there are several matters of detail in our results for which we have no exact account at present. For example, our findings are almost certainly influenced by factors of "time pressure" of the sort investigated by Caspers and van Heuven 1993, especially in the **Sww** and **Sws** cases. A particular puzzle that we might mention here is the fact that the final rise is *longest* in **Sws**, rather than shortest, which is what we might expect if the contour is trying to accommodate a flood plain low and a final rise on the same post-nuclear secondary stressed syllable[5]. None of this, however, undermines the compatibility of our findings with the proposal of Grice et al. (2000). We therefore conclude that our study contributes significantly to the growing body of evidence that the proper analysis of falling and falling-rising nuclear contours involves a "phrase accent", as originally suggested by Pierrehumbert 1980.

STUDY II – "LAB SPEECH" IS REAL SPEECH

While Study I may contribute to a debate in intonational phonology, it does nothing to address the unease of those who are unconvinced by the whole idea that intonation can be appropriately investigated on the basis of "lab speech". It is sometimes suggested that what happens when someone reads sentences aloud in a recording booth does not allow us to draw conclusions about how intonation works in more natural settings. To some extent this view is uncontroversial: it is perfectly obvious that a speaker's choice of intonation contour is influenced by all kinds of factors that are difficult to emulate in the lab. More generally, there are obviously differences between read speech and spontaneous speech, though it is worth noting that it does not appear easy to characterise what makes read speech sound like read speech (see e.g., Blaauw 1995, Laan 1997). Nevertheless, "laboratory phonology" studies of intonation – like Study I – implicitly assume that whatever the differences between read speech and spontaneous conversational speech, they need not undermine conclusions about the phonological categories and structures of intonation and the way they are manifested phonetically.

Yet some writers have more fundamental doubts:

- "In reading texts the speaker is merely articulating structures which have been pre-prepared. His intonation is 'post-syntactic' and does not arise from the sorts of constraints which apply when speech is produced spontaneously. … A reader-aloud first has to assign an interpretation to the text and then to utter it in a way consistent with his interpretation. This is a very different task from the normal processes of speech production in non-goal-directed speech where the speaker has to organise what he wants to say as he is speaking. It must be clear that claims made about intonation on the basis of the study of texts read aloud, should be subject to the most careful scrutiny if there is any suggestion that these will correctly characterise the intonation of spontaneous speech." (Brown, Currie and Kenworthy 1980: 17f)

- "The linguistic study of prosody has recently been reinvigorated by new interest in prosodic and intonational phonology within the generative paradigm. Most research in linguistics proper, however, has worked on the basis of introspective constructed data and has been concerned with the grammatical function of prosody. The categories and methodologies used in this research have been devised to fit this type of data with a grammatical aim in mind........ In sum, despite the long tradition of prosodic research, *its categories and methodologies are inappropriate for handling conversational data*." (Couper-Kuhlen & Selting 1996: 2)

We can readily concede one point: if one's aim is to study conversation, then clearly conversational data, not read speech, are necessary. However, if one is interested in the phonetic realisation of phonological categories and structures, then the source of one's primary data may not be critical. In this connection, Brown et al.'s claim that reading aloud is very different from normal speech production needs closer examination. Current theories of the cognitive processes involved in speech production (e.g., Levelt 1989) do support the idea that the formulation and syntactic planning stages of spontaneous speaking are different from those involved with reading text aloud. However, such theories also involve "lower" levels at which the planned utterance is translated into a phonological/phonetic code, and here the differences between read and spontaneous speech may disappear: the two planning routes may converge to feed into a shared articulation mechanism. Nothing in the presumed cognitive architecture of speech production forces us to expect that the higher-level differences will lead to observable differences at lower levels, especially where matters of fine phonetic detail are concerned. That is, once a speaker has chosen a contour, it is a reasonable assumption that the contour's phonetic properties are largely or wholly predictable from phonetic and phonological factors alone.

Moreover, this is potentially an empirical question, and we wish to take seriously Brown et al.'s suggestion that "the relationship between the intonation of texts read aloud and spontaneous speech needs to be carefully investigated". Almost by definition, we cannot provide a direct test of the assumption that read speech is an appropriate basis for studying intonational phonology and phonetics; nor can we statistically test the hypothesis that there is no difference between read and spontaneous speech. However, we can evaluate the plausibility of the assumption by making rigorous comparisons between "lab speech" data and data from similar materials produced in dialogue. This is what we have done in Study II. While it may be argued that the Map Task dialogues that we use in this study are not entirely spontaneous and do not constitute completely spontaneous data in the sense of Couper-Kuhlen and Selting (1996), the speech produced in Map Task dialogues is clearly not read speech. In the question utterances that we analyse, the questioner is typically requesting information that he or she requires in order to complete part of the task. In addition, the intonation could not be seen as being 'post syntactic' in the sense of Brown, Curry and Kenworthy (1980: 17f).

Method

Our approach was twofold. First, we took all the naturally occurring QFRs in a corpus of conversational Dutch and measured the location of the F0 minima in each of them, to see whether their alignment is consistent with the findings based on "lab speech" presented in Study I. Second, we investigated the various factors – phonological, syntactic and pragmatic – that favour the use of falling-rising questions in dialogue, to see whether in our "lab speech" materials we have effectively matched those conditions. The corpus of conversational Dutch was based on the "Map Task".

The Map Task

The Map Task (Anderson et al. 1991) is a widely used tool for eliciting dialogue while still allowing the investigator some degree of control over the content and structure of what is said. The Map Task works as follows: the two participants to the conversation each have a map showing a variety of named landmarks. Neither speaker can see the other's map. One map (the "Instruction Giver's" map) has a route marked on it, and the task is for the Instruction Giver to explain to the Instruction Follower where the route passes, referring to the various landmarks along the way – accurately enough that the Instruction Follower can reproduce the route on his or her own map. Importantly, the maps may differ slightly in detail. Landmarks present on one map are sometimes missing on the other map, or the same landmark may have two different names on the two maps. This means that the participants frequently need to discuss specific landmarks and check on the possibility of differences between the two maps.

