

# Separating segmental and prosodic contributions to intelligibility

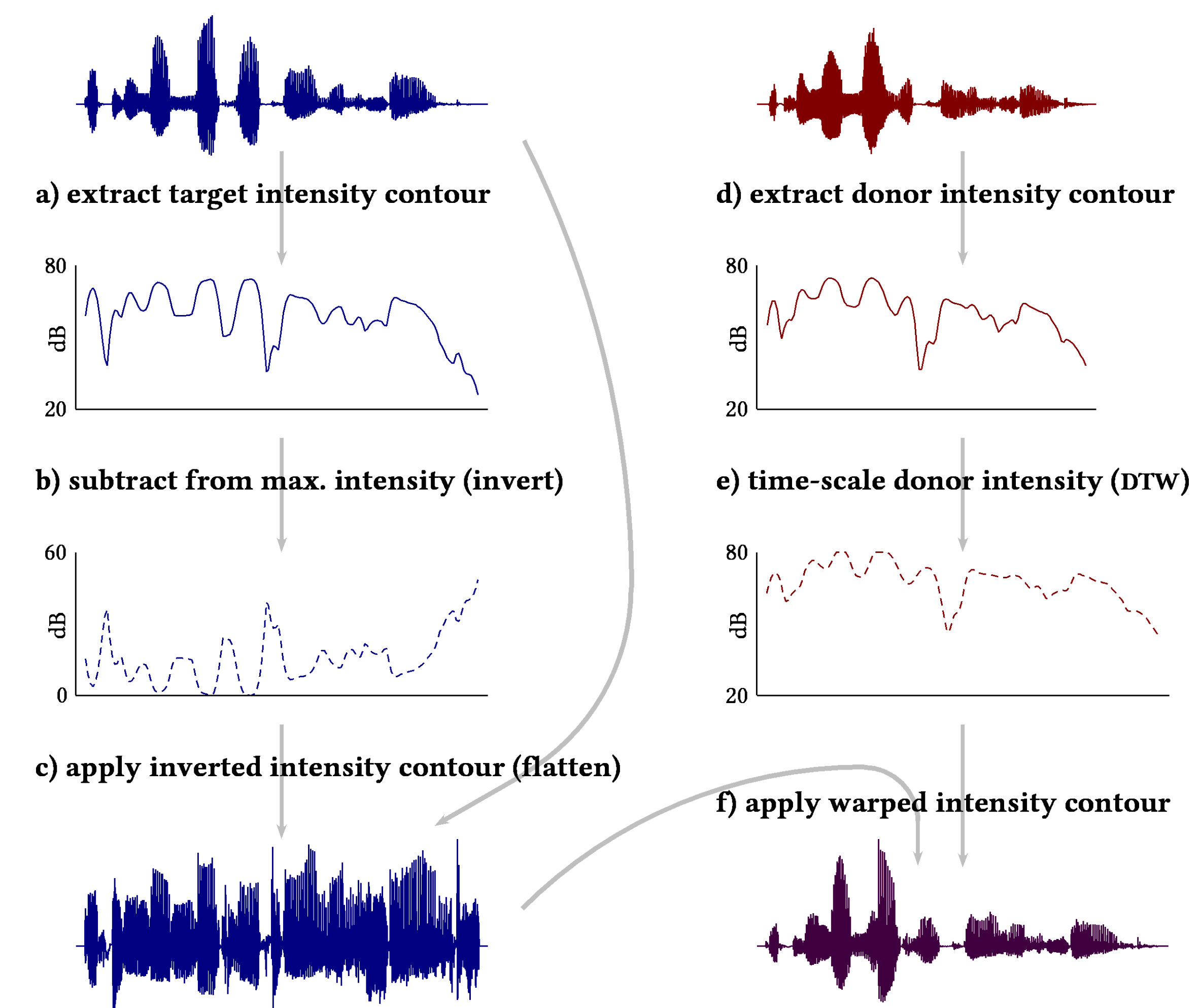
Daniel McCloy • Institute for Learning & Brain Sciences • University of Washington

## Introduction

- PURPOSE** Isolate prosodic contributions to speech intelligibility
- METHOD** Compare intelligibility of prosodically swapped sentences
- STIMULI** Parallel corpus of read sentences from talkers known to vary in intelligibility.  
Three male talkers (same dialect as the 16 listeners), 90 sentences, scored 5 keywords per sentence.

