

# Prosodic Cues for Backchannels and Short Questions: *Really?*

Catherine Lai

January 16, 2008

## Abstract

Short questions can be ambiguous even after considering their preceding contexts. Hence, prosody may be useful for disambiguating different types of questions and their uses. For example, question bias has been linked to the presence of certain pitch accents. This paper presents a corpus study of very short questions and the contribution of prosodic cues to discourse disambiguation. This study focuses on backchannel questions which are by nature highly biased and yet sit between genuine questions and genuine backchannels. The study finds LDA and SVM classifiers do not perform better than random at separating backchannel and question *really* based on these prosodic cues. This means that, while intonation differs between broad categories of questions, theories that try to integrate prosodic cues with semantics and discourse require more than intonation, the final rise and the other usual prosodic suspects like duration and intensity.

## 1 Introduction

A key factor for question interpretation and detection is the match of form and its intonation. Prototypical question types: wh, yes/no and declarative questions, have been characterized as ending in a final fall, final rise, and a higher final rise respectively. This explains wh-questions falls and declarative questions rises. However, the characterization above is not rigid. Linking intonation to question type is much harder if we consider that there is some gradience in what we mean to be a question.

There are clearly more types of questions than the three listed above. Bolinger (1989) distinguishes alternative, tag, and reclamatory questions to name a few. We can consider

a question to be *genuine* if the speaker does not know the answer (Haan, 2002; Hudson, 1975; Banuazizi and Creswell, 1999). That is, the speaker is seeking information from the hearer. However, the line between information seeking and non-information seeking questions is vague. Even if a question seeks information, there can be ambiguity in what type of information is being sought. Moreover, interrogative forms have many other uses. For example, the broad set of questions can be used to make assertions, clarifications, acknowledge turn control, or to express agreement. In particular, it appears that there are questions that are very similar to backchannels.

*Backchannels* are described by Ward and Tsukahara (2000) as ‘short utterances, produced by one participant in a conversation while the other is talking’. Examples of English backchannels are *yeah*, *okay*, and *uh-huh*. More specifically, we can take backchannels to be utterances that do not cause the other speaker to cede the floor and that are passive contributions to the discourse.<sup>1</sup> Backchannels are generally non-information seeking and are often ignored by the other participants. The following example shows an elided polar interrogative that is more like a backchannel than a question.

- (1) B: New Jersey had the stiffest gun laws  
A: oh, do they?  
B: oh yeah  
B: but you know where where are the most  
crimes committed?

In (1), speaker A is questions B’s immediately previous statement. However, this does not appear to be a genuine question. Here, the questioner knows what the listeners response should be since they have just stated it. Short questions of this type are categorized as a *backchannel question* in Jurafsky et al. (1998).

The larger goal of this enterprise is to understand the connection between prosodic cues – such as pitch, intensity and duration – and the interpretation of questions in a discourse. It is not clear what role prosodic cues play or which prosodic cues are salient. Semantic and pragmatic theories focus almost exclusively on pitch. This has led to various forays into describing an *intonational lexicon*. So, we would particularly like to know how much weight should be place on pitch. Kochanski et al. (2005), for example, finds that loudness is a better correlate of prominence than fundamental frequency.

In order to investigate this, we focus very short questions (two words or less). Even with previous context, the interpretation and discourse function of these questions may be

---

<sup>1</sup>MDE Annotation guideline

ambiguous when stripped of sound. The first part of this paper presents a corpus study that looks at the relationship between final pitch rises and different short question types. From this study it appears that pitch slope does pattern with broad question types. However, it is not clear that finer distinction in meaning can be gleaned from the pitch alone.

As such, this paper investigates when such utterances, like the backchannel questions above, are really used for questioning. In the example above, it appears that aux-subj backchannel questions seem to generate a response, although it may be cursory. So, more specifically, the question is when are backchannel questions used to elicit new information from the listener. Beyond probing the semantics/phonetics interface, understanding the use and appearance of backchannels is important for the development of naturalistic spoken dialog systems (Noguchi and Den, 1998; Shriberg et al., 1998).

To shed some light on this, this paper focuses on epistemic *really*. *Really* frequently appears as a one word turn in spontaneous speech which is usually interpreted as a backchannel. That is, something that does not actually contribute content and/or can be safely ignored. However, similar to the the example above, *really* can also be interpreted as an actual response seeking utterance – that is, a question. A further corpus study of the word *really* was carried out investigate this. The result of this study was that something more than standard intensity, duration and pitch movement is required to differentiate these two categories.

The paper is organized as follows. Sections 2 and 3 describes the investigation of final rises and question types in the corpus data. The use of *really* as a backchannel and question is described in Section 4. Section 5 considers some accounts of biased questions from the semantics/pragmatic view that bear on the different interpretations of *really*. Section 6 describes attempts to separate backchannel and question *really* with an extended data set and more prosodic features. Finally Section 7 proposes a pragmatic analysis for *really*'s use as a backchannel. That is, *really* signals an update of the speakers prior beliefs. The magnitude of this update may be related to the prosodic variation seen in the data.

## 2 Intonation and Short Questions in Discourse

Haan (2002) has argued that the less an utterance is syntactically marked as a question, the more intonationally marked it will be (the functional hypothesis). This explains wh-question falls and declarative question rises. However, it is clear that the characterization above is not rigid. For example, there is variability in the rise/fall characterization of polar questions (**numbers from Haan**). We might expect that some of that variability comes from the

fact that questions with the same syntactic form make still have different interpretations and discourse uses. This is particularly salient for elided questions.

Interrogatives are often elided in natural speech. This leads to a natural ambiguity in their meaning and use. For example, elided wh-questions can be used to seek information (e.g. *What (did you do)?*) or to elicit repetitions (e.g. *What (did you just say)?*). These uses multiply when we consider that questions are not always used to elicit information. As mentioned above, utterances with polar interrogative syntax are very similar to backchannels in dialogue. In which case, it is not clear whether they are really questions at all.

