Effects of Voice Quality on Listening Comprehension

Sidney Ma and Kai Jeffers

Background

- Dysphonic voices interfere with cognitive performance at various levels.
- When attending to dysphonic voices, listeners struggle to:
  - Recognize individual phonemes (Schiller et al. 2021)
  - Understand entire sentences (Ishikawa et al. 2017)
  - Follow general tasks (de Silva et al. 2020)
- Creaky voice is often a symptom/effect of dysphonia.
  - Creaky voice might cost more *cognitive load* than modal voice (Imhof et al. 2014)
  - Creaky voice signals *mild affective states* as opposed to strong emotions (Gobl et al. 2003)
  - Speech recognition AI struggles to understand creaky voice (Kane et al. 2013)

Background - cont.

- Most research investigates the effect of *dysphonic voices* on *children's classroom performance*.
- Little research focuses specifically on *voice quality* and its effects on listening comprehension, especially for adults.
- Dysphonia has a variety of different symptoms beyond just "creaky voice."
- Will previous results replicate in an lab setting for adult listeners, specifically focusing on voice quality?

Main Question & Predictions

How does voice quality affect listening comprehension?

- Voice quality = *modal / creaky / glottal*
- Listening comprehension = *correctness and reaction time of responses*

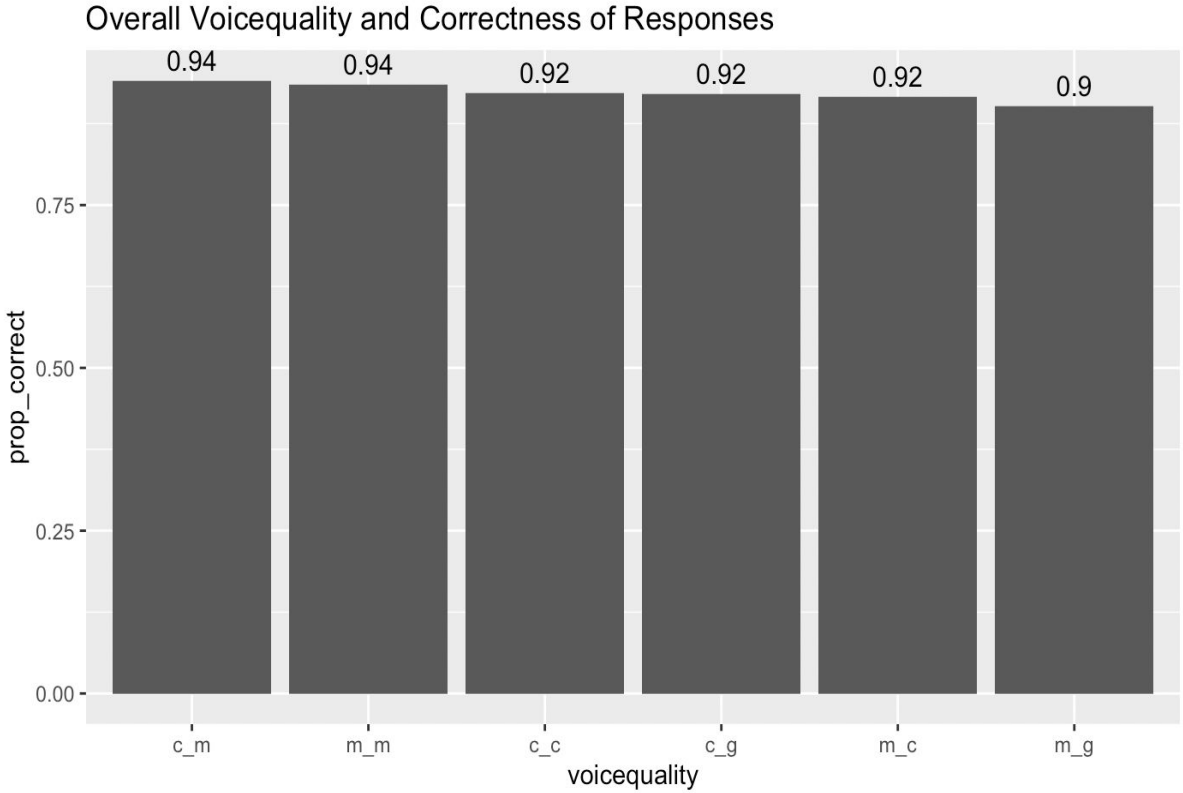
Predictions:

- **Creaky and glottal voice → worse comprehension.**
  - **Correctness of responses should be lower, and reaction times should be slower.**
  - Because creaky and glottal voice resemble dysphonic voice (harder to comprehend).
- This effect should be larger for the *target word*, since it should negatively affect phonological comprehension, and therefore, word recognition.
- No expected difference between creaky and glottal voice – both “sound dysphonic.”