We recorded the Map Task corpus for a variety of purposes as part of the larger research project of which the QFR study forms a part. In the present context, the obvious advantage of the Map Task is that it allowed us to obtain plenty of question intonation data from dialogue speech[6], because in negotiating the route both the participants to the dialogue frequently ask questions (e.g., *Have you got [landmark]?* or *Over to the right?*). Unlike the case of sentences punctuated with question marks and read aloud in the recording studio, we can be reasonably confident of the communicative force of most questions produced by speakers performing the Map Task.

The Dutch Map Task Corpus

The Dutch Map Task corpus from which our natural QFR utterances were drawn was recorded at the Phonetics Laboratory of the University of Nijmegen on the same day as a number of other experimental recordings, including those used in Study I. The speakers were all students at the University and all native speakers of Dutch. They were paid a small sum for their participation. The speakers included three of the speakers recorded in Study I (AC, XC, SK), plus one other (DB). DB had a general Dutch accent.

The maps used were the same basic maps as in the original HCRC Map Task corpus (Anderson et al. 1991), but the landmark names, and most of the landmarks, were changed. We selected landmark names that manifested the phonological structures we were interested in, and that contained consonant types which would permit easy analysis of pitch patterns (e.g., *Vlaamse nonnen* 'Flemish nuns' or *donkere wolvenhol* 'dark wolves' den'). We used the same basic design of the HCRC map task corpus by recording a "quad" of speakers (8 separate dialogues). In a "quad" design each speaker interacts with 3 other speakers, twice as "Giver" (for the same map) and twice as "Follower" (with two different maps).

The overall duration of the 8 conversations is approximately 42 minutes. The entire corpus was orthographically transcribed by Angela Vonk, a native speaker of Dutch and a student at the University of Nijmegen. Fuller details, as well as the corpus itself, are available from www.data-archive.ac.uk; see also several papers by Johanneke Caspers (2000a, 2000b, 2001), which are based on our corpus.

Selection of utterances for analysis

The corpus contains approximately 123 yes-no questions. We say "approximately" because it is occasionally difficult to determine whether an utterance is intended as a question or an instruction, and because it is occasionally difficult to determine whether a stretch of speech within one speaker's turn consists of one question or two. Once we had reached a decision on these indeterminate cases, the 123 questions were labelled as having QFR, rise, or fall intonation on their nuclear accent. (Recall that our category "rise" comprises the three ToDI categories L* H%, L*+H H%, and H* H%.) All three authors listened to all the questions and reached a consensus on the intonation; in most cases this was uncontroversial. We labelled 54 cases as QFR, 55 as rise, and 14 as fall.

In selecting these questions for analysis, we systematically excluded alternative questions (e.g., "left or right?") and WH-questions, even though in some cases these had clear QFR contours. We also excluded two fairly well-defined groups that might be considered yes-no questions: (1) single-word utterances that are simply requests for confirmation that the addressee has understood (these mostly took the form *Ja?* 'yes?' and *Oke?* 'okay?'); (2) utterances that were essentially statements but which ended with a questioning tag (normally *hè?* 'eh?', but also including a few cases with short tag phrases like *bedoel je?* 'do you mean?'). In addition, we had to exclude a few utterance that were clearly intended as yes-no questions but were sufficiently affected by overlapping speech, disfluency, etc., that it was impossible to assign them to one of the three intonational categories. Finally, we note that a few of the 123 questions could not be measured acoustically because of overlapping speech or irregular voicing.

Analysis

We performed two types of analyses. First, we measured the alignment of the F0 minima in the QFR sentences to see whether they are affected by the presence of post-nuclear stressed syllables in the same way as in Study I. Second, we identified three statistically significant factors – one pragmatic, one syntactic, and one phonological – that appear to condition the choice between falling-rising and rising intonation in questions produced in dialogue. The second of these analyses provides information of the sort that is most easily derived from corpus data and would probably be difficult to obtain through experiments based on read speech, yet they can be seen as confirming the laboratory phonology assumption that it is often possible to elicit specific intonation patterns for phonetic study using sensitively constructed materials in a reading task.

Results

In this section we first present the data for the alignment of the F0 minima in QFR sentences from the corpus, directly comparing read speech and dialogue speech with respect to the alignment variables under study. We then present the results of the analysis of pragmatic, syntactic and phonological conditions on the choice between QFR and rise.

Alignment of F0 minimum in QFR produced in dialogue

There were 21 usable cases directly comparable to those in Study I: **Sww** (2 cases)**, Sws** (10 cases)**, Ssw** (4 cases)**,** and **Swsw** (5 cases). Examples include the following:

Heb je beneden de grazende runderen? (Speaker DB, Conversation 3, 150.6s)

have you below the grazing cattle ‘Do you have the grazing cattle at the bottom?’

Accented word *runderen* [✂r✡nd★r★] = **Sww**.

Gaan we richting dennenbos? (Speaker XC, Conversation 6, 122.4s)

go we direction spruce forest ‘Are we going towards the spruce forest?’

Accented word *dennenbos* [✂♎☜■★∩♌⚐♦] = **Sws**

Heb je een oude vuurtoren? (Speaker AC, Conversation 4, 342.9s)

have you an old fire tower ‘Do you have an old lighthouse?’

Accented word *vuurtoren* [✂v⌧⌉❒∩⧫☐⌉❒★] = **Ssw**

We marked H, L, L1 and L2 in these cases in exactly same way as in Study I. Unfortunately there are too few cases to permit full statistical evaluation of the differences in the mean alignment data for the four different subcases, but it can be seen impressionistically from inspection of Figure 4 that the alignment patterns found in the dialogue speech data are strikingly similar to those observed in Study I. (Figure 4 should be compared to Figure 3 above.) Moreover, a few limited statistical comparisons are possible, which strongly suggest that the dialogue data show patterns of alignment that are indistinguishable from the "lab speech".

