

## Introduction

- Differences in voice fundamental frequency (F0) between two talkers aid in their perceptual segregation ( $\Delta F0$  benefit)
- One proposed mechanism is harmonic cancellation [3]
  - Uses masker periodicity to eliminate masker's neural representation
- Cancellation should break down when target F0 is one octave above masker F0 (hereafter called "Target High octave  $\Delta F0$ ")
  - Target and masker share temporal period so masker cannot be cancelled without cancelling target
  - Brokx and Nootboom [2] found little  $\Delta F0$  benefit in this case
  - However, target and masker also have high degree of spectral overlap (i.e., no "spectral glimpses" available)
- This research aimed to determine whether spectral glimpsing or cancellation better accounts for Brokx and Nootboom [2]
  - Compared speech perception at Target High octave  $\Delta F0$  with and without spectral overlap

## Methods

- Speech reception thresholds (SRTs)** measured via 1-up-1-down procedure [4]
- Experiment 1:**
  - Target** was male talker (IEEE corpus [6]) manipulated by STRAIGHT [5]
  - Masker** was white noise at 70 dB SPL
  - 20 UMN students received course credit or \$10/hour
  - 2 lists per condition, list-condition pairing and order randomized
  - Only some combinations of following variables tested:

Name	Levels	Description
Target F0	Intonated (INT)	Natural contour (90 Hz mean)
	80, 95, 160, 190	Monotone (Hz)
Spectral Structure	All Harm	Target has all harmonics
	Odd Harm	Target has only odd harmonics

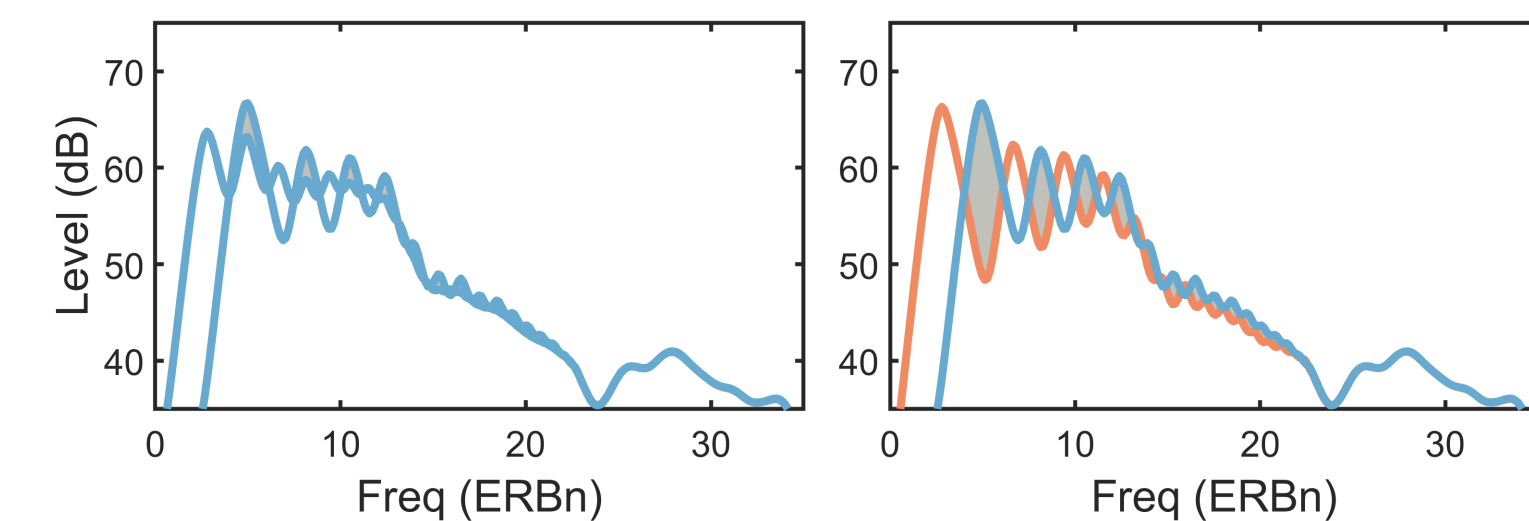
- Experiment 2:**
  - Target** was monotone version of Experiment 1 targets
  - Masker** was random phase harmonic complex tone (HCT) with speech-shaped spectral envelope and monotone F0 at 70 dB SPL
  - 17 UMN students received \$10/hour for participation
  - 2 lists per condition, list-condition pairing and order randomized
  - Fully factorial within-subjects design of following variables:

