The role of pitch and harmonic cancellation in concurrent speech segregation

Daniel Guest, Andrew Oxenham

University of Minnesota, Department of Psychology, Auditory Perception and Cognition Lab

Background

- Multi-talker auditory scenes are commonplace and challenging, but many signal properties offer potential benefits to listeners:
  - e.g., masker temporal modulation, target intonation, & spatial separation (Leclère, Lavandier, and Deroche 2017), target periodicity (Steinmetzger and Rosen 2015), *fundamental frequency (F0) differences (Oxenham 2008)*
- Number of possible mechanisms underlying benefit from F0 differences (ΔF0 benefit)
- Two proposed mechanisms are *spectral glimpsing* (Deroche et al. 2014) and *harmonic cancellation* (Cheveigné 1993)
  - Spectral glimpsing — ΔF0 benefit arises from "listening in the dips" of the masker spectrum that have favorable target-to-masker ratios (TMRs)
  - Harmonic cancellation — ΔF0 benefit arises from the application of masker F0 information to cancel (i.e., inhibit or filter out) the neural representation of the masker
- Brokx and Nooteboom (1982) reported than a fixed octave ΔF0 between two talkers provided little ΔF0 benefit relative to no ΔF0
  - At octave ΔF0, target harmonics overlap with masker harmonics and target/masker share a period
  - *Spectral glimpsing — spectral overlap explains poor octave ΔF0 benefit*
  - *Harmonic cancellation — shared period explains poor octave ΔF0 benefit*
- The present study manipulated target/masker to examine octave ΔF0 under two conditions: normal spectral overlap and absence of spectral overlap
  - *Spectral glimpsing and harmonic cancellation generate different predictions* in latter condition
  - Spectral glimpsing — performance should be good in absence of spectral overlap
  - Harmonic cancellation — performance may remain poor even in absence of spectral overlap because target/masker still share a period

Aims

- Assess extent of ΔF0 benefit with an octave ΔF0
- Examine interactions between ΔF0 benefit and masker temporal modulation
- Determine whether cancellation and/or spectral glimpsing can explain ΔF0 benefit

Methods

- **Outcome measures:**
  - *Speech reception thresholds (SRTs)* measured via 1-up-1-down procedure (Deroche et al. 2014)
- **Stimuli:**
  - *Target:* Male talker speaking IEEE sentences, manipulated to have monotonic F0 via STRAIGHT (Kawahara, Masuda-Katsuse, and de Cheveigné 1999)
  - *Masker:* Random phase harmonic complex tone (HCT) with speech-shaped spectral envelope and monotonic F0
  - Average unresolved stimulus excitation patterns matched across conditions (Fig. 1)
- **Independent variables:**

Name	Levels	Description
ΔF0	0 ST, 3 ST, 12 ST, 15 ST	F0 difference between target and masker
Target Pitch	Target Low	Target F0 fixed at 80 Hz and masker F0 varied
	Target High	Masker F0 fixed at 80 Hz and target F0 varied
Spectral Structure	**All Harm**	Low pitch sound has all harmonics
	Odd Harm	Low pitch sound has only odd harmonics[1]
Masker Type	HCT	Speech-shaped HCT
	Mod HCT	Speech-shaped HCT with broadband speech envelope [2]

1. Odd harmonics removed via IIR comb filter tuned to 2F0, only applied to voiced section of speech
2. Broadband enveloped selected as random sample of concatenated speech stimuli processed with full-wave rectification followed by zero-phase 4th order lowpass filtering

- **Participants & Procedure:**
  - 8 (25 planned) UMN students received $10/hour for participation
  - *Fully factorial within-subjects* design
  - 2 lists (20 sentences) per condition, randomized list-condition pairing and presentation order
  - 64 lists per participant
    - 4 ΔF0 × 2 target pitch × 2 spectral structure × 2 masker modulation × 2 lists per condition
- **Control Experiment:**
  - Target talker against white noise background as function of talker F0 and spectral structure
  - 10 UMN students received course extra credit for participation
  - 2 lists (20 sentences) per condition, list-condition pairing and presentation order randomized
  - 10 lists per participant
    - (4 F0 with **All Harm** + 1 F0 with **Odd Harm**) × 2 lists per condition