## Methodology: Prosodic swapping

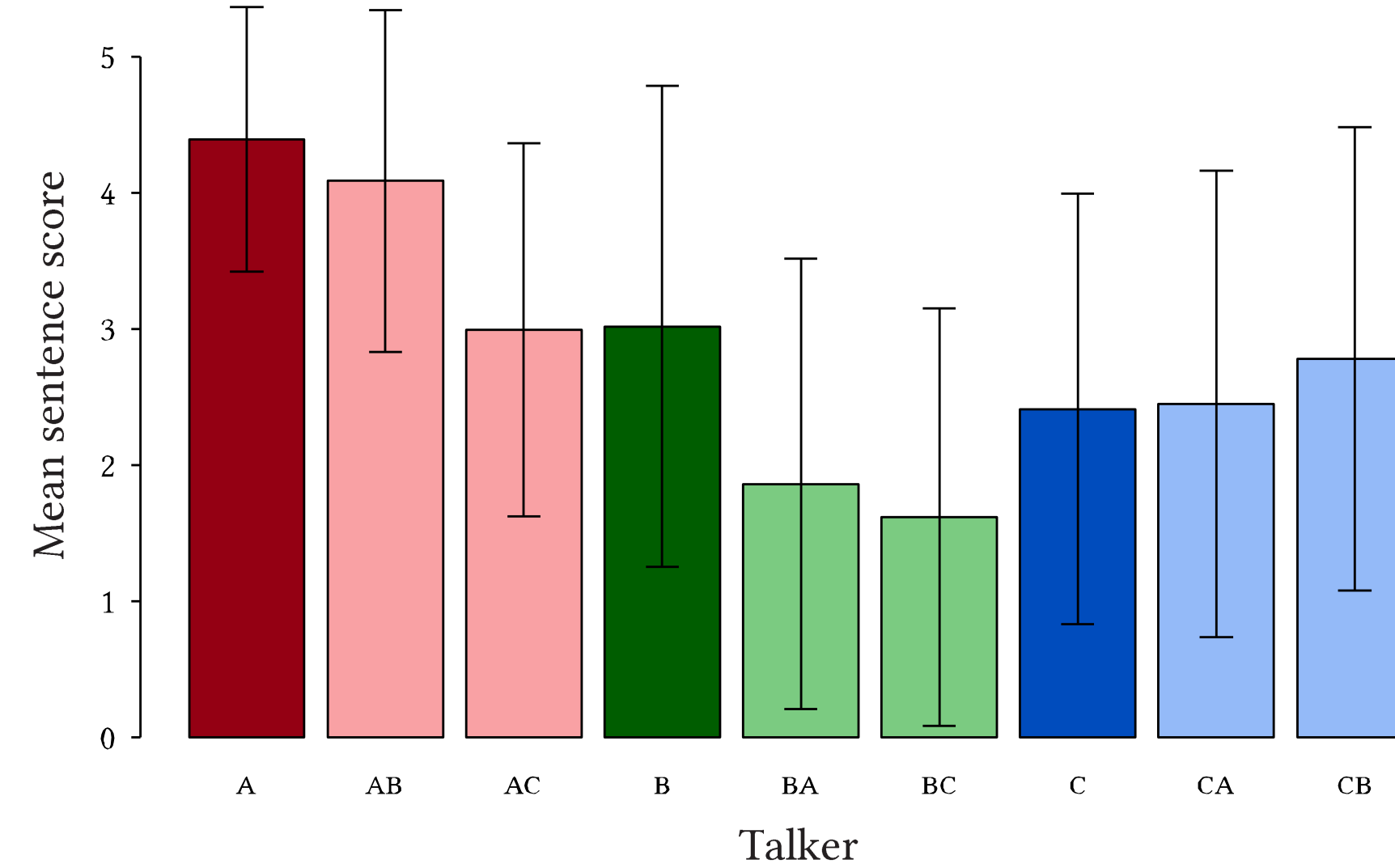
- HOW IT WORKS** Replace intensity contour, pitch contour, and syllable durations of Talker A's sentences with corresponding intensity, pitch, and syllable durations from Talker B (PSOLA™ resynthesis)
- CHALLENGES** Segment misalignment within syllables: *syllabification based on intensity contours*
- Creaky voicing / glottalization: *better resynthesis through hand-correction of pulse epochs*
- Extreme devoicing: *Excluded from corpus.*



## Mixed effects regression model

**MODEL FORMULA**  $\text{lmer}(\text{score} \sim \text{resynth} + \text{signal} + \text{prosody} + \text{trial} + (1|\text{listener}) + (1|\text{sentence}))$

- All fixed effects signif. ( $p < 10^{-3}$ )
- No correlation of fixed effects
- Variation in listener performance minimal
- Variation in sentence difficulty moderate



Predictor	Effect (keywords)
	Talker A baseline
Prosodic donor	Talker B +0.3
	Talker C -0.6
	Talker A baseline
Signal donor	Talker B -1.7
	Talker C -1.3
	Talker A baseline
Resynthesis distortion	-0.7
Task familiarization (trial 1 - 90)	+0.5
Listener variability (standard error)	±0.2
Sentence variability (standard error)	±0.7

