The Impact of Face Masks on Spectral Acoustics of Speech: Effect of Clear and Loud Speech Styles

Thea Knowles, Gursharan Badh

Communicative Disorders & Sciences, University at Buffalo

University at Buffalo

Background

Face Mask Mandates: Due to the COVID-19 pandemic, the United States Center for Disease Control (CDC) has recommended individuals to wear face masks to prevent the spread of viral particles and reduce disease transmission[2]. Surgical Masks and KN95 Masks are widely available, both disposable masks that meet a medical standard and have been recommended by the CDC for use by the general public.

The Effect on Speech: Face masks have been shown to act as a low-pass filter on speech, acting as a barrier to the acoustic signal. Many types of masks have been shown to attenuate acoustic energy at frequencies greater than 1000 Hz[3,13].

Modifying Speech Style: Speaking more clearly or loudly may be one way to overcome the effects of masks on speech. Both styles are produced with greater speech intensity, relative to habitual speech[14] and are associated with an increase in energy in the higher frequency ranges of speech. This leads to a flatter (less negative) spectral slope and greater relative energy in the first formant range[4,14]. Clear speech has been associated with an increase in energy in mid-range frequencies [5, 6, 7, 10, 11, 14].

Purpose

Quantify the effect of face masks and clear and loud speaking styles on spectral acoustics of speech.

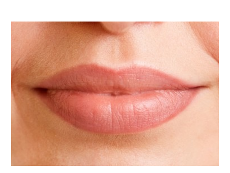
Research Questions

1. What is the impact of face masks on spectral acoustics of speech in unaltered (habitual) speech?
2. What is the relationship between face masks and altered speech styles (clear and loud) on spectral acoustics of speech?

Methods

Participants & Experiment

Mask Conditions



No Mask (nm)

Surgical Mask (sm)

KN95 Mask (kn)

Participants

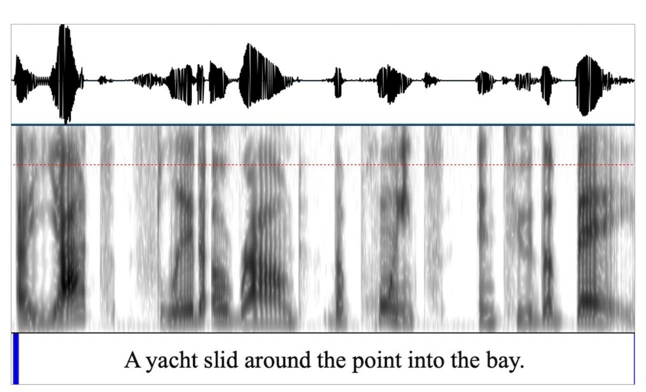
17 healthy adults

- 16 females, 1 male
- Age: 20 – 42 (mean = 26)

Audio Recording

- 6 inches from participant's mouth
- Boom microphone positioned at 6-inch mouth-to-mic distance

Speech Task



Participants read sentences from the Harvard Sentence Corpus[9] in the three mask conditions. These sentence lists and the order of face masks were randomized for each participant.

Speakers began with the habitual condition. Clear and loud conditions were counterbalanced across participants. Instructions given to participants were based on Speech Style:

- **Habitual:** *"Speak in your everyday speaking voice."*
- **Clear**: *"Speak clearly by over-enunciating your speech, similar to how you might speak to someone who is having difficulty hearing you or understanding you."*
- **Loud:** *"Speak at a volume that feels two times louder than your normal speaking voice."*

Analysis

Acoustic Analysis

- **Spectral Moments** (such as center of gravity COG) were selected as they are known to be sensitive to the potential filtering characteristics of the masks and speaking style[1].
- **Mid-range frequency energy: Mean energy in the 1-3 kHz range:** higher amounts of mean energy in the 1-3 kHz range represent increased vocal effort and has been associated with increased intelligibility.[8, 10]
- **Spectral tilt: Difference in energy between 0-1 kHz & 1-10 kHz:** a lower amount of energy in the higher freq. range is captured by a sharper negative spectral tilt which is associated with lower perceived loudness, effort and intelligibility[12].

Extracted from	Acoustic Measurements	Research Question 1	Research Question 2
Utterance	Mean Speech Intensity (Utterance)	✓	✓
The long-term average spectrum (LTAS) of each utterance	Centre of Gravity	✓	—
	Mean energy in the 1-3 kHz range	✓	✓
	Spectral Tilt (difference in energy between 0 to 1 kHz and 1-10 kHz)	✓	✓

Table 1: Acoustic Measurements used for each Research Question

Statistical Analysis

- Modelled acoustic variables as a function of mask condition and speaking style and masks-by-speech style interaction.