Name	Levels	Description
$\Delta F0$	0 ST, 3 ST, 12 ST, 15 ST	F0 difference between target and masker
Target Pitch	Target Low Target High	Target F0 = 80 Hz, masker F0 varied Masker F0 = 80 Hz, target F0 varied
Spectral Structure	All Harm	Low pitch sound has all harmonics
	Odd Harm	Low pitch sound has only odd harmonics
Masker Type	HCT Mod HCT	Speech-shaped HCT Speech-shaped HCT with speech temporal envelope

## Hypotheses

- Experiment 1:**
  - H1:** Shifting F0 away from talker's natural range (approx. 90 Hz) will reduce target intelligibility
  - H2:** Removing target even harmonics will reduce target intelligibility
- Experiment 2:**
  - H1:** Minimal  $\Delta F0$  benefit at Target High octave  $\Delta F0$
  - H2:** Removing masker even harmonics at Talker High octave  $\Delta F0$  will improve speech intelligibility
  - H3:** Masker modulation will improve speech intelligibility

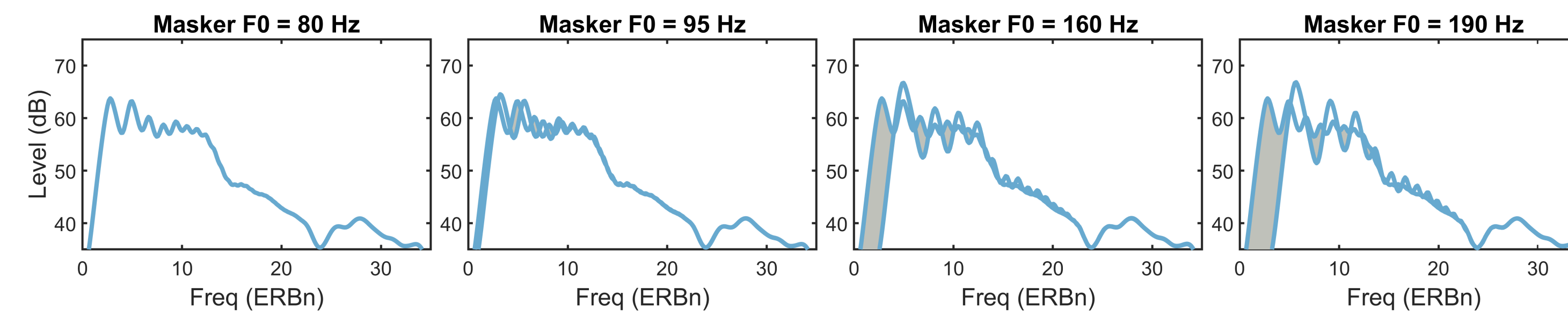
## Odd Harm manipulation creates spectral glimpses at Target High octave $\Delta F0$ (but leaves shared periodicity intact)



**Figure 1:** Average excitation patterns (EPs) showing target with 160 Hz F0 and masker with 80 Hz F0 (**All Harm** in blue on left, **Odd Harm** in orange on right). Spectral glimpsing opportunities for target indicated in gray.