Stimuli & Control Results

Unresolved portion of excitation patterns matched across conditions

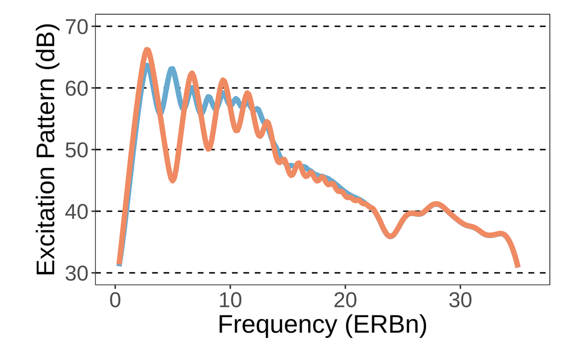


Figure 1: Average excitation patterns for target stimuli with 80 Hz F0. Color indicates spectral structure, with All Harm in blue and Odd Harm in orange.

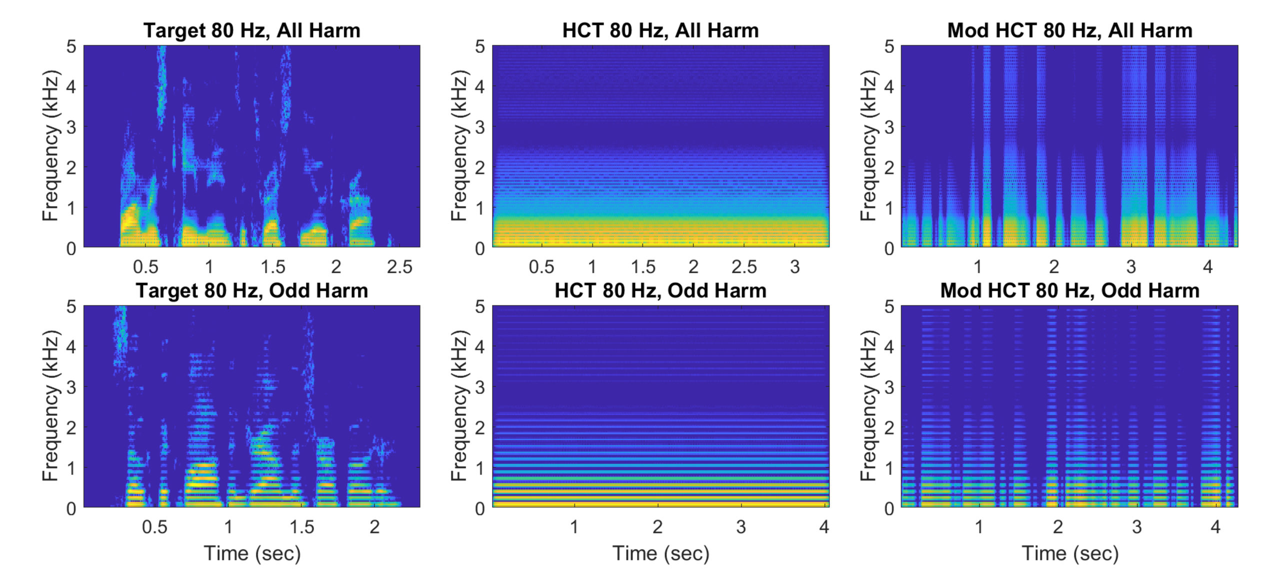


Figure 2: Spectrograms of example stimuli with 80 Hz F0s. From left to right: Target, HCT, Mod HCT. Top row shows All Harm, bottom row shows Odd Harm.

Small effect of F0 & spectral structure variation on intelligibility

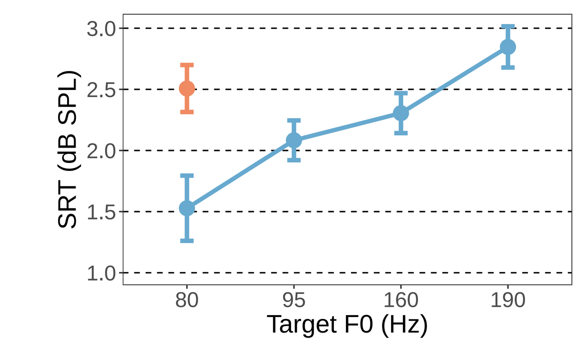


Figure 3: SRT vs target F0 in the control experiment (i.e., with white noise masker). Color indicates spectral structure, with All Harm in blue and Odd Harm in orange. Error bars indicate ±1 standard error of the mean.

