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

Ann K. Syrdal
AT&T Labs – Research
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
  - Wh-what-when-what-why-how interrogatives
  - Example: What game do you want to play?

Part 1: Acoustic Study

- 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 phrase was labeled H-H%
    - Wh-Questions have phrase-final falls and intonationally are like Declarative sentences.
      - Low pitched phrase was labeled L-L%

Phrase-Final Tones: Nuclear Accents for Y/N vs. No Questions

- Nuclear Accent (NA) is final Pitch Accent in an utterance.
- Contrastive Stress: L* and L+H* NA; 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 L-L% contour with expanded pitch range and faster rise rate.

Phrasal Tones: Females

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

Female Speakers exhibited Expected Phrasal Tones more often than Male Speakers

Wh-Question and Declarative Intonation Compared

- Wh-Question intonation patterns 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
    - 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 (5 male and 6 female pairs)
  - 6 Yes/No Questions and 6 Wh-Questions

Listener Preferences for Synthesized Questions

- Preferences reflected overall voice quality
  - Preferred synthetic utterance with fewer abnormalities

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

Study Sentences were Extracted from Recorded Dialogue Readings

Y1. Did you want to fly to London?
W1. When did you need to be in London?
D1. That will have a two-hour layover in London.
Y2. Are you interested in any other earlier/earlier flight?
W2. What’s the weather and data for your future 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.

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?”

Listener Preferences for Yes/No Question Intonation

- Intonation Preferences for Recorded Natural Speech Consistent with Frequency of Usage
  - L-H H-H% 67% versus [L+H] L-L% 33%
  - L* H-H% 79% versus L* L-L% 21%
  - L*+H* 16-P 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

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

Regular Yes/No Questions: rise from 90 to 230 Hz to 330 ms

Examples:

- What day did you want to leave Newark?
  - Incredulous Yes/No Question: rise from 70 to 300 Hz to 260 ms

Test stimuli were 29 same-voice same-question
- 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 (5 male and 6 female pairs)
- 6 Yes/No Questions and 6 Wh-Questions

Listener Preferences for Wh-Question Intonation

- Intonation Preferences for Recorded Natural Speech Consistent with Frequency of Usage
  - L-5 75% versus H-H% 25%
  - L-L% 54% versus L-H% 46%
  - Initial PA on interrogative pronoun (who) 64%
    versus on another word (date) 36%
- Listeners Prefer Synthetic Questions Generally Reflected Overall Quality
  - Unusual tonal patterns were not preferred, however