First, we compared the lab speech data with the dialogue data for the one case where we had enough dialogue cases, namely **Sws**. We performed one-way ANOVAs comparing the dialogue cases (n = 10) with the read cases (n = 138) for three of the dependent variables used in Study I, namely the alignment of the F0 minimum relative to the accentual high (H-to-L1), relative to the beginning of the final vowel (L2-to-final-vowel), and relative to the end of the utterance (L2-to-end). All failed to find any difference between dialogue and lab speech ($F < 1$ in all cases). Obviously, one should be wary of overinterpreting negative results, but at the very least this analysis gives us no reason to doubt the conclusions of Study I.

Second, within the dialogue speech data, we compared the alignment of the F0 minimum in two further sets of QFR cases not represented in Study I, namely cases in which the nuclear syllable followed by only one syllable, either a full-vowel syllable (**Ss**) or a reduced-vowel syllable (**Sw**). We had 7 **Ss** cases (e.g., *wigwam* 'wigwam', *spoorweg* 'railway') and 11 **Sw** cases (e.g., *vijver* 'pond', *ruïne* 'ruins'). The predictions of the phrase-accent analysis are that the low phrase accent should be secondarily associated with the post-nuclear full-vowel syllable in the **Ss** cases and that the F0 minimum should therefore be aligned with the final vowel, whereas in the **Sw** cases the phrase accent has no secondary association and the F0 minimum should be aligned about the beginning of the final syllable. Obviously, with short words in fast speech it may be that time pressure affects the realisation of the falling-rising pitch movement in ways that we cannot predict. Nevertheless, we should at least expect to find significant *differences* in alignment between the two groups of cases.

To test this prediction, we did one-way ANOVAs comparing **Ss** and **Sw** on the same three dependent variables from Study I (H-to-L1, L2-to-final-vowel, and L2-to-end). There was a highly significant difference for the duration of the accentual fall (H-to-L1) (**Ss** mean 139 ms; **Sw** mean 81 ms; $F(1,14) = 24.22$, $p < .001$), suggesting that the F0 minimum is later in **Ss** than in **Sw** . The difference for the alignment relative to the beginning of the final vowel (L2-to-final-vowel) approached significance (**Ss** mean 33 ms; **Sw** mean 4 ms; $F(1,14) = 3.625$, $.05 < p < .10$), again suggesting that the F0 minimum is later in **Ss** than in **Sw**. There was no difference ($F < 1$) in the duration of the final rise (cf. footnote 3). These results, limited though they are, once again give us no reason to doubt or modify the conclusions based on "lab speech" in Study I.

INSERT FIGURE 4 ABOUT HERE

Factors conditioning the choice between QFR and rise

We noted earlier that roughly equal proportions of the questions in the Map Task corpus have QFR and rise. Specifically, we found 44% with QFR, 45% with rise, and 11% with fall. Leaving aside the small number of cases with fall[7], we make some corpus-based observations here on the choice between QFR and rise. We identify three factors that may be at work in this choice, one pragmatic, one syntactic, and one phonological.

Pragmatic factor: We noted in our introductory comments on Dutch question intonation that the QFR may be especially appropriate in questions "at the opening of a transaction or the beginning of a new topic". This impressionistic statement is consistent with our corpus data. The QFR is overwhelmingly used with "first mentions", i.e., the first time a map landmark is named in the dialogue. Roughly two-thirds of the occurrences of QFR in the corpus (35 out of 54 cases, or 65%) were on questions of the general form *Do you have [landmark name]?* By contrast, only 13 of the 55 questions (24%) with rise contours were used for such questions. This difference is statistically significant ($\chi^2 = 18.748$, df = 1, $p < .01$). Rises tended to be used for other kinds of questions that did not involve first mentions of landmark names, including quite a number in which the instruction follower asks for confirmation that they have understood correctly:

Ja, helemaal aan de rechter kant van het ***blaadje****?* (Speaker AC, Conversation 5, 170.0s)
yes, entirely on the right side of the sheet(diminutive) ‘All the way to the right of the paper?’

Drie centimeter naar ***beneden****?*(Speaker XC, Conversation 6, 25.9s)
three centimetres to below ‘Three centimetres down?’

That is, in the Map Task corpus there seems to be a preference for QFR in questions involving new topics or new information and a preference for rise in questions that check or confirm information already in the discourse.

Syntactic factor: Another difference between the two broad intonational types of question may be inferred from the examples we have just given. QFR is overwhelmingly used with short sentences having question syntax (i.e., with the finite verb preceding the subject), whereas rises are found predominantly with verbless utterances consisting of a noun phrase or an adverbial phrase, and with sentences having statement syntax (Haan’s “declarative questions”). Of the 54 QFR cases, 42 (78%) had question syntax but only 21 (38%) of the 55 rise cases did. This difference is statistically significant ($\chi^2 = 17.514$, df = 1, $p < .01$).

Phonological factor: It has been claimed that in German, which like Dutch has both a QFR and a rise question contour, the QFR tends to be avoided on utterance-final accented syllables and replaced with a rise (Féry 1993: 91, Ladd 1996: 133f). Our Map Task data suggest that such a phonological factor is at work in Dutch as well. We counted all the occurrences of QFR and rise in questions where the nuclear accent was on the final, penultimate, or antepenultimate syllable. The results, shown in Table 4, make it clear that the distribution of the two contour types is affected by the position of the nucleus ($\chi^2 = 18.36$, df = 2, $p < .01$) It seems likely that Dutch speakers avoid using the QFR on final accented syllables, as Féry and Ladd suggest for German.

INSERT TABLE 4 ABOUT HERE

It may be worth noting in this connection that 7 of the 14 questions with fall contours had final accented syllables. This may mean that, in questions with final stress that "should have" had a QFR on pragmatic and/or semantic grounds, we frequently find a fall instead. This speculation could be tested in a Map Task corpus in which the stress placement in landmark names is systematically manipulated. Most of our landmark names had penultimate or antepenultimate stress, so we have no basis for discussing this further.