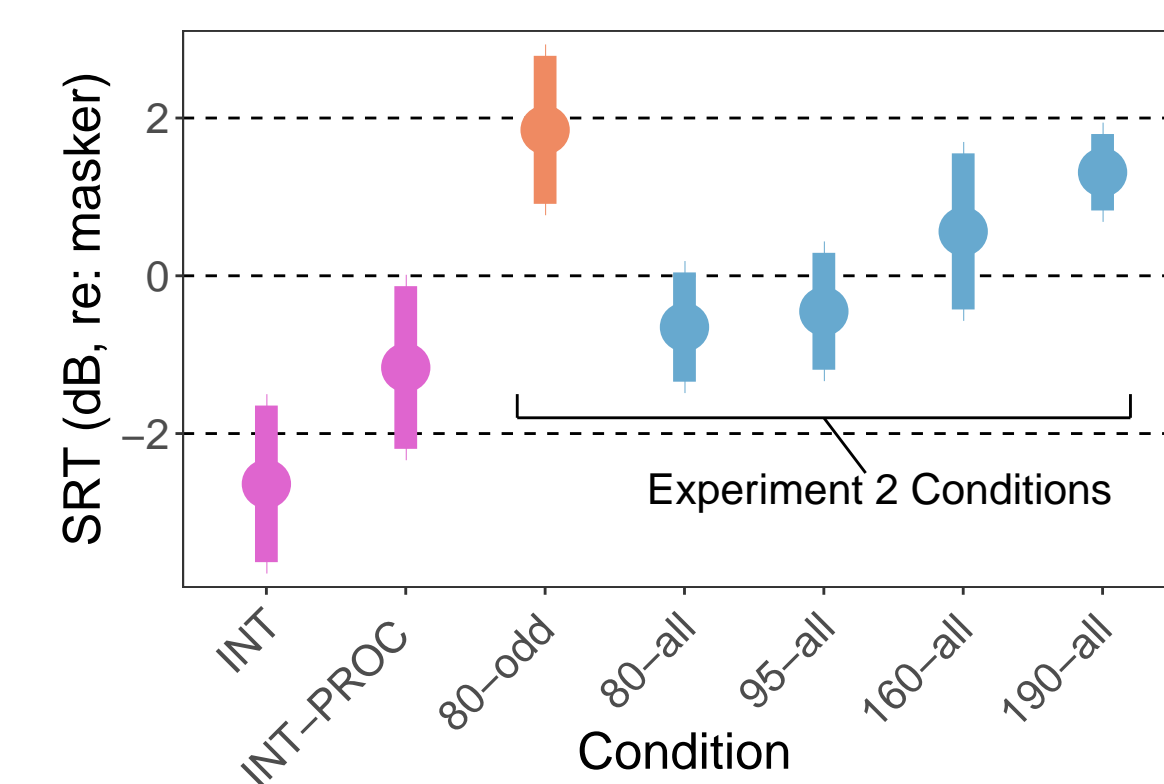
## Increasing masker F0 increases spectral glimpses

**Figure 3:** Average EPs showing **All Harm** target with 80 Hz F0 and **All Harm** masker with F0 increasing from left to right. As masker F0 increases, EP dips between masker partials deepen.



## Results — Experiment 1

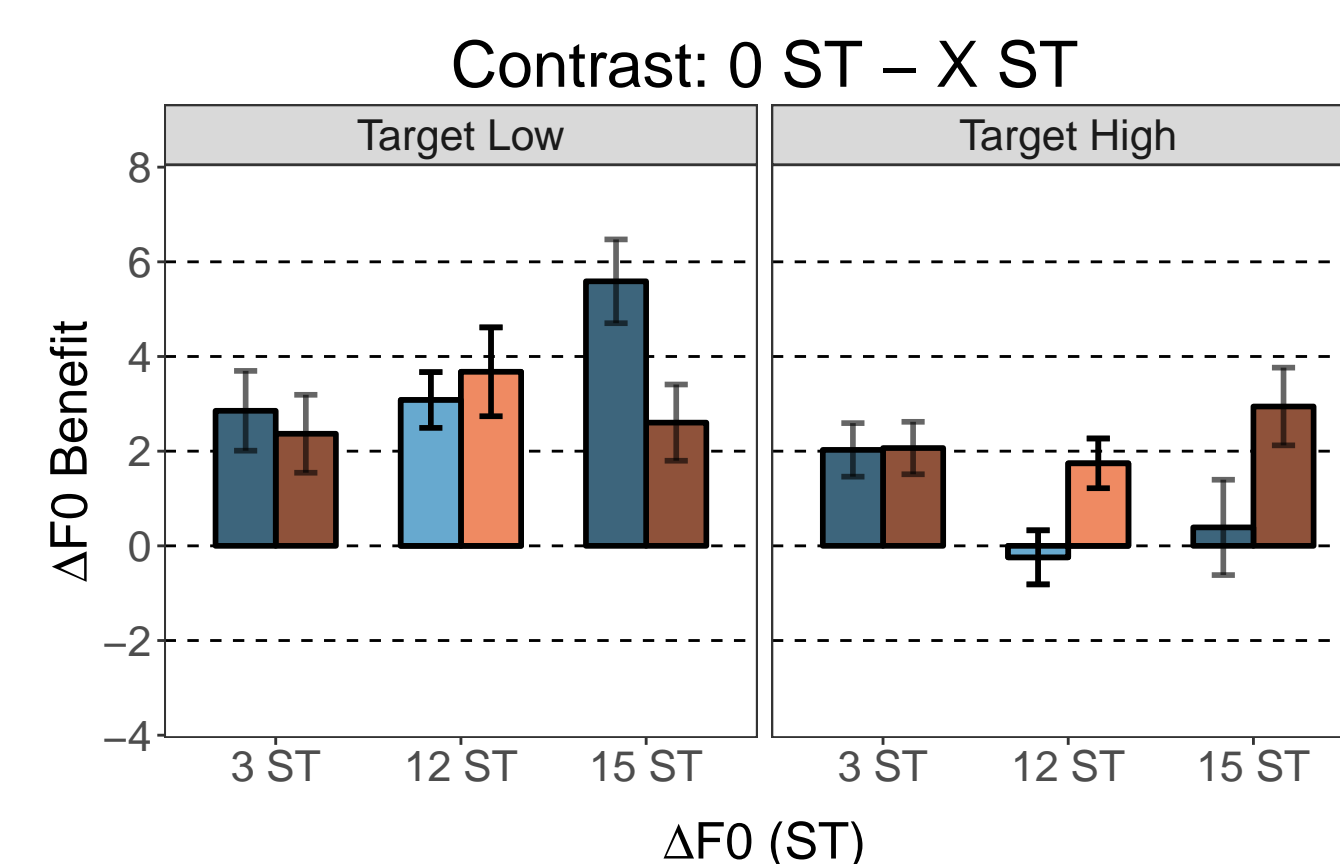
### Stimulus manipulations had small effects on intelligibility.



