

# Acceptability of variations in question intonation in natural and synthesized American English

Ann K. Syrdal  
AT&T Labs – Research  
and  
Matthias Jilka  
University of Stuttgart

## BACKGROUND

There are Two Ways of Asking Questions in American English

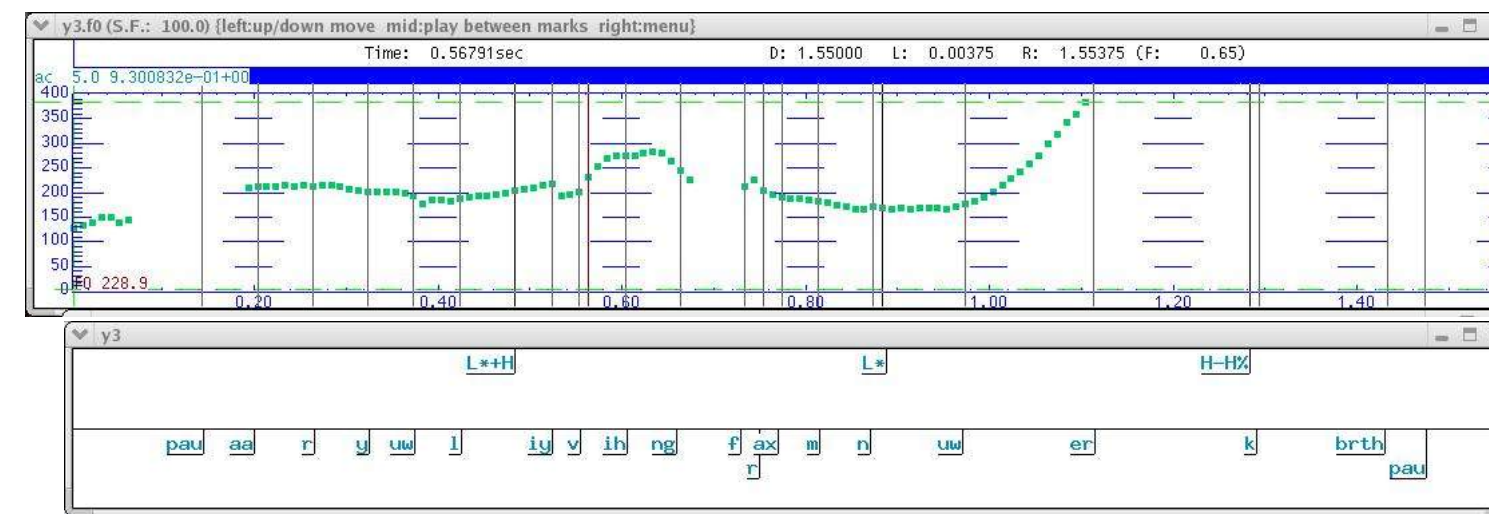
- Yes/No Questions
  - Answerable by *Yes* or *No*
  - Example: *Do you want to play a game?*
- Wh-Questions
  - Who/what/where/when/why/how interrogatives
  - Example: *What game do you want to play?*

The Two Question Types are Different Intonationally

- Intonation is the tune or melodic pattern of the human voice as its pitch rises and falls.
- Expected question intonation patterns
  - Yes/No Questions have phrase-final rises.
    - High pitched phrasal tones labeled **H-H%**
  - Wh-Questions have phrase-final falls and intonationally are like Declarative sentences.
    - Low pitched phrasal tones labeled **L-L%**