Discussion

The comparison of the alignment of the F0 minimum the dialogue speech from the Map Task with the results of Study I gives us no reason to think that the read speech data from Study I are phonetically unrepresentative of what happens in natural interaction. On the contrary, it lends plausibility to the assumption that read speech can be used as a source of evidence in experimental work that addresses phonological and phonetic questions. Indeed, it suggests that it makes practical sense to study the phonetic and phonological issues on the basis of controlled speech materials, rather than recording natural conversations and hoping for the occurrence of appropriate utterances.

This methodological conclusion is strengthened by our corpus-based analysis of the factors affecting the choice between falling-rising and rising questions. We identified three factors favouring the choice of QFR, namely pragmatic newness, question syntax, and non-utterance-final stress. All three of these conditions are met by our *Do you live in X?* sentences. This may indicate that, in reading the sentences, most of our speakers successfully set themselves in a context where the QFR would be the appropriate intonation. The idea that speakers who read aloud do so with a context in mind is sometimes taken as *prima facie* evidence against the validity of read speech data. Brown et al., in the passage cited above, patently take it to be a shortcoming of read speech studies that "a reader-aloud first has to assign an interpretation to the text and then to utter it in a way consistent with his interpretation". Our results put this in a different light: we suggest that readers-aloud do exactly as Brown et al. say *and that this is valuable knowledge for those who would use read speech as an investigative tool.*

This finding is also relevant to the issue of "introspective constructed data" raised by Selting and Couper-Kuhlen in the passage quoted above. These authors are not alone in deprecating the practice of "inventing data" that is said to characterise linguistic work not based on the study of corpora. However, it is important to draw a distinction between "inventing data" – which some linguists may come unacceptably close to doing – and "constructing materials". The sentences we made up for speakers to read in Study I were not our data; they were a means to obtaining our data. If we keep this distinction clearly in mind, and if we suspect that readers-aloud may put themselves in context, then "introspection" actually simplifies the task of investigating phonological and phonetic questions. The careful experimenter with a sensitive ear for the way intonation is used can actually put linguistic intuitions to good use, to obtain contour types for study as efficiently as possible.

CONCLUSION

As we said in the introduction, our broader aim in this paper is to defend the use of "lab speech" – experimental data based on controlled speech materials read aloud in the laboratory – in the study of intonation. We have shown that conclusions based on lab speech, at least insofar as they concern the phonetic realisation of phonological categories, can generalise to natural spontaneous speech. This means that, if one's main aim is to study the way phonological structure affects the phonetic detail of intonation, there is no special merit in preferring naturalistic corpus data over experimental data. On the contrary, the latter have the advantage of being precisely focused on one's hypotheses and may therefore be regarded as methodologically more appropriate.

It should be clear that we are not rejecting the use of corpus data. Indeed, our findings on pragmatic, syntactic and phonological conditions affecting the choice between QFR and rise depend for their validity on the fact that they come from dialogue speech. Our point is simply that there is no justification for assuming, as some seem to do, that spontaneous or task-oriented dialogue speech data are intrinsically superior to controlled speech materials and read speech. Ultimately what is important is to obtain data that bear on one's empirical questions. If one is interested in how different question intonations are used in different contexts, there are clear advantages to the use of corpus data. If, on the other hand, one is interested in how a specific contour type is realised phonetically, then our work suggests that there is no principled objection to the efficient gathering of data in a carefully designed reading task.

Footnotes

1. The description just given applies in detail only to Greek and Romanian. In Hungarian there are minor differences that are not relevant to our point here.

2. Orthographic *n* in the coda of an unstressed syllable is frequently omitted in ordinary Dutch pronunciation, and this is the basis for our transcriptions here.

3. Whereas data analysis was performed using ESPS Waves+ Software, the software used in these images is Wavesurfer (Sjölander & Beskow, 2000)

4. In the MANOVA, significant main effects of STRESSPATTERN and SPEAKER are found, as well as an interaction. However, the effect size for STRESSPATTERN is far larger than for SPEAKER and the interaction [STRESSPATTERN: Wilks' Lambda = .050, $F = 154.59$, $p < .001$, ${}_p\eta^2 = .631$; SPEAKER: : Wilks' Lambda = .567, $F = 10.56$, $p < .001$, ${}_p\eta^2 = .107$; STRESSPATTERN X SPEAKER: Wilks' Lambda = .548, $F = 3.677$, $p < .001$, ${}_p\eta^2 = .113$], and this pattern is repeated throughout the univariate ANOVAS that we report.

5. Part of the explanation for this apparent anomaly may lie in the segmental composition of the final syllable. In our materials many of the final syllables in **Sws** have both a phonologically long vowel and one or more coda consonants (e.g., *-wijk, -doorn, -laan, -hout*) whereas in **Sww**, **Ssw** and **Swsw** the final syllables by definition have a reduced vowel and most of them have no coda. It would require a separate experiment based only on cases like our **Sws** cases, systematically manipulating phonological vowel length and presence or absence of coda, to determine how the makeup of the final syllable affects the duration of the final rise.

6. On the advice of one of the referees, we refer to 'dialogue speech', rather than 'spontaneous speech', which is the term we had originally used.

7. Of the 14 cases of falling questions, only two had question syntax, and fully half (exactly 7 cases) were complete sentences with statement syntax. All 14 were marked as questions by the transcriber and confirmed as such by the authors. One might of course object that a "question" with statement syntax and falling intonation makes any definition of question problematical, which is part of the reason we leave these cases out of further consideration.

References

ANDERSON, A. BADER, M., BARD, E. G., BOYLE, E., DOHERTY, G. M., GARROD, S., ISARD, S., KOWTKO, J., MCALLISTER, J., MILLER, J., SOTILLO, C., THOMPSON, H. S. & WEINERT, R. (1991). The HCRC Map Task Corpus. *Language and Speech* **34**, 351-366.