Attempts have been made to explain intonational variation in questions by looking at the issue in a broader discourse framework. It is understood that prosodic factors can cue the interpretation a question receives and what response is required from the hearer (Bolinger, 1989; Nilsonova, 2006; Gunlogson, 2002). The question status of various types of utterances has mainly been investigated in the semantics and pragmatics literature in term an intonational lexicon and how this interacts with other linguistic structure. Interpretations of the final rise generally revolve around lack of speaker certainty or commitment to the utterance at hand.

Pierrehumbert and Hirschberg Pierrehumbert and Hirschberg (1990) argue that a high boundary tone (H%) signals that that the utterance is to be interpreted with respect to following discourse. However, as noted in Banuazizi and Creswell (1999), this applies to all prefinal utterances in the discourse. So, this distinction does not appear to be sharp enough to tease apart the different levels of meaning we want to understand.

Other approaches correlate boundary tones with certainty on the part of speaker. Steedman Steedman (2000) argues that high boundary tones imply hearer responsibility for the current proposition. Low boundary tones imply speaker responsibility. These approaches predict high rises when the speaker is uncertain of truth of the proposition and falls when they are certain. In a wh-question like *Who ate the cake?* the speaker has responsibility for the proposition *Somebody ate the cake*. Hence, wh-questions do not require a final rise.

Similarly, Gunlogson (2002) uses the idea of speaker commitment to analyze declarative questions. She argues that rising intonation signals speaker commitment. Roughly, the meaning of an utterance is the change it makes to the hearer's beliefs. She argues that both rising and falling declaratives indicate commitment of the propositional content by a discourse participant. The difference between rising and falling declaratives is that the latter commits the speaker, while the former commits the addressee. Thus, both rising and falling declaratives are semantically declaratives. The latter is a typical assertion that commits the

speaker to the content of their utterance. On the other hand, rising declaratives are like questions in that they commit the addressee to the propositional content of the utterance. This, formalized in terms of the Stalnakerian notion of Common Ground. Gunlogson, however, separates out the public beliefs of each discourse participant. So, rising declarative asserts that its propositional content is a public belief of the addressee and vice versa for falling declaratives.<sup>2</sup> has already made an appearance in the discourse.

In the same manner, rising intonation is treated as an intonational adverb expressing uncertainty by Nilsenova (2006). The bias and questioning aspects of rising declaratives are derived as by-products of pragmatic principles. This provides a nicely unified analysis for polar interrogatives and rising declaratives.

The proposed link between speaker commitment and question intonation predicts final falls/rises should help in determining question status and subtype. In fact, Liscombe et al. (2006) found that the presence of a final rise in particular to be the most useful cue for a question bearing turn in a student/computer tutor scenario – although additional intensity and timing information also improved performance. Intonation has also been found to differentiate uses of affirmative backchannels like *okay* (Benus et al., 2007). Although Gravano et al. (2007) finds that contextual conditions seem to override prosodic cues in distinguishing different uses of *okay*.

We would expect elided wh-questions to retain final falls and genuine yes/no questions to rise. True backchannel questions are highly biased, so we would expect them to fall. However, backchannel questions that have a final rise may convey something more than acknowledgement. According to the analyses outlined above, this is mostly likely speaker uncertainty of the previous utterance. These predictions are examine in the following section.

## 3 Final Rises in Short Questions

### 3.1 Data

This study used the Conversational Telephone Speech component of the the MDE RT-04 corpus (LDC2005S16). This comprises of approximately 40 hours of speech from the Switchboard-1 Corpus Release 2. The MDE annotation (LDC2005T24) provides a range of discourse metadata including question and backchannel type turns. 315 questions turns containing two or less words were located.

---

<sup>2</sup>This includes presuppositions and all that jazz.

Question type	Total	Negative slope	Positive slope
Yes/no	53	8 (15%)	45 (85%)
Wh	35	26 (74%)	9 (26%)
Reclamatory	63	9 (14%)	54 (86%)
Confirmation	14	1 (7%)	13 (92%)
Incomplete	8	2 (25%)	6 (75%)
Backchannel	89	21 (24%)	68 (76%)
Tag	26	13 (50%)	13 (50%)
Suggestion	10	0 (0%)	10 (100%)
Complementary	2	1 (50%)	1 (50%)
Alternative	7	5 (71%)	2 (29%)

Table 1: *Slopes and short questions*

F0 contours were extracted from final words using `praat`. Some speakers participated in more than one conversation, so the F0 data for each speaker was combined. Outliers were then removed for each individual question. These were defined as F0 values outside the range  $(q1 - 1.5 \times (q3 - q1), q3 + 1.5 \times (q3 - q1))$ , where  $q1, q3$  are the first and third quartiles of combined F0 data of each individual speaker. The F0 data was further normalized to a log scale between 0-10 according to the following formula:

$$F0_{norm}(x) = (10 / \log_{10}(\max(F0) / \min(F0))) \times \log_{10}(x / \min(F0)).$$

The pitch slope for each final word was fitted from the normalized F0 data using the linear regression function `lm` in R. Normalization eliminated five questions, two more questions were eliminated due to lack of speaker information, while a further question was removed because its transcription clearly did not match the audio recording. This left 307 questions.