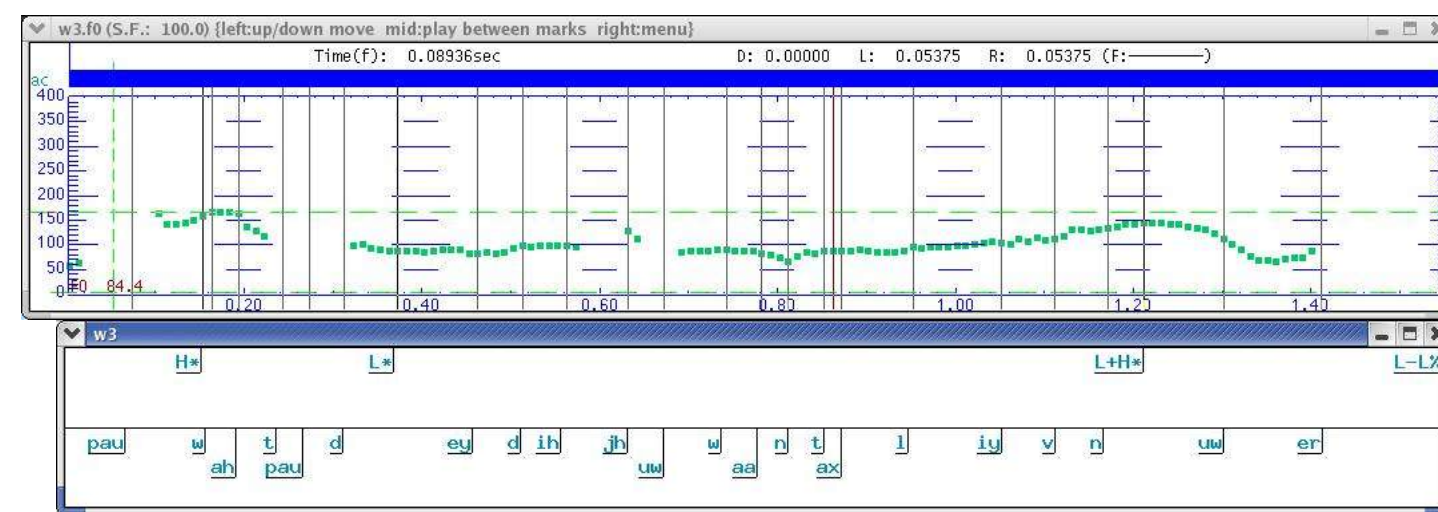
## Example Yes/No Contour: H-H%

“Are you leaving from Newark?”



## Example Wh-Q Contour: L-L%

“What day did you want to leave Newark?”



## Part 1: Acoustic Study

- Recordings of 8 speakers of American English
  - 5 female speakers, 3 male speakers
- All speakers are professional voice talents
- Each speaker read the transcript of an actual dialogue between travel agent and customer
- Acoustic Analysis
  - $f_0$  contours were calculated automatically using a robust pitch-tracking algorithm
- Linguistic Analysis
  - ToBI (Tones and Break Indices) labeling by linguist (MJ) of Pitch Accents and Phrasal Accents for 89 selected utterances

## Study Sentences were Extracted from Recorded Dialogue Readings

- Y1. *Did you want to fly to London?*  
W1. *When did/do you need to be in London?*  
D1. *That will have a two-hour layover in London.*  
Y2. *Are you interested in a(n) earlier/later flight?*  
W2. *What's the month and date for your return flight?*  
D2. *That is a non-stop flight.*  
Y3. *Are you leaving from Newark?*  
W3. *What day did you want to leave Newark?*  
D3. *Okay, you're leaving from Newark.*

Female Speakers exhibited Expected Phrasal Tones more often than Male Speakers

| Phrasal Tone     | Females | Males | All  |
|------------------|---------|-------|------|
| Y/N-Q H-H%       | 83%     | 53%   | 70%  |
| Wh-Q L-L%        | 100%    | 79%   | 91%  |
| Declarative L-L% | 100%    | 100%  | 100% |
| Overall          | 94%     | 74%   | 85%  |

## Phrase-Final Tunes:

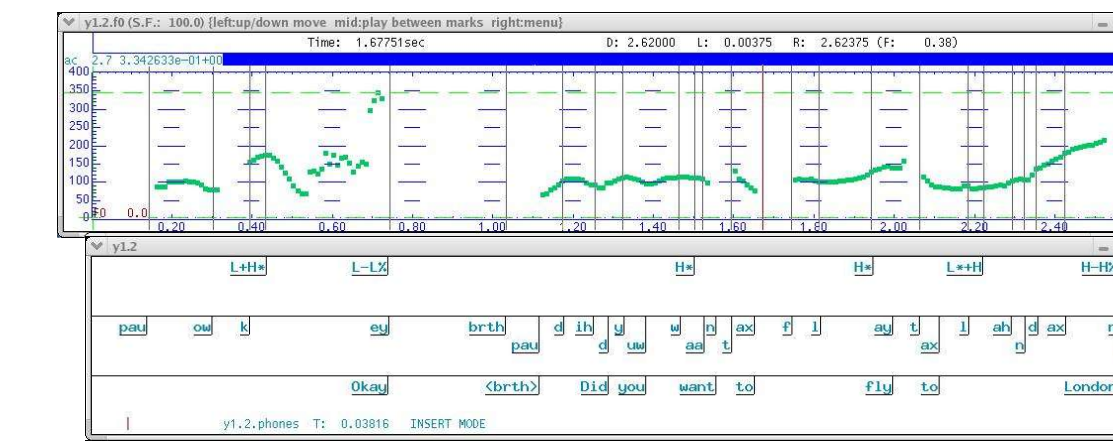
Nuclear Accents for H-H% Yes/No Questions

| Accent | Percent |
|--------|---------|
| L*     | 61%     |
| L*+H   | 17%     |
| L+H*   | 17%     |
| H*     | 4%      |