ARVANITI, A., LADD, D. R. & MENNEN, I. (1998). Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics* **26**, 3-25.

ATTERER, M. & LADD, D. R. (2004). On the phonetics and phonology of "segmental anchoring" of F0: evidence from German. *Journal of Phonetics* **32**, 177-197.

BECKMAN, M. E. & PIERREHUMBERT, J. B. (1986). Intonational Structure in English and Japanese. *Phonology Yearbook* **3**, 255-310.

BLAAUW, E. (1995). *On the perceptual classification of spontaneous and read speech*. PhD thesis, Onderzoeksinstituut voor Taal en Spraak (OTS), Utrecht University.

BROWN, G. CURRIE, K. & KENWORTHY, J. (1980). *Questions of intonation*. London: Croom Helm.

CASPERS, J. (2000a). Melodic characteristics of backchannels in Dutch Map Task dialogues. *Proceedings of ICLSP*, Beijing (no page numbers).

CASPERS, J. (2000b). Pitch accents, boundary tones and turn-taking in Dutch Map Task dialogues. *Proceedings of ICLSP*, Beijing (no page numbers).

CASPERS, J. (2001). Testing the perceptual relevance of syntactic completion and melodic configuration for turn-taking in Dutch. *Proceedings of Eurospeech*, Scandinavia, vol. 2, pp 1395-1398.

CASPERS, J. & VAN HEUVEN, V. J. (1993). Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall. *Phonetica* **50**, 161-171.

COUPER-KUHLEN, E. & SELTING, M. (1996). *Prosody in Conversation*. Cambridge University Press.

D'IMPERIO, M. (2000). *The role of perception in defining tonal targets and their alignment.* PhD Thesis, Ohio State University.

D'IMPERIO, M. (2001). Focus and tonal structure in Neapolitan Italian. *Speech Communication* 33, 339-356.

FÉRY, C. (1993). *German Intonational Patterns.* Tübingen: Niemeyer.

GRABE, E. (1998). *Comparative intonational phonology: English and German.* Ph.D. thesis, Max Planck Institute for Psycholinguistics, Nijmegen.

GRICE, M., LADD, D. R. & ARVANITI, A. (2000). On the place of phrase accents in intonational phonology. *Phonology* **17**, 143-185.

GToBI WEBSITE: http://www.coli.uni-sb.de/phonetik/projects/Tobi/gtobi.php3 (Accessed 11 February 2005)

GUSSENHOVEN, C. (2000). The boundary tones are coming: on the nonperipheral realization of boundary tones. In M. Broe and J. Pierrehumbert (eds.), *Papers in Laboratory Phonology V* (Cambridge: Cambridge University Press), pp. 132-151.

GUSSENHOVEN, C. & RIETVELD, T. (2000). The behavior of H* and L* under variations in pitch range in Dutch rising contours. *Language and Speech* **43**, 183-203.

GUSSENHOVEN, C., RIETVELD, T., KERKHOFF, J. & TERKEN, J. (2003). ToDI: Transcription of Dutch Intonation (2nd Edition). http://todi.let.kun.nl/ToDI/home.htm. Accessed 28 February, 2005.

HAAN, J. (2001). *Speaking of questions: An exploration of Dutch question intonation*. PhD thesis, Netherlands Graduate School of Linguistics (LOT).

HALLIDAY, M. A. K. (1967). *Intonation and Grammar in British English.* The Hague: Mouton.

't HART, J. & COLLIER, R. (1975). Integrating different levels of intonation analysis. *Journal of Phonetics* **3**, 235-255.

't HART, J., COLLIER, R. & COHEN, A. (1990). *A perceptual study of intonation: An experimental-phonetic approach*. Cambridge: Cambridge University Press

ISAČENKO, A. & SCHÄDLICH, H.-J. (1970). *A model of Standard German intonation.* The Hague: Mouton.

ISHIHARA, T. (2003). A phonological effect on tonal alignment in Tokyo Japanese. *Proceedings of ICPhS*, Barcelona, pp. 615-618

LAAN, G. P. M. (1997). The contribution of intonation, segmental durations, and spectral features to the perception of a spontaneous and a read speaking style. *Speech Communication* **22**, 43-65.

LADD, D. R. (1983). Phonological features of intonational peaks. *Language* **59**, 721-759.

LADD, D. R. (1996). *Intonational Phonology.* Cambridge University Press.

LADD, D. R. FAULKNER, D., FAULKNER, H., & SCHEPMAN, A. (1999). Constant "segmental anchoring" of F0 movements under changes in speech rate. *Journal of the Acoustical Society of America* **106**, 1543-1554.

LADD, D. R., MENNEN, I. & SCHEPMAN, A. (2000). Phonological conditioning of peak alignment in rising pitch accents in Dutch. *Journal of the Acoustical Society of America* **107**, 2685-2696.

LADD, D. R. & SCHEPMAN, A. (2003). Sagging transitions" between high pitch accents in English: experimental evidence. *Journal of Phonetics* **31(1)**, 81-112.
LEVELT, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge MA: MIT Press.

O'CONNOR, J. D. & ARNOLD, G. F. (1973). *Intonation of Colloquial English* (2nd ed.). London: Longman.

PHEBY, J. (1975) *Intonation und Grammatik im Deutschen*. Berlin: Akademie-Verlag.

PIERREHUMBERT, J. B. (1980). *The Phonology and Phonetics of English Intonation.* PhD thesis, MIT.

PIERREHUMBERT, J. B. & BECKMAN, M. (1988). *Japanese Tone Structure*. Cambridge, MA: MIT Press.

PRIETO, P., VAN SANTEN, J. & HIRSCHBERG, J. (1995). Tonal alignment patterns in Spanish. *Journal of Phonetics* **23**, 429-451.

SCHEPMAN, A., LICKLEY, R. J. & LADD, D. R. (in press). Effects of vowel length and “right context” on peak alignment in Dutch nuclear accents. To appear in *Journal of Phonetics.*

SILVERMAN, K. & PIERREHUMBERT, J. (1990). The timing of prenuclear high accents in English. In J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology I* (Cambridge: Cambridge University Press), pp. 72-106.