**Figure 4:** SRTs vs. condition. **INT** is in purple, while monotone **All Harm** is in blue and **Odd Harm** is in orange. Error bars are 95% confidence intervals.

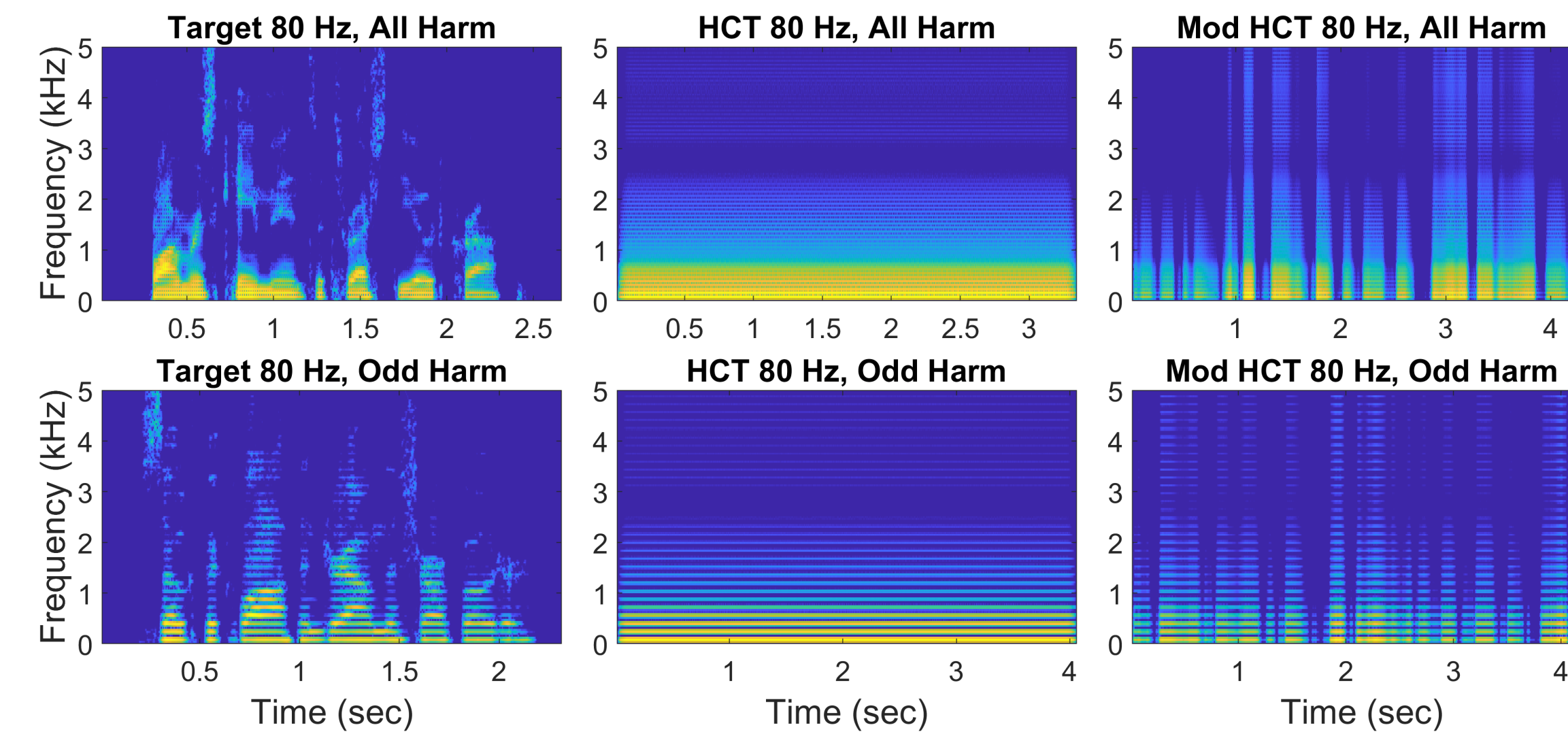
- Mixed-effects model (via `lme4` [1]) revealed significant main effect of condition ( $p < 0.001$ )