Statistical Model

- Mixed-effects linear regression
- Implemented using `lme4` in `R` (Bates et al. 2015)
- **Fixed effects ($\boldsymbol{\beta}$):** ΔF0, target pitch (TP), spectral structure (SS), masker type (MT)
- **Random effects ($\boldsymbol{u}$):** Participant, list
- Fit via penalized maximum likelihood estimation

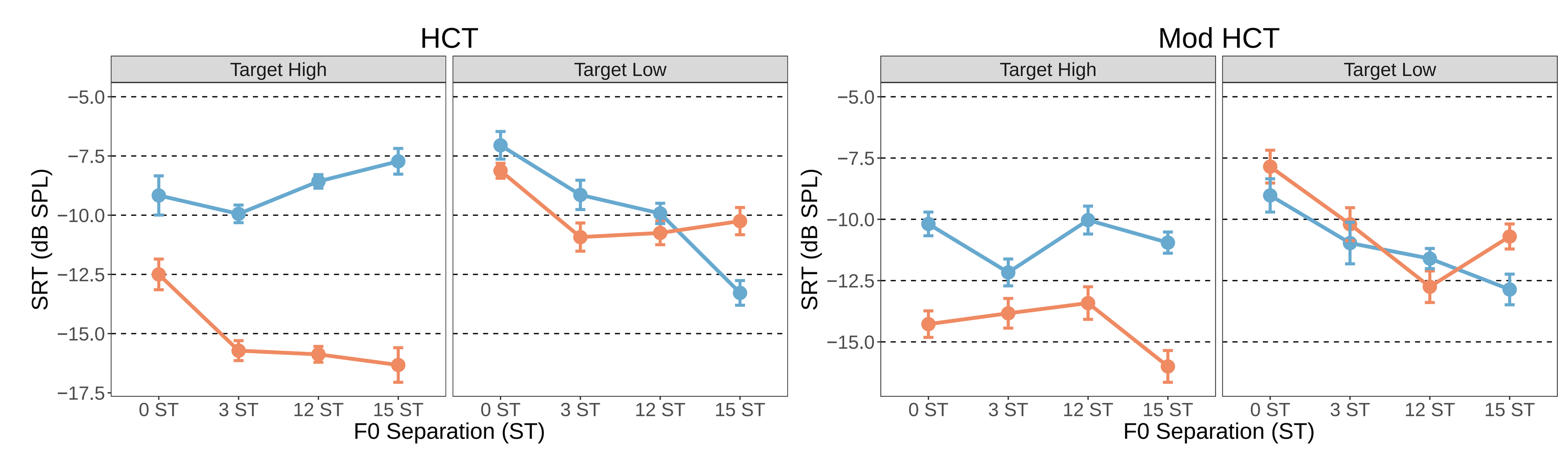
$$y = X\boldsymbol{\beta} + Z\boldsymbol{u} + \epsilon \tag{1}$$

$$\hat{y}_{i,p,l} = \sum_j \beta_j x_{i,j} + \delta_p + \gamma_l \tag{2}$$

…where i indexes observations, j indexes fixed-effects coefficients, p indexes participants, and l indexes lists

- Satisfactory independence of residual errors assessed graphically
- Significant terms (by Type III ANOVA with Wald F test):
  - SS ($p = 0.02$), ΔF0×TP ($p < 0.001$), ΔF0×SS×TP ($p = 0.01$), SS×TP×MT ($p = 0.03$)

Results



No octave ΔF0 benefit in Target High condition

Figure 4: SRT vs ΔF0, averaged across listeners. Panels indicate target pitch. Color indicates spectral structure, with All Harm in blue and Odd Harm in orange. The left-hand figure shows data with HCT, the right-hand figure shows data with Mod HCT. Error bars indicate ±1 standard error of the mean.