SILVERMAN, K., BECKMAN, M., PITRELLI J., OSTENDORF, M., WIGHTMAN, C., PRICE, P., PIERREHUMBERT, J. & HIRSCHBERG, J. (1992). TOBI: A standard for labeling English prosody. *Proceedings of ICSLP* 2: 867-870. Banff, Alberta, Canada.

SJÖLANDER, K. AND BESKOW, J. (2000). WaveSurfer - an Open Source Speech Tool. *Proceedings of ICSLP*, Beijing, (No page numbers).

WILLIS, E. W. (2003). Prenuclear low tone alignment in Dominican Spanish. *Proceedings of ICPhS*, Barcelona, pp. 2941-2944.

XU, Y. (1998). Consistency of tone-syllable alignment across different syllable and speaking rates. *Phonetica* **55**, 179-203.

APPENDIX

List of placenames used in Study I. All were produced in the context *Woon je in … ?* (*Do you live in …?*), and elicited in random order, interspersed with other non-question utterances not reported in this study. Columns represent the four different stress patterns used.

Sww	**Sws**	**Ssw**	**Swsw**
Drimmelen	Dwingeloo	Nijmegen	Amerzoden
Gameren	Hengelo	Antwerpen	Amerongen
Ommeren	Hillegom	Driewegen	Westerhoven
Dommelen	Anderlecht	Linschoten	Zevenhoven
Someren	Denekamp	Schoonhoven	Zevenhuizen
Engelen	Veenendaal	Eindhoven	Zevenbergen
Drongelen	Doornenburg	Pijnacker	Valkenhuizen
Staveren	Voerendaal	Werkhoven	Hindeloopen
Zwinderen	Almelo	Leeuwarden	Nistelrode
Beveren	Hummelo	Kerkrade	Vrouwenpolder
Tongeren	Angerlo	Driehuizen	Vrouwenakker
Sevenum	Barneveld	Veldhoven	Lichtenvoorde
Gaanderen	Boekelo	Nieuwnamen	Lindenheuvel
Haalderen	Boelenslaan	Berlicum	Oudewater
Vlaanderen	Dodewaard	Huijbergen	Wissenkerke
Steenderen	Apeldoorn	Kaatsheuvel	Illikhoven
Genderen	Ellecom	Lanaken	Beertenshoven
Donderen	Roosendaal	Steenbergen	Benzenrade
	Zwartewaal	Bunschoten	Herkenrade
	Ulvenhout		Eikenheuvel
	Oisterwijk		Eppenhuizen
	Middelbeers		Hobbelrade
	Bennebroek		IJzendijke
	Oldenzaal		Kleverskerke
	Oldebroek		Lagenheuvel
	Rodenrijs		

		H to L (ms)		*H to L1 (ms)*	
Stress Pattern	N	Mean	SD	Mean	SD
Sww	105	202	47	162	53
Sws	138	253	53	177	53
Ssw	93	268	72	179	55
Swsw	122	332	74	213	63

Table 1. Distances in milliseconds from nuclear F0 maximum to post nuclear F0 minimum (L) and F0 turning point (L1) for each stress pattern.

		L to VF onset (ms)		*L2 to VF onset (ms)*	
Stress Pattern	N	Mean	SD	Mean	SD
Sww	105	-45	42	-13	43
Sws	138	14	35	51	43
Ssw	93	-144	61	-69	35
Swsw	122	-154	54	-68	35

Table 2: Distances in milliseconds from post nuclear F0 minimum (L) and F0 turning point (L2) to onset of final vowel, for each stress pattern.

		L to utterance end (ms)		*L2 to utterance end (ms)*	
Stress Pattern	N	Mean	SD	Mean	SD
Sww	105	-180	35	-148	35
Sws	138	-258	71	-220	76
Ssw	93	-260	61	-185	36
Swsw	122	-272	63	-186	42

Table 3: Distances in milliseconds from post nuclear F0 minimum (L) and F0 turning point (L2) end of utterance, for each stress pattern.

Contour Type	S#	Sx#	Sxx#	Total
Rise	18	19	12	49
QFR	1	21	19	41

Table 4: Distribution of Rise and Question Fall Rise contours as a function of the nuclear accent relative to the end of the utterance.

Figure 1. ToDI notation for question intonation (Haan, 2001, pp 111-112). Boxed areas roughly indicate the CV part of the accented syllable. In this paper we refer to H*LH% as Question Fall Rise (QFR) and the other three types as Rise.

Figure 2: Two examples of Question Fall Rise utterances. *Woon je in = do you live in*. H and L mark approximate loci of F0 Maximum and Minimum. L1 and L2 mark locations of beginning and end of floodplain. (A) *Gameren* is Sww, (B) *Steenbergen* is Ssw.[3]

Figure 3: Mean Syllable durations and loci of F0 maxima, minima and flood plains for each stress pattern in the read data. For clarity, onset consonant durations are included for stressed and final syllables.

Figure 4: Mean Syllable durations and loci of F0 maxima, minima and flood plains for each stress pattern in the dialogue (Map Task) data, for comparison with read data in Figure 3, above. For clarity, onset consonant durations are included for stressed and final syllables.