Methodology

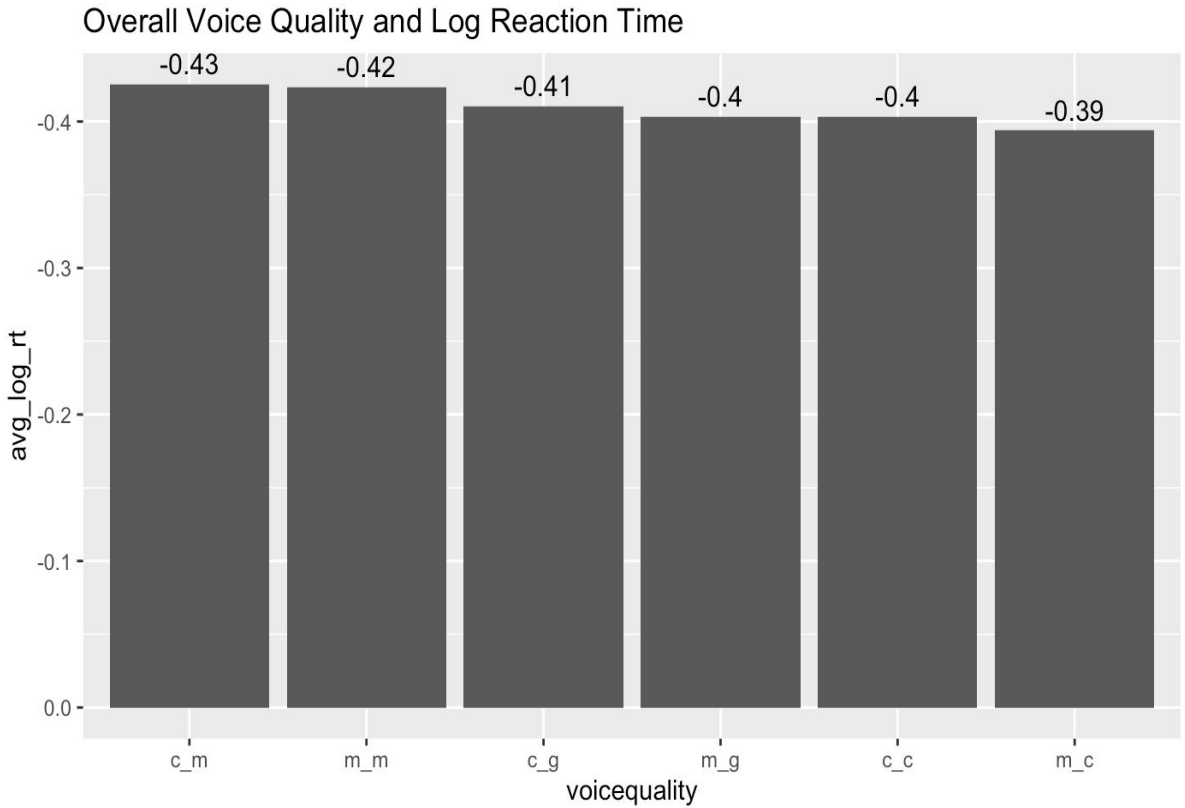
- Collect data from Maxine’s study
- Include filler trials in dataset, no attention trials
- Save information about correctness and reaction time for each trial
- Analyze trends between voice quality and comprehension
    - Baseline voice quality (VQ1), target voice quality (VQ2), and overall voice quality (interaction)