Research Question 1 (habitual speech):
`DV ~ Group * Speech Style (…)`

Research Question 2 (all speech conditions):
`DV ~ Group * Speech Style * Mask Condition (…)`

- All models included random by-participant and by-item intercepts. Models for RQ2 included random by-participant slopes for speech style.

Results & Discussion

Acoustic Measures Across all Speakers

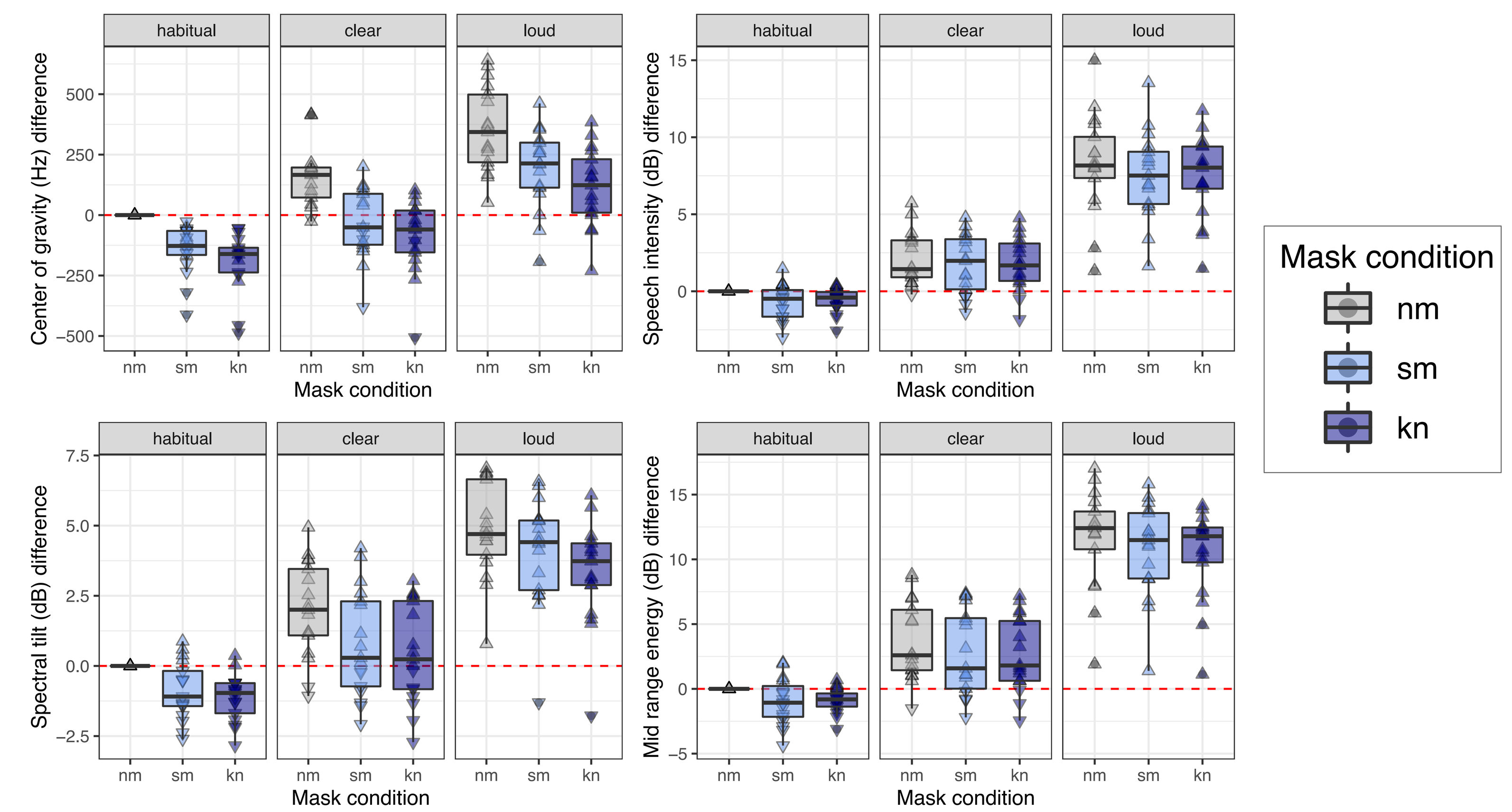


Figure 1: Acoustic measures of interest by speech style (habitual, clear, loud) and mask type (no mask, surgical mask, KN95 mask). Horizontal dashed line reflects individual participants' baseline (habitual speech without a face mask).