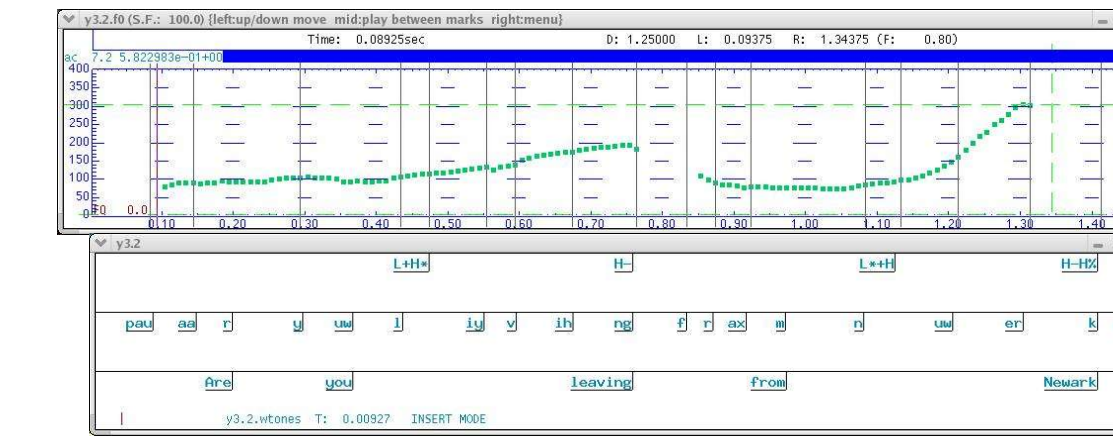
- Nuclear Accent (NA) is final Pitch Accent in an utterance.
- Contrastive Stress: L\* and L\*+H NAs focus salience of question on accented word. Example: *Are you leaving from Newark?* Vs. *Are you leaving from Newark?*
- Incredulity is conveyed by L\* H-H% or L\*+H H-H% contour with expanded pitch range and faster rate of rise.

Comparison of Yes/No Contours with L\*+H H-H%:

Regular Yes/No Question : rise from 80 to 210 Hz in 330 ms



Incredulous Yes/No Question : rise from 70 to 300 Hz in 260 ms



## Wh-Question and Declarative Intonation Compared

- Wh-Q phrase-initial pitch accents were more prominent (36-41 Hz higher on average) than those in Declarative sentences
- Both Wh-Questions and Declaratives normally ended in “final fall” L-L% phrase tones.

## Part 2: Listener Judgments of Question Acceptability

- Forced choice preference judgments were made by 12 listeners
- Test stimuli were 29 same-voice same-question but intonationally different pairs
  - 17 question pairs were recordings of natural speech (11 male and 6 female pairs)
    - 9 Yes/No Questions and 8 Wh-Questions
  - 12 question pairs were synthetic speech (6 male and 6 female pairs)
    - 6 Yes/No Questions and 6 Wh-Questions

## Listener Preferences for Yes/No Question Intonation

- Intonation Preferences for Recorded Natural Speech Consistent with Frequency of Usage
  - L\* H-H% 67% versus [!]<sup>1</sup>H\* L-L% 33%
  - L\* H-H% 79% versus L\* L-H% 21%
  - L\*+H/L+H\* H-H% 46% versus \* L-L% 54%
- Intonation Patterns were Generally Unrelated to Listener Preferences for Synthetic Questions
  - Preferences reflected overall voice quality
  - Preferred synthetic utterance with fewer abnormalities

## Listener Preferences for Wh-Question Intonation

- Intonation Preferences for Recorded Natural Speech Consistent with Frequency of Usage
  - L-L% 75% versus H-H% 25%
  - L-L% 54% versus L+H% 46%
  - Initial PA on interrogative pronoun (*what*) 64% versus on another word (*day*) 36%
- Listener Preferences for Synthetic Questions Generally Reflected Overall Quality
  - Unusual tonal patterns were not preferred, however

## SUMMARY

- Expected phrase-final question intonation patterns were the most common patterns spoken for both Y/N and Wh-question types
  - Female speakers exhibited expected intonation patterns more often than male speakers
  - More intonational variation observed in Y/N Questions
- Listeners usually preferred expected intonation patterns in natural speech questions
- Preference for synthesized questions was based more on overall voice quality than on intonation