Figure 1

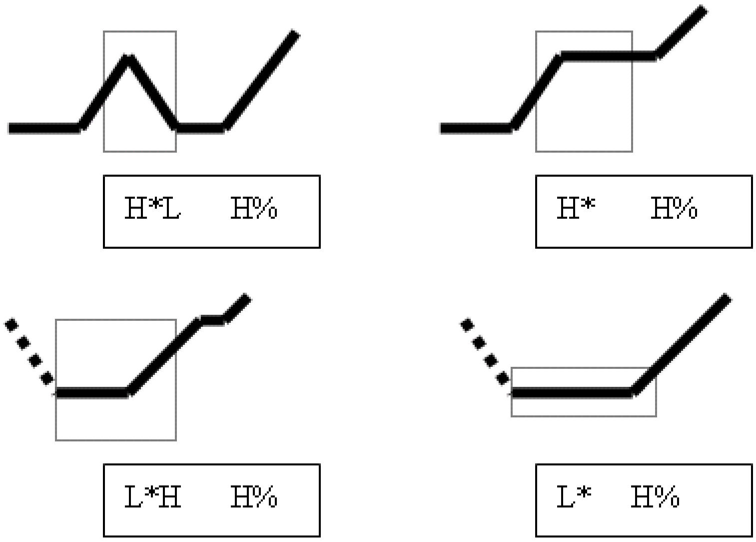
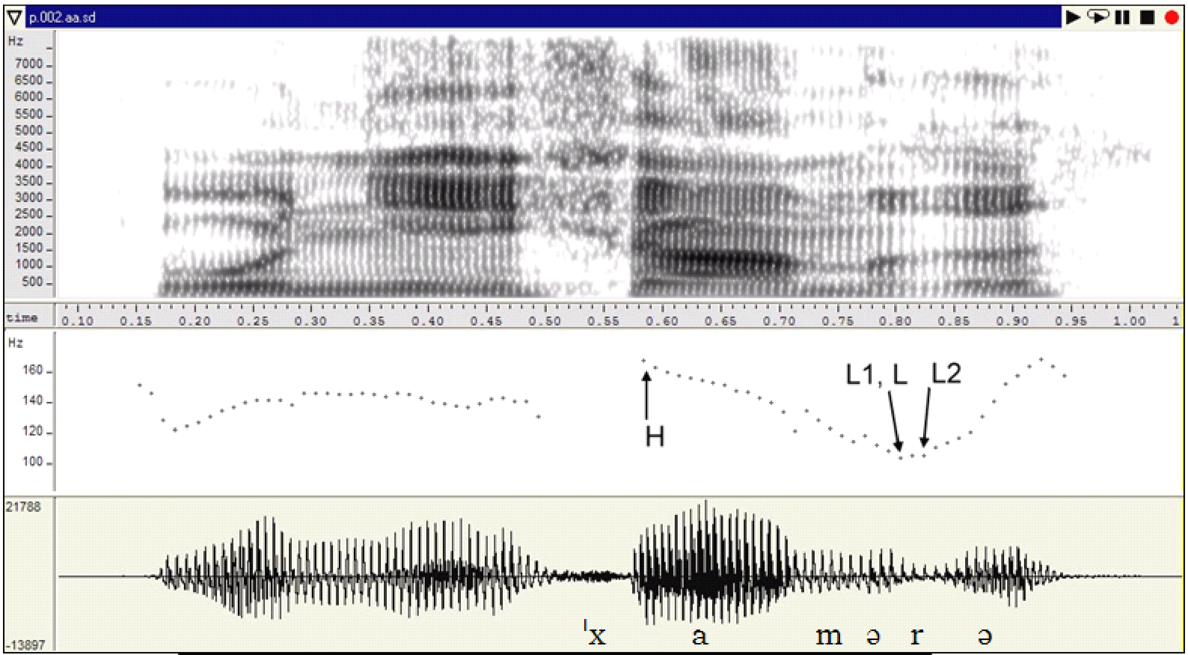
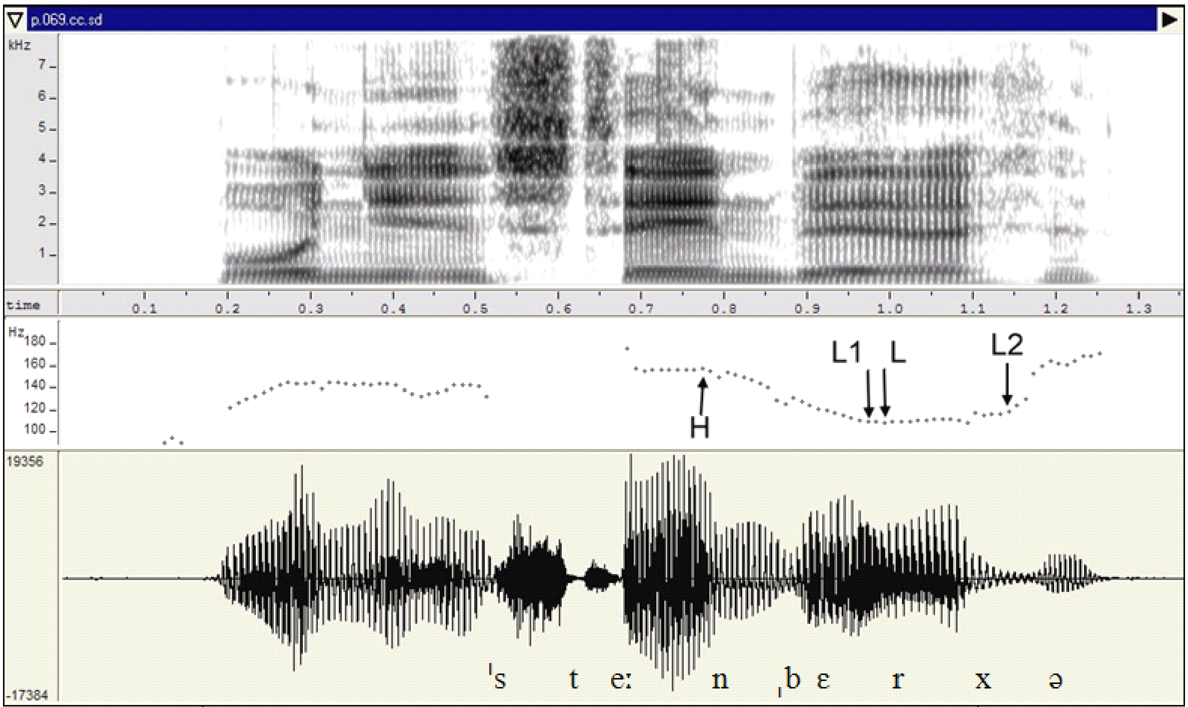


Figure 2



A: Woon je in Gameren?