- Linear contrasts between 0 ST and 12 ST collapsed across Masker Type revealed ΔF0 benefit at octave in Target Low but not Target High conditions

Removal of even harmonics in Target High condition provided large release from masking

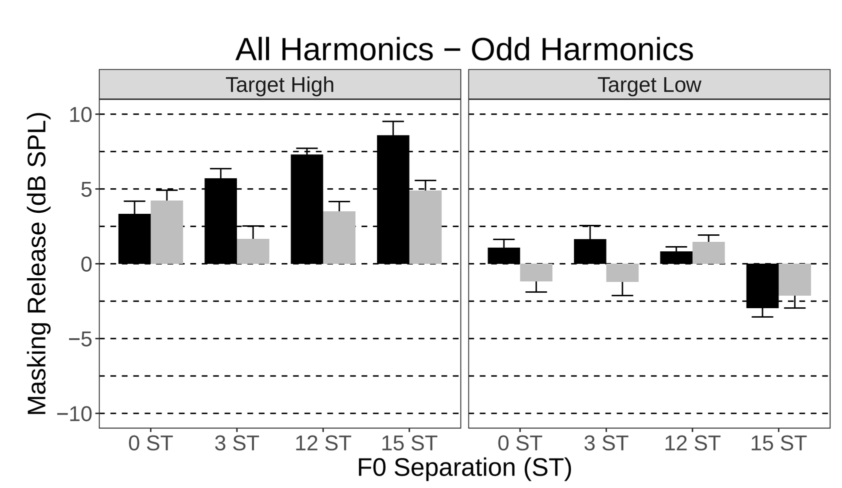


Figure 5: Masking release by removal of even harmonics vs ΔF0. Calculated for each listener, then averaged across listeners. Panels indicate target pitch. Color indicates masker type, with **HCT** in **black** and **mod HCT** in **gray**. Error bars indicate ±1 standard error of the mean.

- Linear contrasts collapsed across Masker Type confirmed significant release from masking in all Target High conditions, no release from masking in any Target Low conditions
- Results consistent with spectral glimpsing theory

Broadband masker modulation provided small but consistent release from masking in All Harm

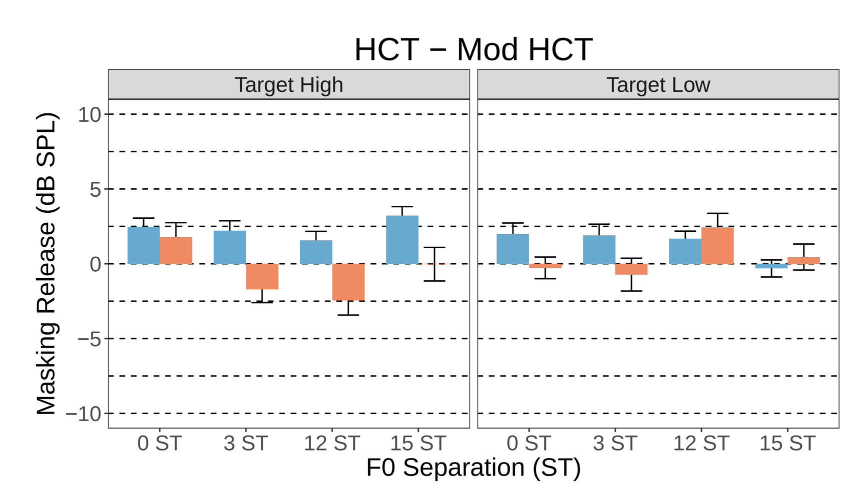


Figure 6: Masking release by masker temporal modulation vs ΔF0. Calculated for each listener, then averaged across listeners. Panels indicate target pitch. Color indicates spectral structure, with All Harm in blue and Odd Harm in orange. Error bars indicate ±1 standard error of the mean.

- Linear contrasts collapsed across ΔF0 revealed significant release from masking with All Harm but not Odd Harm
- Similar in magnitude to previously reported benefits in similar task (Leclère, Lavandier, and Deroche 2017)

In All Harm conditions and with 15 ST ΔF0, Target Low was easier than Target High