Genuine yes/no questions (Y) and wh-questions (W) were manually identified along with a number of other question types. Reclamatory (R) questions elicited repetitions of the previous utterance. Confirmation (C) questions clarified the current topic of discourse. Incomplete (I) questions attempted to elicit non-specific speech from the hearer (e.g. *Hello?*). In tag questions (T), speakers questioned their own prior statement. Backchannel questions were utterances where the speaker questioned an immediately prior statement of the hearer. Speakers offered possible but indefinite options to the hearer in suggestion (S) questions. Complementary (Cp) questions elicit the same responses as wh-questions but without the

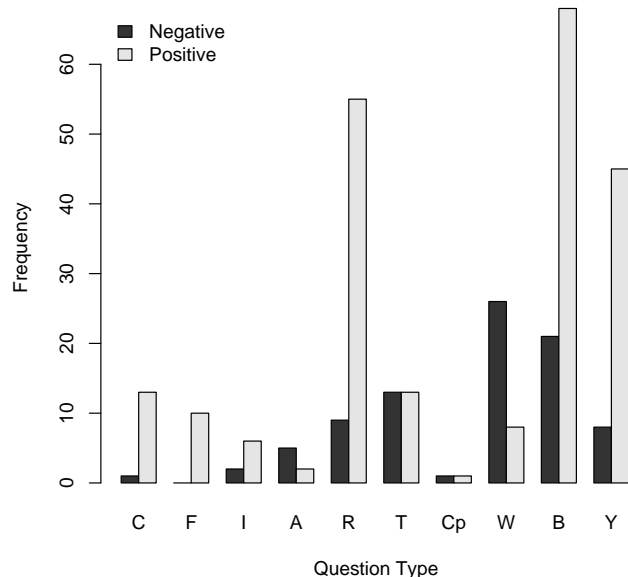


Figure 1: *Proportions of samples with negative and positive slopes in different question types*

wh-word. Alternative (A) questions present a list of possible alternatives.

### 3.2 Observations

The proportions of utterances with positive and negative slope across question types is shown in the Figure 1. Pearson’s  $\chi^2$  test was used to see if there were any differences between positive and negative slope counts across gender, age (10 year groups) and accent. For these cases, the test did not show a significant differences for the question data as a whole. Similarly, no significant difference in age or sex was found when looking at specific question types using the Fisher exact test (due to the small amount of data). However, there was a significant different in comparing accent type and reclamatory questions (p-value = 0.03).

Although the amount of data is relatively small, we can make some general observations. Yes/no questions do appear to have final rises while wh-questions have final falls in the majority of cases. Confirmation, incomplete, and suggestion questions tend to end in a

rise. These questions are really declarative questions so this result is inline with Haan's functional hypothesis and also Gunlogson's analysis of rising declaratives. The ambivalence associated with these types of questions also agrees with the association of rising and speaker uncertainty. However, it appears that wh-reclamatories are distinguished from genuine wh-questions by their final rise. This seems to group wh-reclamatories with other echo questions like the Gunlogson's rising declaratives mentioned above. That is, asks the hearer to commit the question they just asked to the common ground.

This categorization shows that broad categories of questions do have different final rise characteristics. However, it is not at all clear that pitch movement can make finer grained distinctions. In particular, it is not clear what the presence of a final rise means for disambiguating the different uses of backchannel-like questions discussed above. Backchannel questions were used to convey a range of signals from acknowledgement to denial and surprise and with varying pitch curves. Figure 2 shows pitch tracks for examples (1) and (2) and one additional 'surprised' *really*. It seems plausible that pitch movement could be an indicator of deviation backchannel status alongside a number of other prosodic factors.

The rest of this paper is concerned with how listeners can distinguish backchannel and question *really*. The next section provides a qualitative study of how *really* is used in dialogue as both a question and a backchannel. We also consider some literature on question bias that directly bears on the interpretation of *really*. Section 6 investigates whether backchannel and question *really* are distinguishable based on a number of prosodic features.

## 4 *Really* as backchannel and question

**Backchannels:** \* Agreements are also confused with backchannels Shriberg et al. (1998).

As mentioned above, *really* is used in dialogue as both a backchannel or as something more like a question. In the following dialogue (1), *really* appears to be a *question* (speaker B justifies his question in the last line).

- (2) B : You like Lubbock better than Dallas  
A : Yeah  
B : Why?  
A : Uh, because people are so much nicer  
B : Really?  
A : Yes  
B : Well people are nice here in Dallas

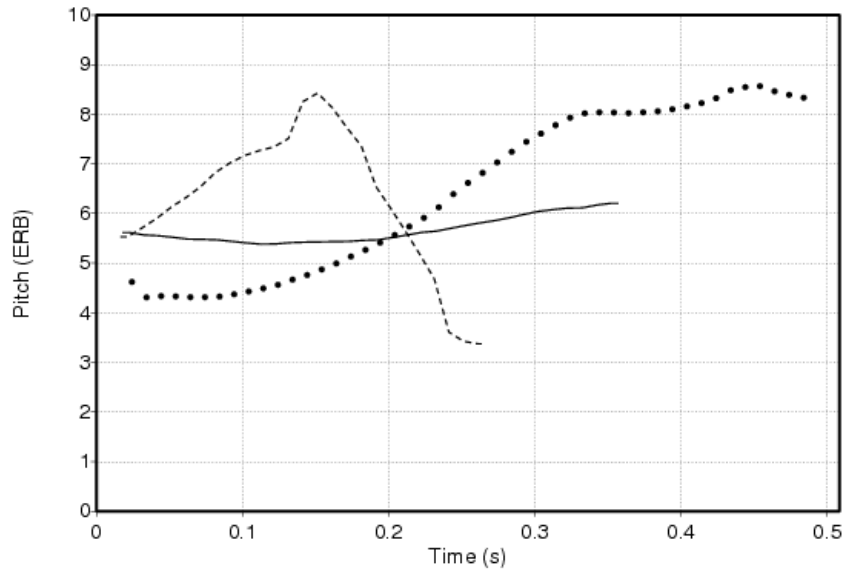


Figure 2: *Two types of really*: dashed, speckled lines are surprised, the solid line is a backchannel.

