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

Catherine Lai

Department of Linguistics
University of Pennsylvania

Speech Prosody
May 8, 2008
A key factor for question interpretation and detection is the match of form and its intonation. Wh-questions fall, yes/no questions rise, declarative questions have a higher final rise. 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, questions can be used to make assertions, clarifications, acknowledge turn control, or to express agreement. They sometimes also act as backchannels. This paper investigates when a question can be interpreted as a backchannel.
Outline

- *Really* as a backchannel and a question.
- Question meaning and prosody.
- Final rises in short questions
- Prosodic cues for *really* as a question and backchannel
- A pragmatic account?
Backchannels

- Ward and Tsukahara (2000): ‘short utterances, produced by one participant in a conversation while the other is talking’.
- For example, *yeah, okay, uh-huh*.
- More specifically: backchannels are utterances that do not cause the other speaker to cede the floor and that are passive contributions to the discourse.
- Backchannels are generally non-information seeking and are often ignored by the other participants.
- However, sometimes short questions act as backchannels – i.e. they’re backchannel questions! (Jurafsky et al., 1998).
An example of a real question?:

(1) 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
Really as a backchannel

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

(2) 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
Something in between?

The following was annotated as a question. 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.

(3) 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?

In the following, I will assume that really is structurally a question. The question is why is it used as a backchannel and when is it used to actually elicit information.
Really like a rising declarative?

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.

(4)  *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.
The goal here is to investigate the interaction of prosodic cues – such as pitch, intensity and duration – and the interpretation of questions in a discourse.

Semantic and pragmatic theories focus almost exclusively on pitch. This has led to various forays into describing an *intonational lexicon*.

What components of pitch do most of the work for us? (c.f. Kochanski et al. (2005))

Backchannel questions are a good probe for this problem.
What else?

▶ 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 (like backchannel questions).
▶ The second part investigates when backchannel questions are interpreted as questions. Our testing ground is epistemic *Really*. 
An Intonational Lexicon?: A Very Brief Look

- Attempts have been made to explain intonational variation in questions by looking at the issue in a broader discourse framework (Pierrehumbert and Hirschberg, 1990; Bolinger, 1989).

- In various forms: high boundary tones imply hearer responsibility for the current proposition. Low boundary tones imply speaker responsibility (Steedman, 2000). These can trigger different update rules in the common ground (c.f. Gunlogson (2002))

- Nilsenova (2006): rising intonation is an intonational adverb expressing uncertainty. Questioning and question bias are by-products of pragmatic principles.

- Reese (2007): $L^*+H H^\%$ can trigger biased interpretations of negative polar questions. This corresponds to metalinguistic negation: such questions express denial or counterevidence to something in the (immediately) preceding discourse.
Liscombe et al. (2006): The presence of a final rise was the most useful cue for a question bearing turn in a student/computer tutor scenario (intensity and timing information also helped).

Intonation has also been found to differentiate uses of affirmative backchannels like okay (Benus et al., 2007).

However, Gravano et al. (2007) finds that contextual conditions seem to override prosodic cues in distinguishing different uses of okay.
What does the final rise/fall tell us about a short question? We at least have a bunch of predictions:

- Elided wh-questions should retain final falls.
- Genuine yes/no questions should have a final rise.
- *Backchannel* backchannel questions are highly biased, so we would expect them to fall.
- Backchannel questions that have a final rise may convey something more than acknowledgement – this is most likely speaker uncertainty of the previous utterance.
Final Rises/Falls in Short Questions

- 315 questions turns containing two or less words were located from the Conversational Telephone Speech component of the MDE RT-04 corpus (LDC2005S16): 40 hours of Switchboard-1 Corpus Release 2.

- The MDE annotation (LDC2005T24) provides a range of discourse metadata including question and backchannel type turns.

- F0 contours were extracted from final words using praat, normalized to a log scale (0-10), outliers were removed.

- The pitch slope for each final word was fitted from the normalized F0 data using the linear regression function `lm` in R.

- 307 questions were left after normalizations, lack of speaker info etc.
Results: Question and rise type

Figure: Proportions of samples with negative and positive slopes in different question types: Confirmation, Suggestion, Incomplete, Alternative, Reclamatory, Tag, Complementary, Wh, Backchannel, Y/N
Observations

- Broad categories of questions do appear to have different final rise characteristics.
- No significant differences for the question data as a whole between positive and negative slope counts across gender, age (10 year groups) and accent. (Pearson’s $\chi^2$ test)
- No significant difference in age or sex was found when looking at specific question (Fisher exact test)
- However, there was a significant different in comparing accent type and reclamatory questions (p-value = 0.03).
Final Rises and Backchannel Questions

- It is not clear what the presence of a final rise means for disambiguating the different uses of backchannel-like questions. This ranges from acknowledgement to denial and surprise.

![Figure: Two types of really: dashed, speckled lines are surprised, the solid line is a backchannel.](image)
Prosodically Distinguishing $really_b$ and $really_q$

- Expanded the previous data set (MDE 2003 annotations (LDC2004T12) and audio (LDC2004S08))
- Instances of *really* labelled as a backchannel (450) or a question (130) were extracted via MDE annotations.
- Pitch and intensity measurements were done as previous.
- The correlation between raw intensity and pitch at 10ms intervals was also derived ($\text{Corr}$).
- Overall pitch range ($\text{prange}$). Duration ($\text{Dur}$) and relative time position of pitch minimum ($\text{p.min.d}$) and maximum ($\text{p.max.d}$) were also recorded.
- 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.).
PCA 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).