### No significant Target High octave $\Delta F0$ benefit, large Target Low octave $\Delta F0$ benefit



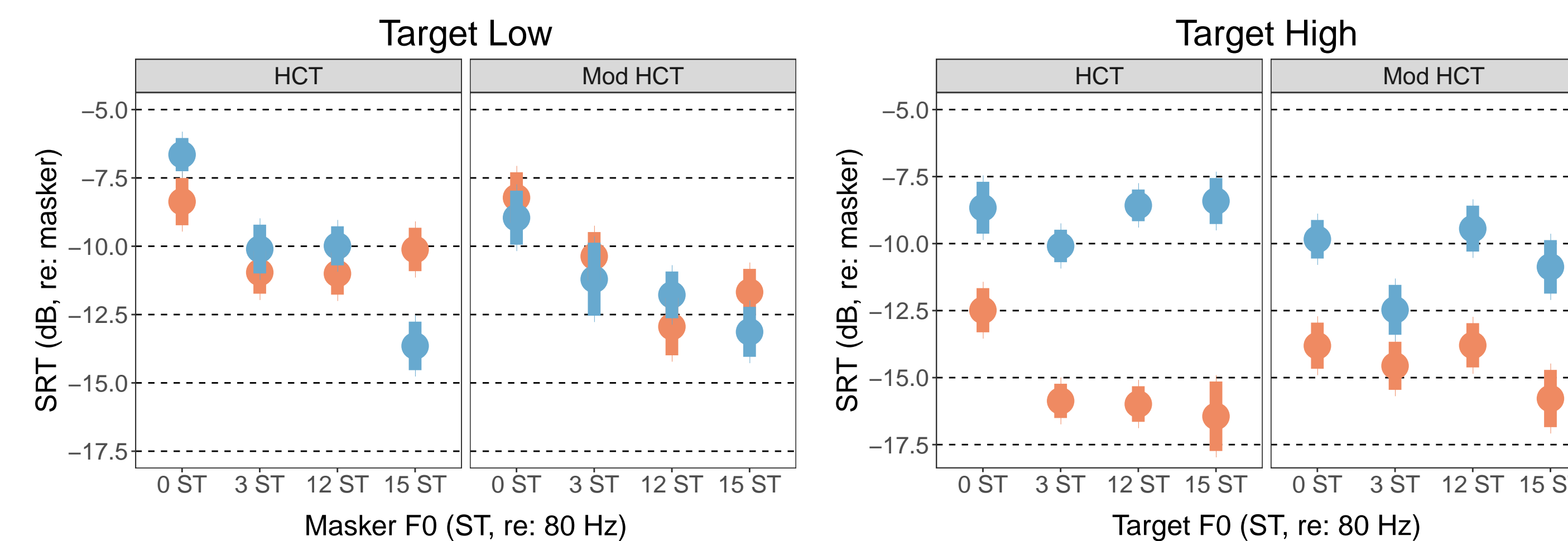
**Figure 6:** SRTs relative to 0 ST vs.  $\Delta F0$ . **All Harm** is in blue, while **Odd Harm** is in orange. Left figure shows Target Low data, while right figure shows Target High data. Error bars are 95% confidence intervals. Octave  $\Delta F0$  data is highlighted.

