Prosody as the main cue to differentiate wh-in-situ questions

An Nguyen[1], William Howe, Geraldine Legendre
Johns Hopkins University, Cognitive Science Department

Prosody plays an important role in speech perception and comprehension. Previous studies have shown that children can rely on prosody to resolve ambiguous prepositional-phrase attachments (Snedeker, 2008) as well as ambiguous structures due to noun/verb homophones (Carvalho et al., 2016). In this study, we conduct an experiment to investigate if children as young as 4 years of age can rely on prosody to differentiate and correctly infer the meaning of two types of wh-in-situ questions that are structurally similar: probing questions (PQs) versus echo questions (EQs).

Example 1. a. (PQ) A: That's a what? B: A laptop. b. (EQ) A: That's a computer. B: That's a WHAT? A: A computer.

Both PQs and EQs are present in child-directed speech, with PQs being slightly more dominant than EQs in full sentential structures (65% versus 35%, respectively). EQs are typically used to ask for a repetition or clarification of a previous utterance (Authier, 1993), while PQs are used to ask for new information, similarly to the regular fronted wh-questions (Nguyen and Legendre, 2020). On the surface, both types of questions can have the same structure, but they are different prosodically. Specifically, EQs have a L+H* intonation with a HH% boundary tone (Artstein, 2002). In contrast, PQs have a falling pitch accent and the wh-word does not receive stress. An acoustical analysis of 50 PQs and EQs shows that they are significantly different from each other in duration and pitch contour (Figure 1 & Table 1).

Figure 1. Prosody of EQs and PQs.

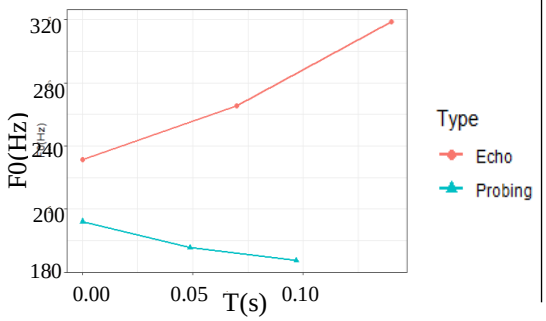


Table 1. Duration & pitch contour of EQs and PQs

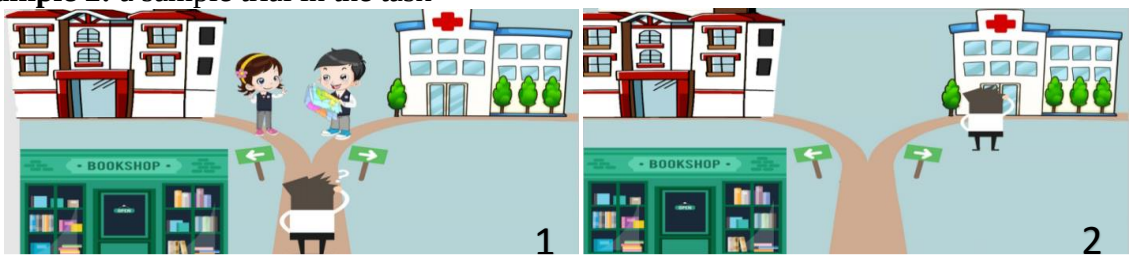
	Echo	**Probing**	
	Mean	**Mean**	p-value
Duration (ms)	0.14	0.097	< **.001** ***
$\Delta F0_{offset-onset}$ (Hz)	93.74	-24.52	< **.001** ***

Twenty native English-speaking children (M_{age} =4.01, age range = 3;06-5;06) participated in a comprehension experiment. Twelve adults were recruited to serve as a control group. The task contained 12 trials, including two practice trials not included in the data analysis. In each trial, children saw a story unfolded on a screen while hearing a character made a true but under-informative statement about the story (Example 2). At the end of the story, they heard either a PQ or EQ (both have the same surface structure but differ prosodically). Half of the scenarios involved PQs and half involved EQs. The questions were asked directly by the researcher playing the storyteller role to ensure the attention and interaction with children. To ensure that the prosody is consistent in every trial and experiment, we conducted a post-hoc acoustical

[1] Corresponding author: an.nguyen@jhu.edu

analysis to exclude any trials in which the prosody of the question is significantly different from the rest of the sample. In total, we have 168 data points for children and 120 for adults.

Example 2: a sample trial in the task



Participants see the man walk into the hospital while hearing "the man went to the white building", although there are two white buildings in the story (the hospital and the apartment). In a PQ trial, the experimenter asks: "the man went where?" and the target answer is "the hospital" as it is the most specific, while "the white building" is a non-target answer as it is not wrong but not informative enough. In an EQ trial, the experimenter asks: "the man went WHERE?" and the target answer is "the white building". The wrong answer for both questions would be "the book store" or "the apartment".

We expect that participants will generally prefer to be informative and precise (Grice, 1975) and thus would give the target answer (based on the information they see on the screen) over the under-informative answer (which they hear from a character in the story) in a PQ trial. At the same time, in an EQ trial, participants should know that the storyteller is interested in what the character has said instead of their own opinion, thus they can only repeat the character's answer even though they may perceive it as being under-informative. Results show that indeed, children gave new-information answers to PQs and repetition answers to EQs with an accuracy significantly above chance level (M = 61.5%, t(147)=2.86, p=0.004). None of the participants consistently produced only one type of answer to all the questions throughout the experiment. In other words, every child used both types of answers (echo-appropriate and probing-appropriate) at least once.

Table 2. Distribution of the answers by category

	Target	Non-target	Wrong
Children	61.5%	33.7%	4.8%
Adults	91.7%	8.3%	0%

A target answer means the right information to the right type of question (e.g. repetition to EQs). A non-target answer means the right information to the wrong type of questions (e.g. repetition to PQs). A wrong answer means the wrong information.

If children assumed that PQs and EQs have a similar goal, or if children ignored the prosody cues and consistently interpreted wh-in-situ questions as either PQs or EQs, we would expect the percentage of target answers to be the same as non-target answers. The fact that they provided almost twice as many target (61.5%) as non-target (33.7%) suggests that children as young as 3;06 years old are able to use prosody to aid comprehension and assign the right intention to the right type of question with moderate accuracy. However, children overall performance was worse than adults, suggesting that children may not be as sensitive to the prosody of questions as adults are. This is in line with previous studies, which claim that although children are able to use prosodic information in sentence processing, they use such information less effectively than adults do to infer the intended meaning (Snedeker, 2008).