Acoustic Measures of Individual Speakers

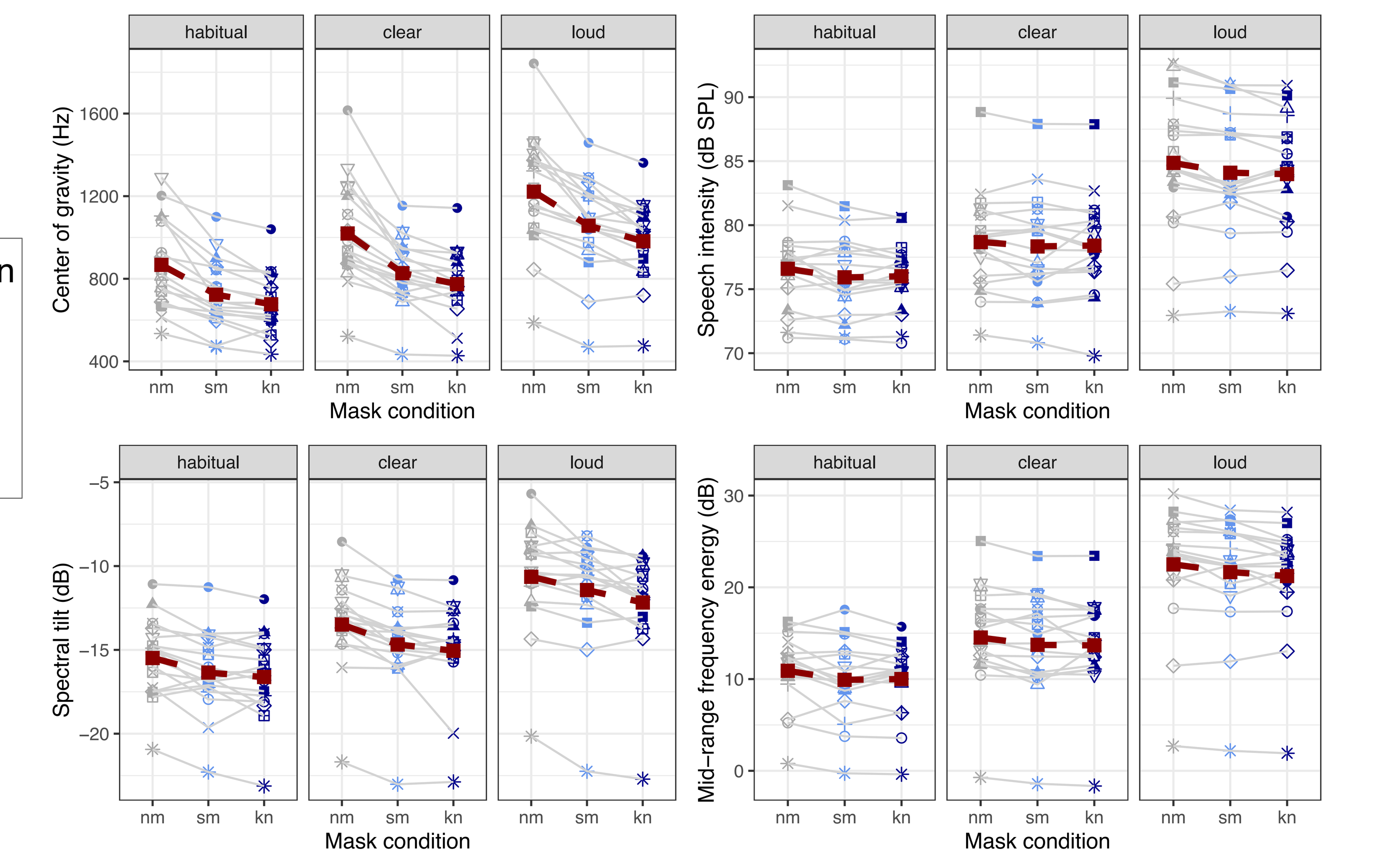


Figure 2: Differences in acoustic measures of interest for each individual speaker compared to baseline (habitual speech without a face mask) by speech style (clear, loud) and mask type (surgical mask, KN95 mask). Red dashed line reflects group mean.