## Stimuli



**Figure 2:** Spectrograms of example stimuli with 80 Hz F0s. From left to right: Target, HCT, Mod HCT. Top row shows **All Harm**, bottom row shows **Odd Harm**.

## Results — Experiment 2

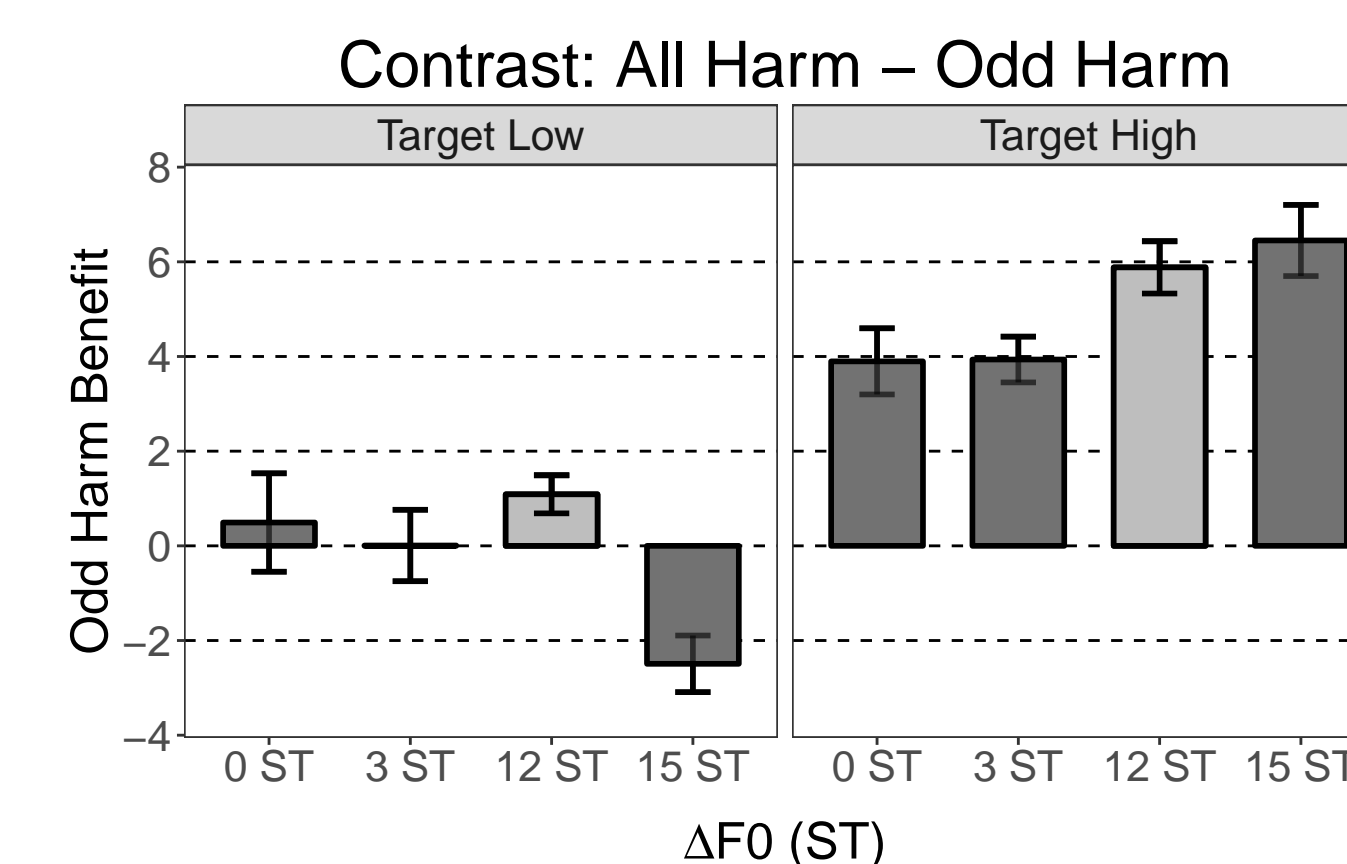


**Figure 5:** SRTs vs.  $\Delta F0$ . **All Harm** is in blue, while **Odd Harm** is in orange. Left figure shows Target Low data, while right figure shows Target High data. Within each figure, left panel shows data with HCT masker, while right panel shows data with Mod HCT masker. Error bars are 95% confidence intervals.