## What predicts different dimensions of intelligibility?

**OVERALL PATTERN (TALKER A > B > C):**

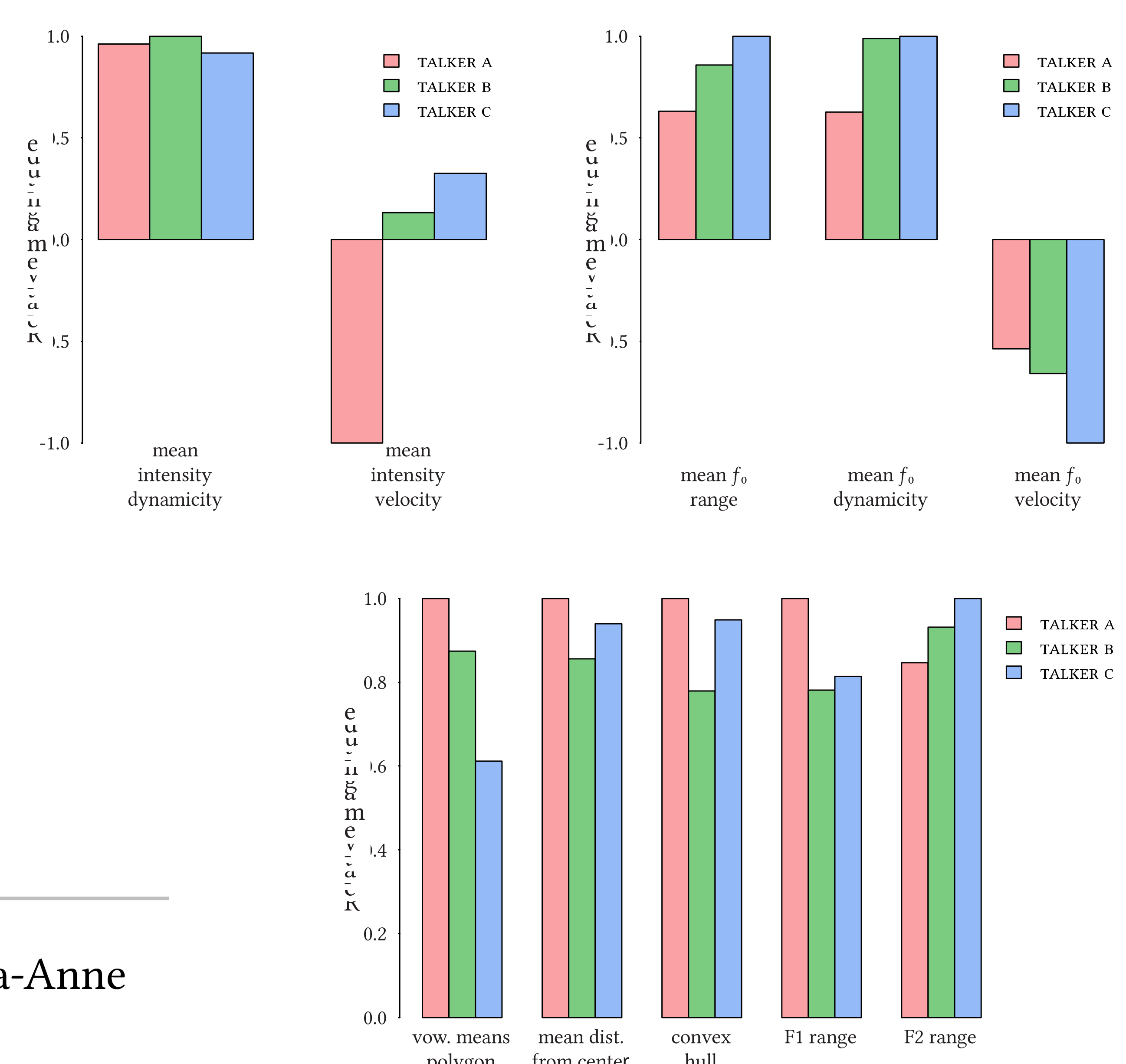
- Proportion of stop consonants that are unreduced
- Vowel formants: area of polygon formed by vowel means

**NON-PROSODIC PATTERN (A > C > B):**

- Vowel formants: convex hull area, mean distance from center, F1 range

**PROSODY-BASED PATTERN (B > A > C):**

- No perfect match; closest: mean  $f_0$  dynamicity, mean  $f_0$  range
- Pitch measures confounded by Talker C's creaky voicing



**Acknowledgments** Richard Wright, Erick Gallun, Sharon Hargus, Gina-Anne Levow, Pamela Souza, KC Lee, UW Linguistic Phonetics Laboratory members