Long-Term Average Spectra

Across Face Mask Conditions

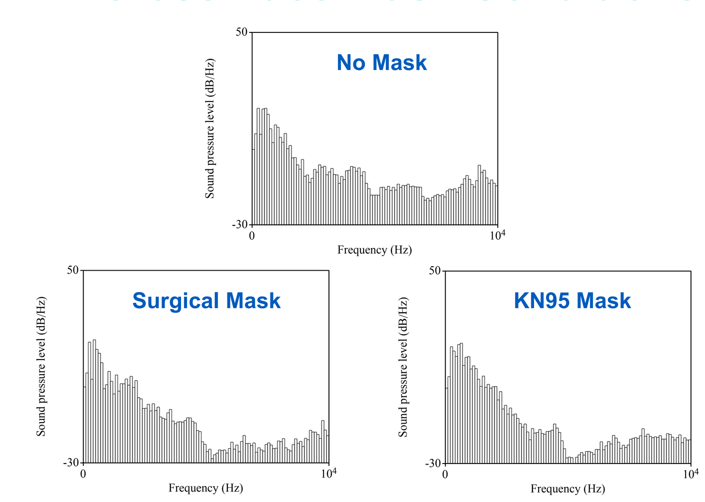


Figure 3: Long-Term Average Spectra across mask conditions while using habitual speech

Across Speech Style Conditions

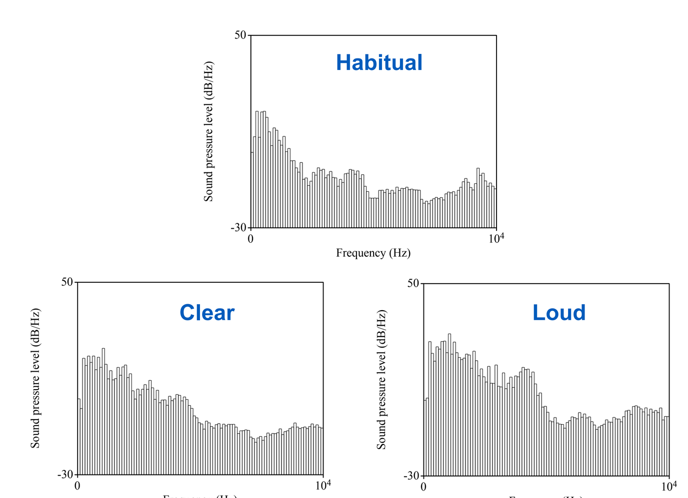


Figure 4: Long-Term Average Spectra across speech style condition while wearing no mask

Research Question 1

Mask vs No Mask: Wearing a face mask resulted in:

- **Lower…**
  - Center of gravity (large effect size), intensity, spectral tilt (medium effect size), mid-range frequency energy (negligible effect size)

Overall the presence of masks demonstrated a systematic, significant effect on all spectral measures compared to not wearing a mask.

Surgical Mask vs KN95 Mask: Wearing a KN95 mask resulted in:

- **Lower spectral tilt** (negligible effect size)
- **No significant differences in the following measures:**
  - Intensity
  - Mid-range frequency energy (1-3 kHz)

Overall there was a greater high-frequency filtering effect of the KN95 mask compared to the surgical mask.

Research Question 2

Habitual Speech vs. Clear and Loud Speech:

- **Intensity:** Clear and loud speech was associated with higher intensity. Large effect size.
- **Mid-range Frequencies (1-3 kHz):** Clear and loud speech was associated with greater mid-range frequency energy. Large effect size.
- **Spectral Tilt:** Clear and loud speech was associated with flatter spectral tilt. Large effect size.

Clear vs. Loud Speech:

- **Intensity:** Intensity was higher in the loud speech condition. Large effect size.
- **Mid-range Frequencies (1-3 kHz):** Loud speech was associated with greater mid-range frequency energy when compared to clear speech. Large effect size.
- **Spectral Tilt:** Loud speech was associated with flatter spectral tilt when compared to clear speech. Large effect size.

The overall effects of the masks were the same in clear and loud speech as in habitual speech. Overall, speech style conditions had a greater effect on acoustics than the mask conditions. Apart from speech intensity, there were no significant mask-by-speech style interactions found for any of the measures. Clear and loud speaking styles were found to be successful in overcoming the filtering effect of masks, though the general relationship between masked and unmasked speech remained intact within these altered speech styles.

Summary

- In habitual speech, findings confirmed a low pass filtering effect of face masks. The KN95 mask showed a (somewhat) greater effect than the surgical mask. This pattern persisted across speech styles.
- Talkers were most successful at overcoming the effects of the masks when speaking more loudly than when speaking more clearly.
- Findings may have implications for talkers with degraded speech acoustics due to disordered speech or voice production. Our lab is currently analyzing similar data (acoustic and perceptual) from speakers with Parkinson's Disease.

SCAN ME