This contrasts with the following dialogue where *really* was annotated as a backchannel rather than a question.

- (3) B: Oh I've got some Chinese Hollies  
that are just outrageous  
B: They they are very sharp  
A: Oh really  
B: Do you do your own uh lawn maintenance?  
A: Yeah

Speaker A's *really* did not require nor elicit a response from speaker B.

However, although *really* in (4) was marked as a question it is very similar to a backchannel. Even though it apparently required a response (*Yeah*), there did not appear to be any need for either speaker to justify their statements any further.

- (4) B: I kind of enjoyed that boat  
I looked at today  
B: It's nice and clean

A: Really?  
B: It wasn't [interrupted]  
B: Yeah  
A: Did it have a cabin?

Note, *really* does not necessarily have the same backchannel effect as agreement type backchannels like *uh-huh*. It is quite hard to get the questioning effect with *uh-huh*:

(5) A: You have to be a bit careful when removing the tyre or you might tear it.  
B: uh-huh?  
A: #Yes.

So, *really* is used as both a question and backchannel, where questionhood clearly involves some gradience. In fact, it appears that the *really*'s even shorter forms of the elided polar interrogatives involving *really*. These also appear as backchannel questions in spontaneous speech.

(6) A: I don't subscribe to any magazines  
A: Do you?  
B: I subscribe to lots of magazines  
A: Do you really?  
B: Yes  
B: My family w[ell] well we're we're kind of renowned for being  
big magazine subscribers

*Really* is used as both a question and backchannel, where questionhood clearly involves some gradience. In the following, I will assume that *really* is structurally a question. The question this is why is it used as a backchannel and is it used to actually elicit information. Since the speaker has the answer from the previous utterance, *really* must be a highly biased question if it is interpreted as one at all. That is, the speaker already has beliefs about what the answer will be. Possible sources of bias and the contribution of intonation are reviewed in the next section.

## 5 Biased Questions

The backchannel questions considered above seem to express a mixture of bias (in terms of beliefs) and uncertainty at the same time. These provide a framework for understanding the backchannel/question distinction.

*Really* has been linked to question bias in the case of negative polar questions. A number of studies have also observed that negative polar interrogatives can also be biased towards both *yes* and *no* answers. These two types of bias were categorized as inner (expect *no*) and outer (expect *no*) negation by Ladd (1981). These readings can be forced by the inclusion of Negative and Positive Polarity Items (NPIs, PPIs). Examples are given in the following:

- (7) A: Wow look at all the steakhouses! I don't think we're going to find a place to eat here.
- a. B: Aren't there (any) vegan restaurants around here (at all)?  
(Inner negation: expect no)
- b. B: Aren't there (some) vegan restaurants around here (too)?  
(Outer negation: expect yes)

More generally, if we observe the question  $?p$ , INPQs can be thought of as double checking  $\neg p$  while ONPQs double check  $p$ . In this sense, the proposition presented in an ONPQ can form a denial or counterevidence to assertions made by the listener previously. This dual function of check and denial is similarly to that rising declaratives and backchannel question *really* discussed above. A number of analyzes that try to account for this data have been put forward.

Romero (2006) account for these two readings via a scope ambiguity introduced by the interaction of negation and the VERUM operator. VERUM is given the same meaning as epistemic *really*:

$$\begin{aligned}
 (8) \quad \llbracket \text{VERUM}_i \rrbracket^{gx/i} &= \llbracket \text{really} \rrbracket^{gx/i} \\
 &= \lambda p_{\langle s, t \rangle} . \lambda w . \forall w' \in \text{Epi}_x(w) [\forall w'' \in \text{Conv}_x(w'') [p \in \text{CG}_{w''}]] \\
 &= \text{FOR-SURE-CG}_x
 \end{aligned}$$

The fact that NPIs are not licensed ONPQs is then due to the intervening VERUM. They note that the presence of NPIs/PPIs is not explained by the decision theory based account of van Rooy and Safarova (2003). The two readings above have the following logical forms given below.

- (9) Aren't there vegan restaurants around here?
- a. INPQ: [ Q VERUM [not [There are vegan restaurants around here]]]  
 $= \{ \text{FOR-SURE-CG}_x(\text{There are no vegan restaurants around here}),$   
 $\neg \text{FOR-SURE-CG}_x(\text{There are no vegan restaurants around here}) \}$

- a. ONPQ: [ Q [not [VERUM [There are vegan restaurants around here]]]  
 = {FOR-SURE-CG<sub>x</sub>(There are vegan restaurants around here),  
 ¬FOR-SURE-CG<sub>x</sub>(There are vegan restaurants around here) }

This derives the ‘double checking’ duality of negative polar questions. Romero and Han (2004) argue that *really* (VERUM) carries a negative epistemic implicature when it appears in positive polar questions. That is, it expresses that speaker originally believed the answer to be no. The epistemic implicature can be either positive or negative in negative polar questions because of the scope ambiguity claimed above.

As a double-checker, *really* expects a *yes* answer and hence aligns with ONPQs. We can analyze *really* questions as elided positive polar questions. Here the analysis of *really* as VERUM will provide the same LF as that of an ONPQ.

- (10) A: The crocodile bit my head off!  
 B: Did it really bite you?

- a. [ Q VERUM [There crocodile bit you]]  
 = [Q not [ VERUM [There crocodile bit you]]]  
 = {FOR-SURE-CG<sub>x</sub>(The crocodile bit you),  
 ¬FOR-SURE-CG<sub>x</sub>(The crocodile bit you) }

B’: The crocodile (really) bit you?

The previous example also shows that rising declaratives have much the same effect as the ONPQ and the *really* flavored positive polar question.

Gunlogson (2002) observes that rising declaratives are licensed in situations where content of the rising declarative has already been asserted. We can see in the following example that *really* can have the same effect as the rising declarative.