Main Results (1/4) - Overall VQ and Correctness



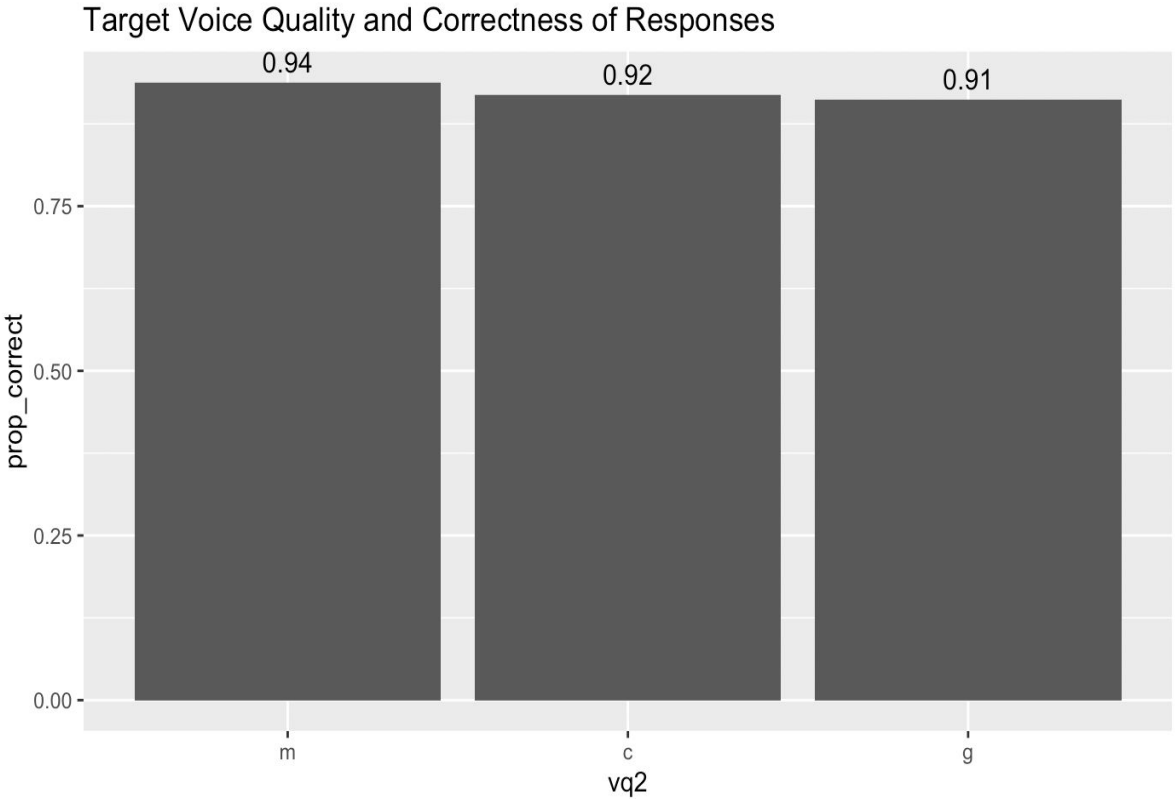
- With m_m as default, **significant effects of all levels except c_m**
- Correctness always very high
- Highest when VQ2 is modal
- Other than that, no obvious patterns
- Somewhat aligns with predictions

Main Results (2/4) - Overall VQ and Reaction Time



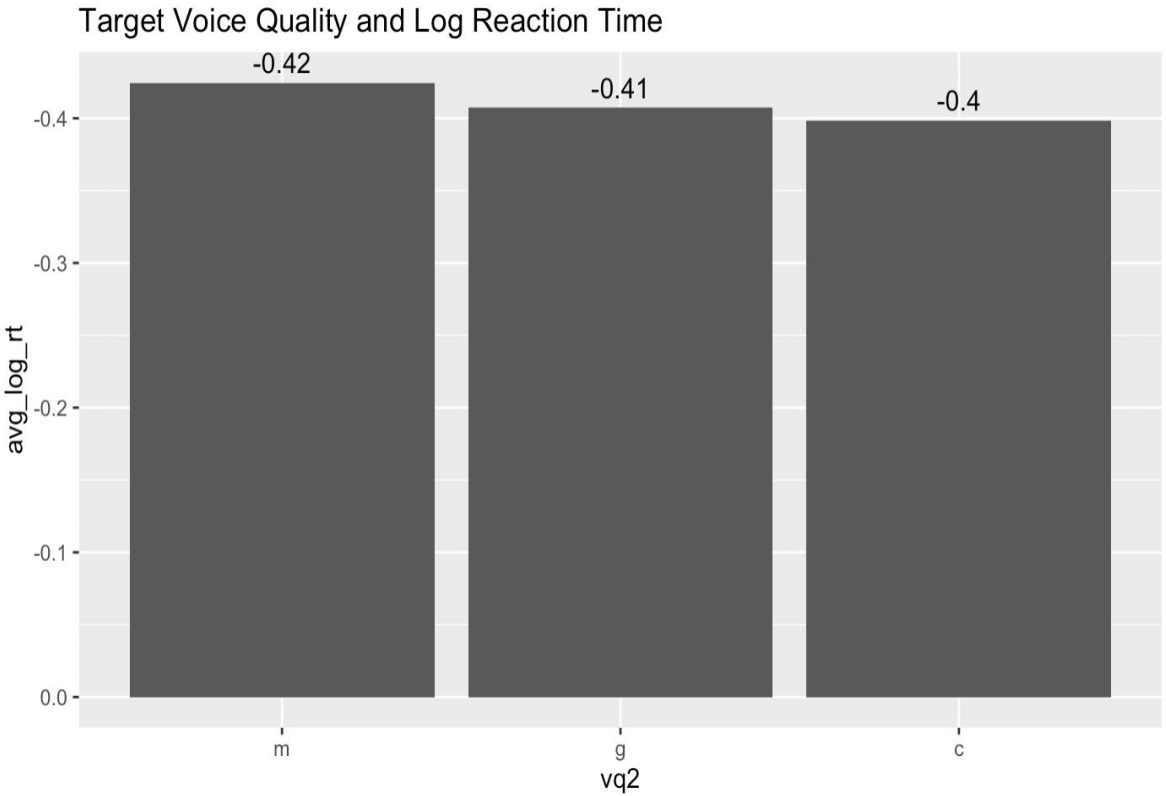
- **No significant effect on reaction time**
- Still, noticeable pattern
- Fastest when VQ2 is modal
- Fastest when VQ1 is creaky (surprising)