`p1`, pitch slope, did not seem to account for much of the variance in the data!
Principal Components Analysis

Figure: Projection on to the first two dimension of the PCA space.
Significant differences? Pitch range

Figure: Probability densities for the prange feature. 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.).
Significant differences? Correlation conditioned on rise/fall

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

The means for positive p1 samples ("rising") 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)).
Separating Prosodic Cues

The overlapping distributional data above suggests that it unlikely that the prosodic features described above can differentiate $really_q$ and $really_b$.

To further test this hypothesis, two classifiers were built in an attempt to separate the data.

- Linear Discriminant Analysis (LDA) as implemented in R (lda).
- A Support Vector Machine (SVM) classifier with radial basis function kernel (libsvm via R).
Classification results

The 10-fold cross-validation error rates are shown in Table 1 alongside bootstrap estimates bias, standard error, and 95% confidence intervals (1000 bootstrap samples).

<table>
<thead>
<tr>
<th></th>
<th>Error (Std. Error)</th>
<th>Bias</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>0.245 (0.018)</td>
<td>-0.005</td>
<td>(0.21,0.28)</td>
</tr>
<tr>
<td>LDA</td>
<td>0.244 (0.019)</td>
<td>-0.013</td>
<td>(0.21,0.28)</td>
</tr>
<tr>
<td>SVM</td>
<td>0.267 (0.019)</td>
<td>-0.041</td>
<td>(0.23,0.30)</td>
</tr>
</tbody>
</table>

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

- LDA and SVM error rates do not fall outside the 95% confidence interval for the cross-validation error of the baseline.
Discussion

- This supports the hypothesis that these two categories are not separable on the basis of these features.
- 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.
- In any case, the difference between question and backchannel does not quite seem to fall out of theories of question bias and intonation.
What now? A pragmatic account

- Proposal: *really* signals that adding the previous proposition \( P \) requires the speaker to update their current beliefs.

- *Really* is interpreted as a backchannel (i.e. ignorable) when the addressee determines that \( P \) already had a high probability in the speaker’s beliefs.

- This does not necessarily require the hearer to commit anything to the common ground. Nor that the speaker did not believe \( p \) or even that \( p \) was viewed as unlikely.

- We may expect that the magnitude of the update is what is signalled by prosodic factors.

- Agreement backchannels (e.g. *uh-huh*) signals satisfaction of Gricean Quality, while *really* signals Quantity and Relevance?
Conclusion

- Backchannel questions can be used both as an acknowledgement and as real questions. They can also project many shades of meaning in between.

- In that sense, the second part of this paper was an attempt to find out if some element of an intonational lexicon could systematically differentiate $really_q$ and $really_b$ 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.
But for now...

- 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.
- 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.

Thanks!

Especially to: Jiahong Yuan, Mark Liberman, Andrew Clausen and the Penn Phonetics lab gang.
Here’s an elided polar interrogative that is more like a backchannel than a question:

(5) 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?
Really and other backchannels

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:

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

A: Yes.
Really appears to be an even shorter form of an elided polar interrogatives: These also appear as backchannel questions in spontaneous speech. For example:

(7) 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 in biased questions

- *Really* has been linked to question bias in the case of negative polar questions (Romero and Han, 2004).
- 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).
Question Bias

(8) 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)
Check and denial

More generally, if we observe the question \( ?p \)

- INPQs double check \( \neg p \)
- ONPQs double check \( p \) (denial or counterevidence to previous assertions).

Romero (2006) accounts for these two readings via a scope ambiguity between negation and the \textsc{verum} operator where \textsc{verum} is given the same meaning as epistemic \textit{really}. This derives some NPI licensing facts.

\[
\text{VERUM}_\lambda \equiv \text{really}_\lambda \\
\lambda p_{<s,t>} \cdot \lambda w . \forall w' \in \text{Epi}_x(w)[\forall w'' \in \text{Conv}_x(w'')[p \in \text{CG}_{w''}]] = \text{FOR-SURE-CG}_x
\]
(10) Aren’t there vegan restaurants around here?

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

a. ONPQ: \[ \text{Q \ [not \ [\text{VERUM \ [There \ are \ vegan \ restaurants around here]]]} \]
\[= \{\text{FOR-SURE-CG}_x(\text{There \ are \ vegan \ restaurants around here}), \neg \text{FOR-SURE-CG}_x(\text{There \ are \ vegan \ restaurants around here}) \} \]

Romero and Han (2004) argue that \textit{really} (VERUM) carries a negative epistemic implicature when it appears in positive polar questions.
Reese and Asher: Gluing the discourse together

- Reese (2007): Prosody does distinguish ONPQs from INPQs.
- ONPQs = metalinguistic negation: such questions express denial or counterevidence to something in the (immediately) preceding discourse.
- This can triggered by the presence of an L*+H H% nuclear pitch accent (the ‘contradiction contour’ (Pierrehumbert and Hirschberg, 1990)).
- What about other prosodic features?
INPQs/ONPQs

Really can precede both INPQs and ONPQs in certain contexts:

(11) 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:

(12) A: So much choice for the non-cow eater! Let’s eat!
    B: (#Oh really?) Aren’t there some vegan restaurants around here?

*Really* appears to be more about signalling the speakers surprise than asking the listener to check something into the common ground.
Getting inside their heads!

- There are at least some clear cases where the addressee would have good evidence that the speaker requires further evidence or elaboration.
- e.g. When *really* is a response to the answer to an apparently genuine information seeking question.

(13) 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

- 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. An example is shown below.

(14) 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?

That is, more updating and more evidence is required than a simple yes or no answer.


Haan, J. (2002). *Speaking of Questions: An Exploration of Dutch