B: Woon je in Steenbergen?

Figure 3

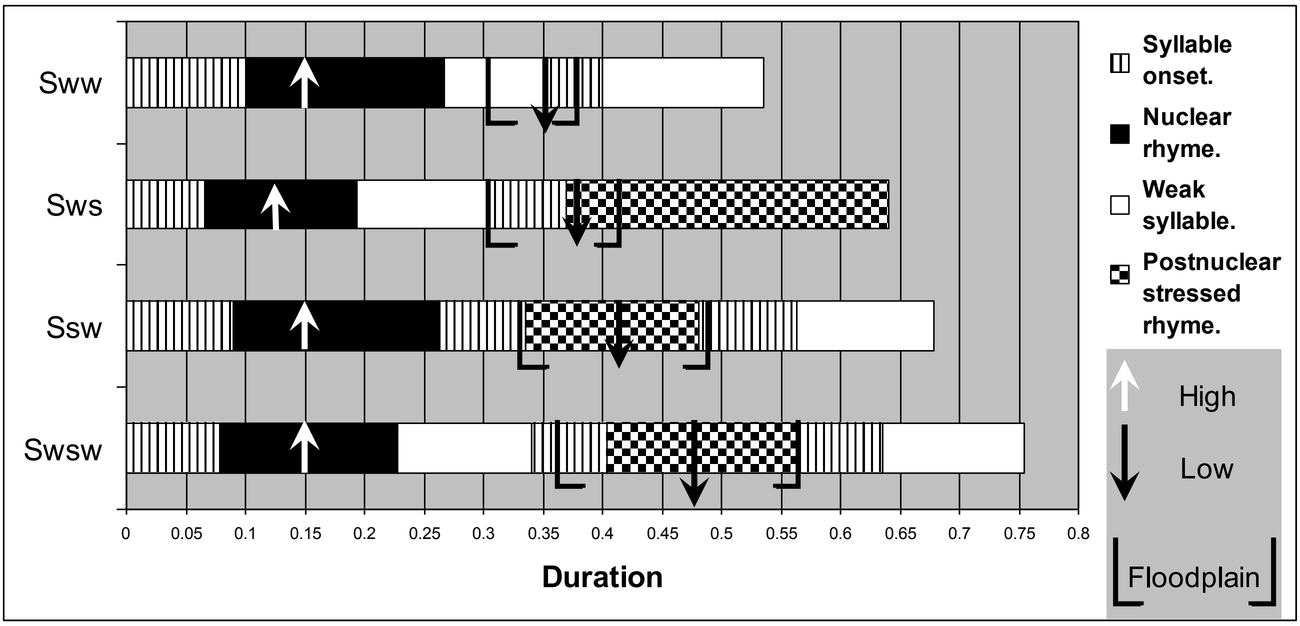
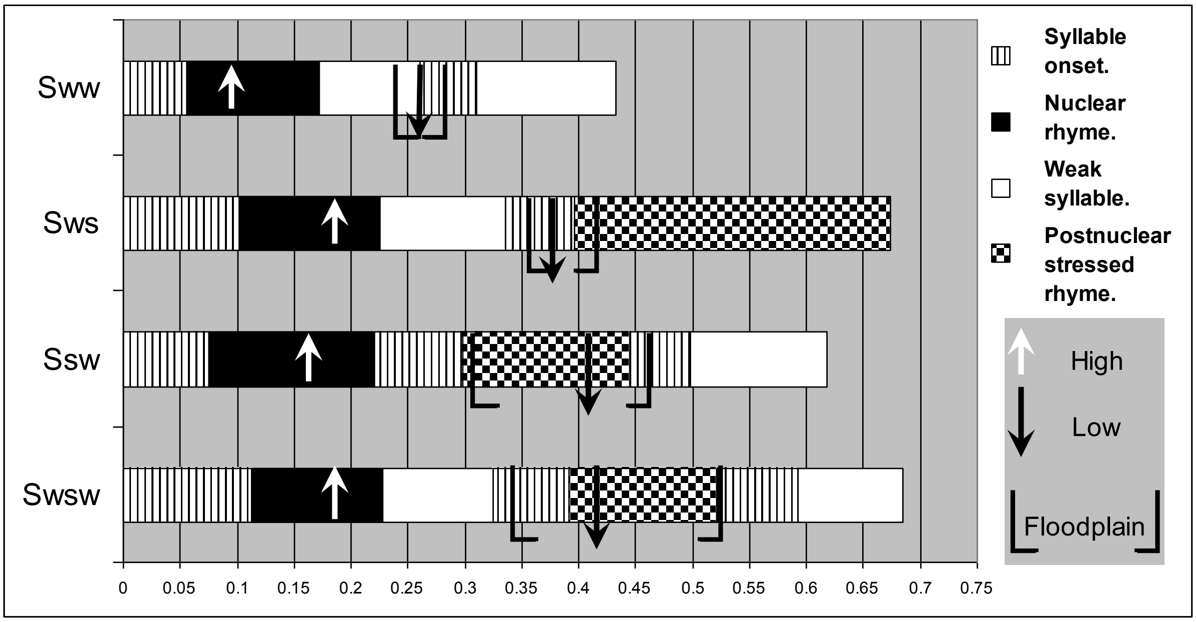

Sww
Sws
Ssw
Swsw
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55
0.6
0.65
0.7
0.75
0.8
Duration
Syllable onset.
Nuclear rhyme.
Weak syllable.
Postnuclear stressed rhyme.
High
Low
Floodplain

Figure 4


Sww
Sws
Ssw
Swsw
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55
0.6
0.65
0.7
0.75
Syllable onset.
Nuclear rhyme.
Weak syllable.
Postnuclear stressed rhyme.
High
Low
Floodplain