- (11) *John and Mary on their first date:*

J: I’m a member of the communist party.

- a. M: #Are you a member of the communist party?  
 b. M: You’re a member of the communist party?  
 c. M: #?You’re a member of the communist party.  
 d. M: Really?

e. M: Aren't you an investment banker? (ONPQ)

*Really* appears to act as a check on the addition of the previous proposition into the common ground. Like rising declaratives, in this case it is up to the *really*-listener to perform the final commit. In some ways, rising declaratives make more vehement, pointed denials than their ONPQ counterparts or *really*. However, they do not seem to provide the counterevidence denials that are possible with ONPQs (11e). Counterevidence denials are also limited with *really*— they can only make appeal to the general notion of reality rather than specific counterexamples.

However, the ignorability of *really* as a backchannel indicates that the listener does not necessarily have to respond to this call to commit the proposition in this particular case. So, something more fine grained is necessary to give us the story on backchannel *really*. It seems plausible that prosody could play a role differentiating *really* on the spectrum from check to denial. Here, checks align with the backchannel interpretation while denials are more like real questions.

It is interesting to note that here that *really* can precede both INPQs and ONPQs in certain contexts. Consider the following:

(12) Then the crocodile came towards Tim. He ran away but I just laughed.

a. Oh really? Didn't the croc scare you *at all*? (INPQ)

b. Oh really? Didn't the croc scare you *too*? (ONPQ)

However, *really* seems infelicitous when previewing ONPQs that don't have the element of surprise. For example, the following gives an ONPQ that is a suggestion rather than a denial.

(13) A: So much choice for the non-cow eater! Let's eat!

B: (#Oh really?) Aren't there some vegan restaurants around here?

That is, *really* appears to be more about signalling the speakers surprise than asking the listener to check something into the common ground.

Along these lines Reese (2007) argues that prosody does distinguish ONPQs from INPQs, but in a different way to that proposed by Romero and Han. His analysis focuses on the fact that a *no* answer to an ONPQ (or *really*?) does not really mean that the addressee is unsure of whether the proposition should be placed in the common ground. Rather, that the proposition is untrue.

Instead, Reese proposes that the outer negation interpretation of a negative polar interrogative is really metalinguistic negation: such questions express denial or counterevidence to something in the (immediately) preceding discourse. This a completely different entity to inner negation which seem to be restricted to checks or confirmations of the common ground. This can be triggered by the presence of an L\*+H nuclear pitch accent (the ‘contradiction contour’ (Pierrehumbert and Hirschberg, 1990)). This allows him to present a unified theory of negative bias in negative polar interrogatives and emphatic focus questions. The latter being positive polar questions that receive rhetorical type readings with emphatic focus. So, *really* with intonation might actually be signalling some sort of metalinguistic negation on the previous utterance.

Given this and previous invocations of the ‘intonational lexicon’, one might expect a similar prosodic distinctions to separate backchannel *really* from question *really*. The differentiability of backchannel and question really is investigated in Section 6. The goal here is to see if the prosodic features of *really* make a reliable contribution to the interpretation along the lines proposed in the theoretical linguistics literature. Prosodic features such as intensity and duration are often ignored in semantic analyses, which focus almost exclusively on pitch. In this manner, this study is an attempt to find out whether semanticists are need to broaden the idea of an intonational lexicon to incorporate more elements of

In fact, it appears that the usual prosodic suspects – pitch, intensity, duration – are *not* sufficient to differentiate these two categories of *really*.

## 6 Prosodically Distinguishing Backchannel and Question *Really*

This section argues that the usual suspects of prosody – intonation, duration and intensity – do not provide useful cues for distinguishing backchannel questions from pure backchannels. The second subsection explores the prosodic differences, and the final subsection tests whether these differences provide useful cues.

### 6.1 Data

These experiments expanded the previous data set to include MDE 2003 annotations (LDC2004T12) and audio (LDC2004S08) from the Switchboard I. Instances of *really* labelled as a backchannel (450) or a question (130) (*really<sub>q</sub>*, *really<sub>b</sub>* in the following) were

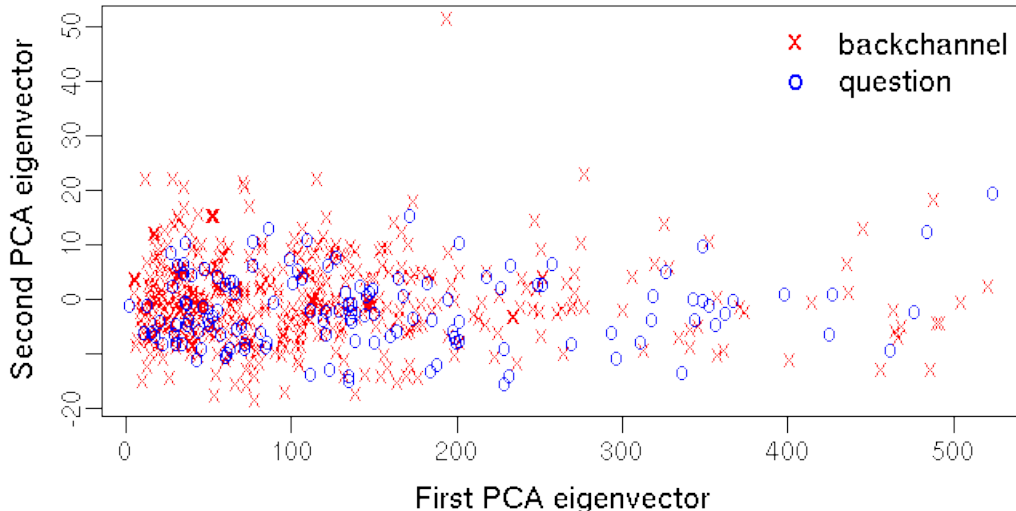


Figure 3: *Projection on to the first two dimension of the PCA space.*

extracted using timing information from the MDE annotations.