- Mixed-effects model revealed all main effects significant (all  $p < 0.01$ ), significant interactions between  $\Delta F0$  and target pitch ( $p < 0.001$ ),  $\Delta F0$  and spectral structure ( $p = 0.012$ ), and target pitch and spectral structure ( $p < 0.001$ )

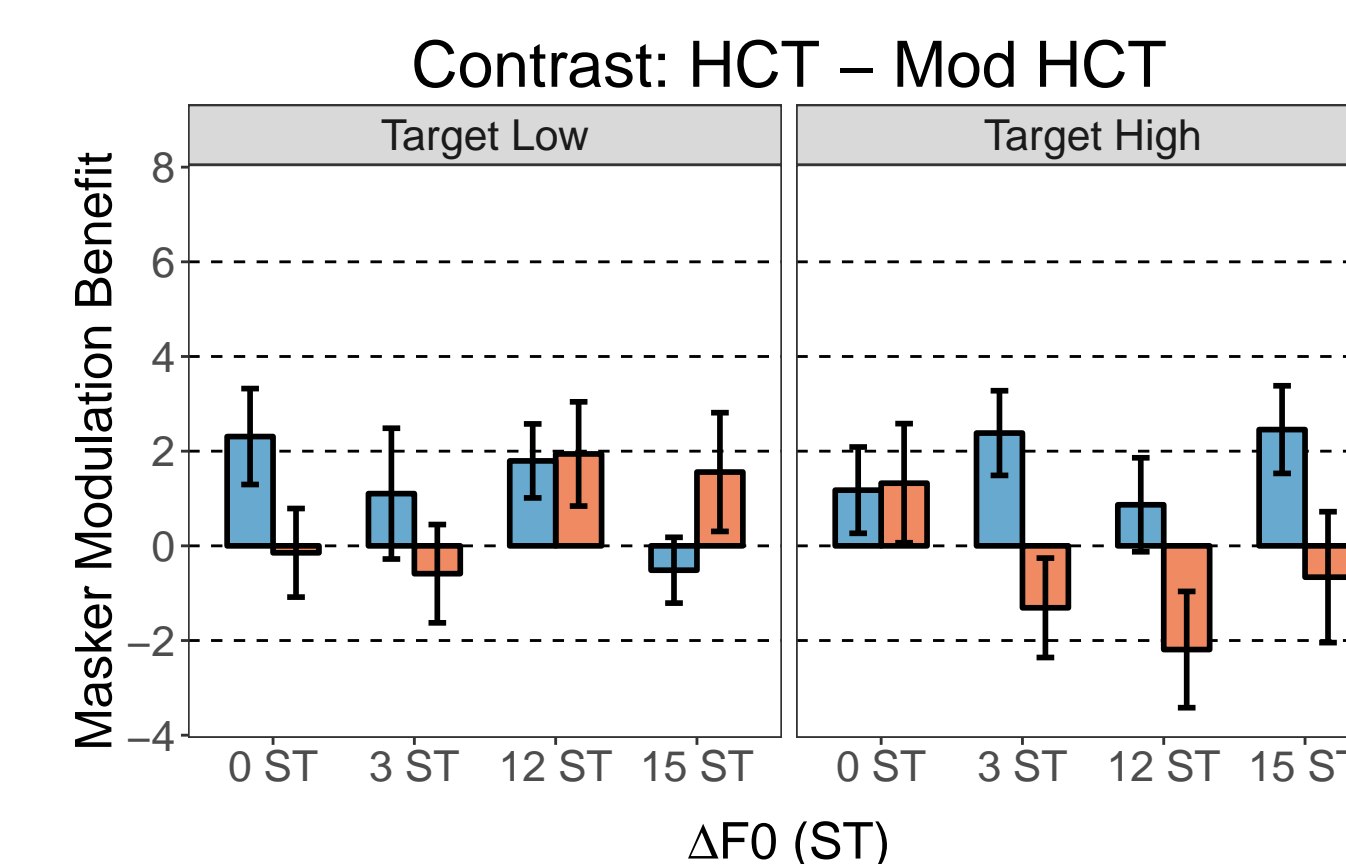
## Analysis — Experiment 2

### Removing masker's even harmonics improved speech intelligibility at octave $\Delta F0$



**Figure 7:** SRTs in **All Harm** relative to **Odd Harm** vs.  $\Delta F0$ . Left figure shows Target Low data, while right figure shows Target High data. Error bars are 95% confidence intervals. Octave  $\Delta F0$  data is highlighted.