Main Results (3/4) - Target VQ and Correctness



- With modal as default, **significant effect of both creaky and glottal voice on correctness**
- **No significant difference between creaky and glottal**
- Aligns with predictions perfectly!

Main Results (4/4) - Target VQ and Reaction Time



- With modal as default, **significant effect of both creaky and glottal voice on reaction time**
- **No significant difference between creaky and glottal**
- Aligns with predictions perfectly!
- Surprising that glottal is better than creaky...

Discussion of Main Results

- For the most part, our results support our predictions:
  - Creaky and glottal voice hinder listening comprehension
  - Effect was larger for target voice quality
  - No significant difference between creaky and glottal
- There were some unexpected results:
  - No significant effect of baseline voice quality – “spotlight” effect?
  - Not a clear meaningful interaction between baseline and target voice quality
- Limitations
  - Effect is largely due to target sentences (a minority). Hard to know if this effect applies to all kinds of words/sentences or just vowel-initial words.
  - Answers are very easy and sometimes predictable due to repetition – very few wrong answers, even when reaction time is negative.
  - Can confirm THAT comprehension is worse for creaky and glottal voice, but not WHY.

Bonus Findings and Directions for Future Research

- **Observations of worst-scoring files**
  - Any reason why some files are harder to understand than others?
  - Issues with file, or can be explained phonetically?
- **Possible interactions between baseline and target voice quality**
  - Is there any pattern in the overall voice quality's effect on comprehension?
- **Possible effect of speaker (and maybe, gender?)**
  - Are some speakers easier to understand than others?
  - Is one gender easier to understand?
  - Does this depend on voice quality?

Which files have the worst comprehension? Why?

All files with < 80% correctness include:

- Words from target trials (w-initial words, e.g. wake, wage, weight)
  - Hard to distinguish between vowel-initial and w-initial words (this is intentional!)
- “I heat” files
  - When the vowel is lengthened, it suggests that the coda should be voiced [d] – “heed.”
- “My dad” and “no gate” files
  - When the onset [d] is pitched down, sounds more like [b] – “bad.”
  - When the coda [t] is pitched down, it sounds more like [k] – “gake.”
- “My mold” and “my pod” files
  - Difficult because the [d] isn’t audibly released – hard to distinguish from the competitor.

Possible directions for future research:

- Speech normalization with respect to the pitch of consonants
- The role of vowels in consonant-based minimal pairs (heat vs. heed, pod vs. pog)

Any interactions between VQ1 and VQ2?

- Comprehension is best when VQ2 is modal (both in correctness and reaction time).
  - Easy to understand when the target is clear
- m_g has the lowest correctness
  - Sudden glottal voice quality might falsely indicate a glottal onset
- m_c has the slowest reaction time
  - Similarly, might indicate a glottal onset
- c_g has surprisingly high comprehension (by both metrics)
  - Maybe listeners aren't affected by the sudden glottal voice when they are already normalized to a creaky voice baseline?

Any effect of speaker (or gender)?

Correctness of answers varies (slightly) between speakers:

- AT files: 95% correct
- BL files: 94% correct
- SC files: 84% correct

(No statistical tests yet, not enough data anyway)

Possible directions for future research:

- Do individual or gender-based differences in speakers (or listeners) affect comprehension?
- Does this interact with voice quality?

Future Research

- Replicate our study with sentences of appropriate difficulty
- Replicate our study focusing on other acoustic effects of dysphonia
- Speech normalization with respect to the pitch of consonants
- The role of vowels in consonant-based minimal pairs
- Effect of individual or gender-based differences in speakers (or listeners) on comprehension
- Interaction between gender and voice quality