Pitch and intensity measurements were made using `Praat` with samples at 1ms intervals (to provide enough points for curve fitting). The pitch data was normalized to a log scale from 0-10 as previously. The mean intensity of the speaker for their entire conversation was subtracted from the intensity measurements. Pitch and intensity curves were approximated using orthogonal polynomial curve fitting with order 5 Legendre polynomials (c.f. Kochanski et al. (2005)). Six coefficients were recorded for each pitch and intensity curve (p0-p5, i0-i5 resp.). Legendre polynomial fitting has the nice property that coefficients derived from this process are not fraught with the correlation problems of those from standard polynomial fitting. They also have intuitive interpretations: the first coefficient indicates general bias, the second indicates overall slope, the third indicates convexity and so on. Information about the utterance may be signalled by from non-linear characteristics of pitch and intensity the curve, such as convexity, so this is a desirable property.

Beyond this, the correlation between raw intensity and pitch at 10ms intervals was also derived (`Corr`) for each utterance, as well overall pitch range (`prange`). Duration (`Dur`) and relative time position of pitch minimum (`p.min.d`) and maximum (`p.max.d`) were also recorded.

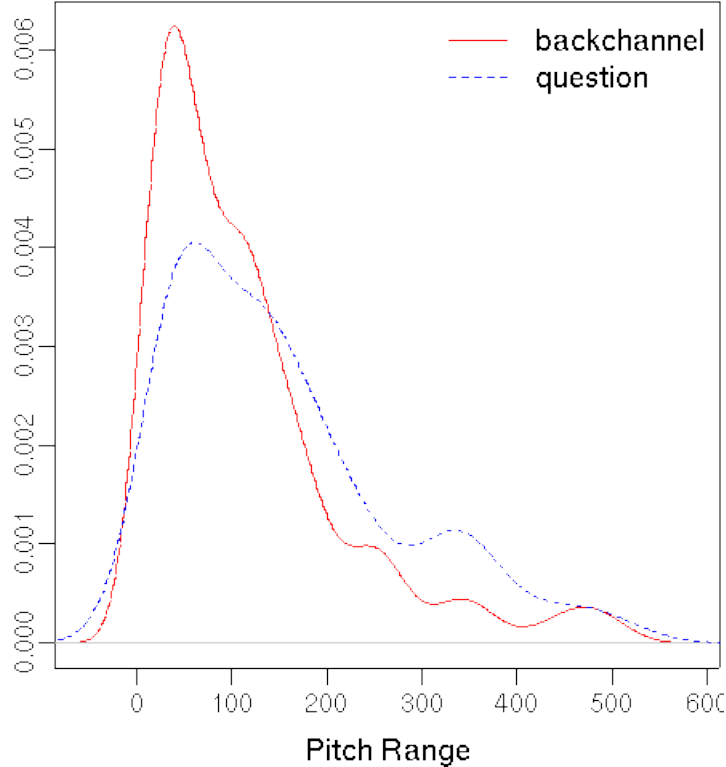


Figure 4: *Probability densities for the `prange` feature.*

## 6.2 Exploring the Differences

Principal components analysis was carried out on all the numeric features using the R function `prcomp`. The principal component with the largest standard deviation (109.4) was dominated by `prange`. The second component (standard deviation, 7.7) pointed predominantly in the direction of `i0` (intensity bias) and `i1` (the linear coefficient for intensity). Notably, `p1`, pitch slope, did not seem to account for much of the variance in the data. This suggests that theories that rely on the direction of pitch movement will not distinguish *really<sub>q</sub>* and *really<sub>b</sub>*.

In anycase, it does not appear that any of these components differentiate backchannel and question *really*. This can be seen from Figure 3 which shows the overlapping distribution of the data transformed to the space spanned by the principal components and then projected onto the first two components.

Analysis of the data feature by feature suggests that there are differences in how *really<sub>q</sub>*

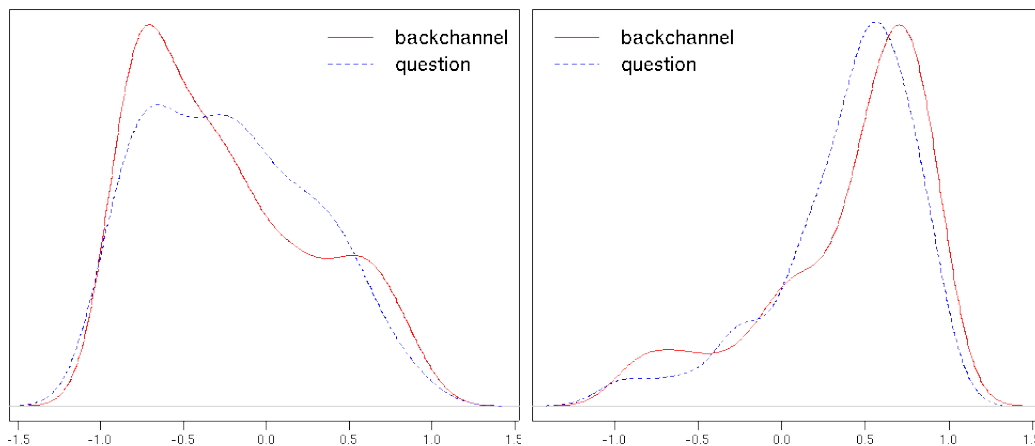


Figure 5: *Empirical probability densities of pitch/intensity correlation for rising (left) and falling (right) pitch.*

and  $really_b$  are produced. However, the amount of distributional overlap leaves the hypothesis that the listener can actually differentiate the data based on these cues, somewhat weak. For example, density plots for the pitch range data are shown in Figure 4. The non-parametric bootstrap was employed to see if the means of these two distributions differed as the data is clearly not normally distributed. Both sample means (115Hz, 154Hz) fall outside the 99% bootstrap confidence intervals for the other class ((102.7, 128.3) and (124.3, 184.4) resp.). However, it is clear from the density plot that the overlap in distributions is great.