### Small, but somewhat inconsistent, benefit of masker temporal modulation



**Figure 8:** SRTs in Mod HCT relative to HCT vs.  $\Delta F0$ . **All Harm** is in blue, while **Odd Harm** is in orange. Left figure shows Target Low data, while right figure shows Target High data. Error bars are 95% confidence intervals.

## Summary — Experiment 1

- H1:** Shifting F0 away from talker's natural range had small impact on intelligibility ( $< 2$  dB SRT; Fig. 4)
  - Magnitude comparable to Deroche et al. [4]
- H2:** Removing target even harmonics reduced talker intelligibility by about 2.5 dB SRT (Fig. 4)
  - Possibly due to reduced naturalness or sparser sampling of spectral envelope
- Processed and monotone speech less intelligible than natural speech (Fig. 4)

## Summary — Experiment 2

- H1:** No octave  $\Delta F0$  benefit at Target High octave  $\Delta F0$  (Fig. 6)
  - At least part of this effect ( $\sim 1$  dB) may be due to reduced intelligibility of target talker at higher F0 values (Fig. 4)
  - Resembles findings of Brokx and Nootboom [2]
- H2:** Removing masker even harmonics at Target High octave  $\Delta F0$  improved speech intelligibility (Fig. 7)
  - Inconsistent with explanation of Brokx and Nootboom [2] based on shared periodicity interfering with cancellation
  - Likely due to introduction of spectral glimpses in masker in this condition (Fig. 1)
- H3:** Mixed evidence for masker modulation (Fig. 8)
  - Overall small but significant benefit of masker modulation (average benefit = 0.8 dB,  $p = 0.014$ )
  - Inconsistent across conditions, but interactions not significant

## Significance

- Hearing-impaired (HI) listeners' reduced  $\Delta F0$  benefit may play a role in their difficulty with multi-talker scenes [7]
- This research suggests that spectral glimpsing plays important role in  $\Delta F0$  benefit — HI listeners may not see these benefits due to broadened auditory filters

## Acknowledgements

- Special thank you to Hideki Kawahara for STRAIGHT code
- Supported by NIH R01 DC005216, UMN CLA Graduate Fellowship, and NSF NRT-UtB1734815

## References

- Douglas Bates et al. "Fitting Linear Mixed-Effects Models Using lme4". In: *Journal of Statistical Software* 67.1 (2015), pp. 1–48. DOI: 10.18637/jss.v067.i01.
- J. P. L. Brokx and S. G. Nootboom. "Intonation and the perceptual separation of simultaneous voices". In: *Journal of Phonetics* 10 (1982), pp. 23–36.
- Alain de Cheveigné. "Separation of concurrent harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model of auditory processing". In: *The Journal of the Acoustical Society of America* 93.6 (1993), pp. 3271–3290.
- Mickael L. D. Deroche et al. "Roles of the target and masker fundamental frequencies in voice segregation". In: *The Journal of the Acoustical Society of America* 136.3 (2014), pp. 1225–1236. DOI: 10.1121/1.4890649.
- Hideki Kawahara, I. Masuda-Katsuse, and Alain de Cheveigné. "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction". In: *Speech Communication* 27 (1999), pp. 187–207.
- E. H. Rothausen et al. "IEEE Recommended Practice for Speech Quality Measurements". In: *IEEE Transactions on Audio and Electroacoustics* 17.3 (1969), pp. 225–246. DOI: 10.1109/TAU.1969.1162058.
- Van Summer and Majorie R. Leek. "F0 Processing and the Separation of Competing Speech Signals by Listeners With Normal Hearing and With Hearing Loss". In: *Journal of Speech, Language, and Hearing Research* 41 (1998), pp. 1294–1306.