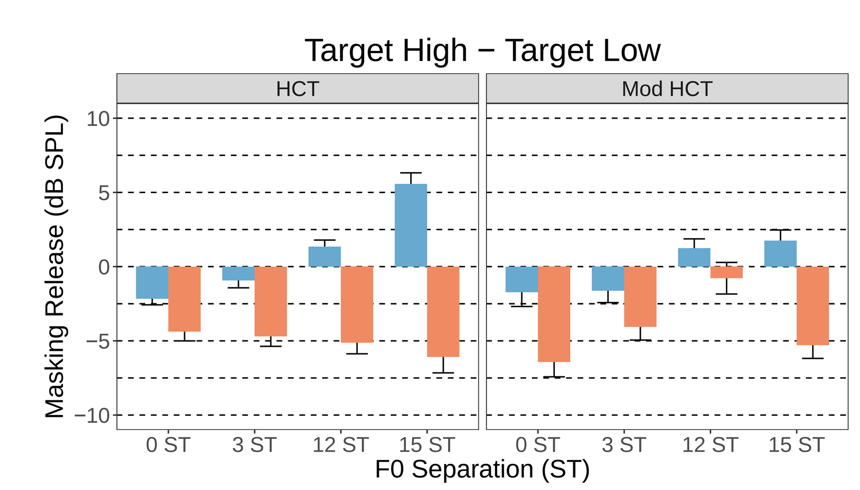


Figure 7: Masking release by target pitch vs ΔF0. Calculated for each listener, then averaged across listeners. Panels indicate masker type. Color indicates spectral structure, with All Harm in blue and Odd Harm in orange. Error bars indicate ±1 standard error of the mean.

- Linear contrasts revealed significant release from masking at 15 ST ΔF0 in All Harm
- Consistent with previous evidence from similar task (Deroche et al. 2014) and spectral glimpsing theory

Conclusions

- Findings similar to Brokx and Nooteboom (1982) — octave ΔF0 provided little benefit in Target High conditions (Fig. 4)
  - At least some of this effect (≥ 1 dB) is attributable to decrease in "intrinsic intelligibility" of target as target F0 increases (Fig. 3)
- Removal of even harmonics provided large release from masking in Target High conditions (Fig. 5)
  - Consistent with spectral glimpsing
  - Inconsistent with naive cancellation mechanism — if it exists, it must not operate here or must operate after spectral glimpsing takes place
- Significant interactions of ΔF0 and target pitch reveal target-masker F0 asymmetry (Fig. 7)
  - Low target F0 and high masker F0 easier than vice versa
  - Also consistent with spectral glimpsing and some previous literature (Deroche et al. 2014)
- Interactions between spectral and temporal glimpsing may exist (Fig. 6, Fig. 7), but interpretation not entirely clear

Future Directions

- Finish data collection and duplicate task with speech maskers
- Further investigate interactions between ΔF0 benefit and temporal glimpsing
- Build ideal observer and/or physiological models to explain results

Significance

- Hearing-impaired (HI) listeners' reduced ΔF0 may play a role in their difficulty understanding speech in multi-talker scenes (Summers and Leek 1998)
- This research suggests that spectral glimpsing underlies ΔF0 benefit — HI listeners may not see these benefits due to broadened auditory filters

Acknowledgements

- Thank you to the members of the APC lab and our participants
- Thank you to the `lme4` group, the R Core Team, and Hideki Kawahara for the code that made this project possible
- Supported by NIH R01 DC005216 and CLA Graduate Fellowship

Resources

Bates, Douglas et al. (2015). "Fitting Linear Mixed-Effects Models Using lme4". In: *Journal of Statistical Software* 67.1, pp. 1–48. DOI: 10.18637/jss.v067.i01.

Brokx, J. P. L. and S. G. Nooteboom (1982). "Intonation and the perceptual separation of simultaneous voices". In: *Journal of Phonetics* 10, pp. 23–36.

Cheveigné, Alain de (1993). "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, pp. 3271–3290.

Deroche, Mickael L. D. et al. (2014). "Roles of the target and masker fundamental frequencies in voice segregation". In: *The Journal of the Acoustical Society of America* 136.3, pp. 1225–1236. DOI: 10.1121/1.4890649.

Kawahara, Hideki, I. Masuda-Katsuse, and Alain de Cheveigné (1999). "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction". In: *Speech Communication* 27, pp. 187–207.

Leclère, Thibaud, Mathieu Lavandier, and Mickael L. D. Deroche (2017). "The intelligibility of speech in a harmonic masker varying in fundamental frequency contour, broadband temporal envelope, and spatial location". In: *Hearing Research* 350, pp. 1–10. DOI: 10.1016/j.heares.2017.03.012.

Oxenham, Andrew J. (2008). "Pitch Perception and Auditory Stream Segregation: Implications for Hearing Loss and Cochlear Implants". In: *Trends in Amplification* 12.4, pp. 316–331. DOI: 10.1177/1084713808325581.

Steinmetzger, Kurt and Stuart Rosen (2015). "The role of periodicity in perceiving speech in quiet and in background noise". In: *The Journal of the Acoustical Society of America* 138.6, pp. 3586–3599. DOI: 10.1121/1.4936945.

Summers, Van and Majorie R. Leek (1998). "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, pp. 1294–1306.