Similarly, Figure 5 shows empirical density plots for the pitch/intensity correlation data. These are separated according to whether pitch rising or falling ( $p1$ ) to account for the fact that intensity generally falls at the end of an utterance. This plot indicates that the distributions for rising pitch backchannels and questions may be different. It is tempting to say that a speaker is more likely to maintain intensity when producing  $really_q$  than  $really_b$ . However, once again there is almost total overlap in the distributions, so it is unlikely that listeners use this cue to determine that a given  $really$  is a backchannel or question. The means for positive  $p1$  samples are not significantly different at the 0.05 level ( $really_q$ : -0.221 (-0.320, -0.124),  $really_b$ : -0.246 (-0.312, -0.188)).

	Error (Std. Error)	Bias	95% CI
Baseline	0.245 (0.018)	-0.005	(0.21,0.28)
LDA	0.244 (0.019)	-0.013	(0.21,0.28)
SVM	0.267 (0.019)	-0.041	(0.23,0.30)

Table 2: *Estimates for classification errors and 95% confidence intervals from bootstrapped bias-corrected cross-validation.*

### 6.3 Testing Prosodic Cues

The overlapping distributional data above suggests that it unlikely that the prosodic features described above can differentiate *really<sub>q</sub>* and *really<sub>b</sub>*. To further test this hypothesis, two classifiers were built in an attempt to separate the data. The first was a classifier based on Linear Discriminant Analysis (LDA) as implemented in R (`lda`). The second was a Support Vector Machine (SVM) classifier with radial basis function kernel (`libsvm` via R). The 10-fold cross-validation error rates are shown in Table 2 alongside bootstrap estimates bias, standard error, and 95% confidence intervals (1000 bootstrap samples).

The classifiers clearly do not fare much better than the baseline that simply categorizes everything as a backchannel. They certainly do not reach error rates outside the 95% confidence interval for the cross-validation error of the baseline. This supports the hypothesis that these two categories are not separable on the basis of these features.

The prosodic features examined in the previous section do not exhaust the possible contributions of prosody to meaning. Certainly, the intuition remains that some *really*'s do sound like they need an answer. The fact that this is not captured by the pitch, intensity and duration measurements considered here does not mean this is the end of the line for prosody and *really*. Further investigation of the prosodic characteristics is necessary. In anycase, the difference between question and backchannel does not quite seem to fall out of theories of question bias. The next section discusses the use of *really*. This identifies some circumstances where *really* may be interpreted as a backchannel and/or ignored. This also suggests what separating prosodic features may contribute to meaning if they are identified in further work. That is, something beyond speaker commitment.

## 7 *Really* as a backchannel: keeping aligned and updated

This section proposes that the basic use of *really* is to signal that, adding the previous proposition will require the speaker to update their prior beliefs. This is different signalling hearer commitment, because it does not necessarily require the hearer to commit anything to the common ground.

We will continue to assume that one-word-turn *really* comes from an elided positive polar interrogative which partitions possible worlds in the manner of VERUM (cf. (8)). This is in line with the negative epistemic implicature assigned to *really* by Han and Romero in many cases. However, we still need to account for *really*'s use as a backchannel.

*Really* seems to be treated as a backchannel (i.e. ignorable) when the addressee determines that the proposition that *really* is applied to,  $p$ , already had a high probability in the speaker's beliefs. This means that it is not necessarily the case that the speaker did not believe  $p$  or even that  $p$  was viewed as unlikely. Rather, that adding  $p$  to the common ground necessarily involved updating their beliefs, increasing the probability of  $p$ . In order to interpret *really* as a backchannel, the addressee has to infer that their previous statements regarding  $p$  were sufficiently convincing. If not, they should provide further evidence to facilitate discourse alignment. That is, *really* is interpreted as a question. We may expect that the magnitude of the update is what is signalled by prosodic factors.

We can now consider why *really* is used as a backchannel. We saw previously that *really* does not have identical function with agreement backchannels like *uh-huh* (cf. Example (5)). Agreement backchannels from one discourse participant, A say, signal that the contribution of participant B was in or inferable from A's beliefs. In this sense, agreement backchannels signal a lack of surprise on the part of the backchanneler. This makes sense from the viewpoint that discourse participants derive utility from knowing that their mental models are well aligned (Pickering and Garrod, 2004). However, pragmatic theories argue that that utterances should be *informative* and *relevant*. Given the previous discussion, one might describe *really<sub>b</sub>* as a backchannel of Gricean Quantity and Relation. On the other hand, agreement backchannels can be seen as signals of *Quality* or *truth*. That is, different types of backchanneling occurs because discourse participants derive utility from knowing that they are not breaking different conversational maxims.

*Really* is interpreted as information seeking when misalignment between discourse par-

ticipants is ‘big enough’<sup>3</sup> There are at least some clear cases where the addressee would have good evidence that the speaker requires further evidence or elaboration. This occurs where *really* is a response to the answer to an apparently genuine information seeking question. That is a question unmarked by bias intonation or negation properties discussed above.

- (14) A: What is your very favorite song that Billy Joel has done?  
B: Probably Piano Man  
A: Really?  
B: Yeah  
B: I uh m[aybe] maybe just because it is like on i think it is  
the first one on the on the CD

In this case, *really* allows the listener to do post-hoc inference on the speakers prior beliefs. That is, that they were biased with respect to the partition induced by that question differently from the listener. In the example, above the listener learns that probability assigned to someone having Piano man as their favorite Billy Joel song was low for the speaker.

In fact, listeners may simply not have enough information to make this inference. Speakers follow up *really* with further questions. That is, more updating and more evidence is required than a simple yes or no answer. An example is show below.

- (15) A: People think Charlotte is the big town you know and it’s small  
B: oh wow  
B: really?  
A: yeah  
B: What’s the population?  
A: uh Charlotte?

## 8 Conclusion

Short questions can be ambiguous in a way that is not always resolvable from the previous context. This suggests that prosody has a part to play in disambiguating different uses for different types of questions. As we saw, backchannel questions can be used both as an acknowledgement and as real questions. They can also project many shades Of meaning

---

<sup>3</sup>Perhaps that should be: *really* is labelled as a question by annotators!

in between. However, they all seem to be clearly geared towards expressing the speaker's uncertainty above previous utterances in the discourse. Hence, backchannel questions pose an interesting testing ground for theories positing an intonational lexicon and its ramifications for discourse.

In that sense, the second part of this paper was an attempt to find out if some element of this intonational lexicon could systematically differentiate *really<sub>q</sub>* and *really<sub>b</sub>* in the similar manner to the way final rises change the interpretation of declarative sentences. It seems clear at this point that the prosodic features considered (including intonation) are not enough to make this distinction. The fact that *really<sub>q</sub>* and *really<sub>b</sub>* do not seem separable on the features selected above, does not mean that there is no role for prosody in distinguishing the different uses of *really*. Of course, there may be latent cues that were not covered by this analysis. The results of this *really* study strongly suggest, however, that theories that try to integrate prosodic cues with semantics and discourse should go beyond intonation, the final rise and the other usual prosodic suspects: plain duration and intensity.

However, the fact that listeners can differentiate between *really<sub>q</sub>* and *really<sub>b</sub>* suggest many avenues for further study. In the end, I proposed that the primary role of *really* was to allow the speaker to signal that they had received new information from the addressee. That is, the speaker had updated their prior beliefs. This sort of turn requires a response when the addressee judges that the update was a substantial one. In order to maintain a smooth and aligned discourse, the addressee then provides further elaboration or explanation of the previous claim.

Further perception studies on *really* and other short questions with multiple uses will help tease out prosodic cues and their relation to updates of the common ground and speaker beliefs.

## References

- Banuazizi, A. and Creswell, C. (1999). Is that a real question? Final rises, final falls and discourse function in yes-no question intonation. In *Papers from the 35th Regional Meeting of the Chicago Linguistic Society*.
- Benus, S., Gravano, A., and Hirschberg, J. (2007). The prosody of backchannels in american english. In *Proceedings of ICPHS 2007*, pages 1065–1068.

- Bolinger, D. (1989). *Intonation and its uses : melody in grammar and discourse*. Stanford University Press, Stanford, Calif.
- Gravano, A., Benus, S., Chavez, H., Hirschberg, J., and Wilcox, L. (2007). On the role of context and prosody in the interpretation of okay. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 800–807. Association for Computational Linguistics.
- Gunlogson, C. (2002). Declarative questions. In Jackson, B., editor, *Proceedings of Semantics and Linguistic Theory XII*. CLC Publications.
- Haan, J. (2002). *Speaking of Questions: An Exploration of Dutch Question Intonation*. PhD thesis, Utrecht: LOT Graduate School of Linguistics.
- Hudson, R. A. (1975). The meaning of questions. *Language*, 51(1):1–31.
- Jurafsky, D., Shriberg, E., Fox, B., and Curl, T. (1998). Lexical, Prosodic, and Syntactic Cues for Dialog Acts. In *Proceedings of ACL/COLING-98 Workshop on Discourse Relations and Discourse Markers*, pages 114–120.
- Kochanski, G., Grabe, E., Coleman, J., and Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America*, 118:1038.
- Ladd, D. (1981). A First Look at the Semantics and Pragmatics of Negative Questions and Tag Questions. In *CLS 17*, pages 164–171.
- Liscombe, J., Venditti, J., and Hirschberg, J. (2006). Detecting Question-Bearing Turns in Spoken Tutorial Dialogues. In *Interspeech'06*. ISCA.
- Nilsenova, M. (2006). *Rises and Falls. Studies in the semantics and pragmatics of intonation*. PhD thesis, University of Amsterdam.
- Noguchi, H. and Den, Y. (1998). Prosody-Based Detection of the Context of Backchannel Responses. In *Fifth International Conference on Spoken Language Processing (ICSLP'98)*. ISCA.
- Pickering, M. and Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(02):169–190.

- Pierrehumbert, J. and Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In Cohen, P., Morgen, J., and Pollack, M., editors, *Intentions in Communication*. MIT Press, Cambridge.
- Reese, B. (2007). *Bias in Questions*. PhD thesis, University of Texas at Austin.
- Romero, M. (2006). Biased Yes/No Questions: The Role of VERUM. *SPRACHE UND DATENVERARBEITUNG*, 30(1):9.
- Romero, M. and Han, C. (2004). On Negative Yes/No Questions. *Linguistics and Philosophy*, 27(5):609–658.
- Shriberg, E., Bates, R., Taylor, P., Stolcke, A., Jurafsky, D., Ries, K., Coccaro, N., Martin, R., and Meteer, M. (1998). Can Prosody Aid the Automatic Classification of Dialog Acts in Conversational Speech? *Language and Speech*, 41(3-4):443–492.
- Steedman, M. (2000). Information Structure and the Syntax-Phonology Interface. *Linguistic Inquiry*, 31(4):649–689.
- van Rooy, R. and Safarova, M. (2003). On polar questions. *Proceedings of Semantics and Linguistic Theory*, 13.
- Ward, N. and Tsukahara, W. (2000). Prosodic features which cue back-channel responses in English and Japanese. *Journal of Pragmatics*, 32(8):1177–1207.