PROBING **VOICE QUALITY'S** CONTRIBUTION TO TONE PERCEPTION:

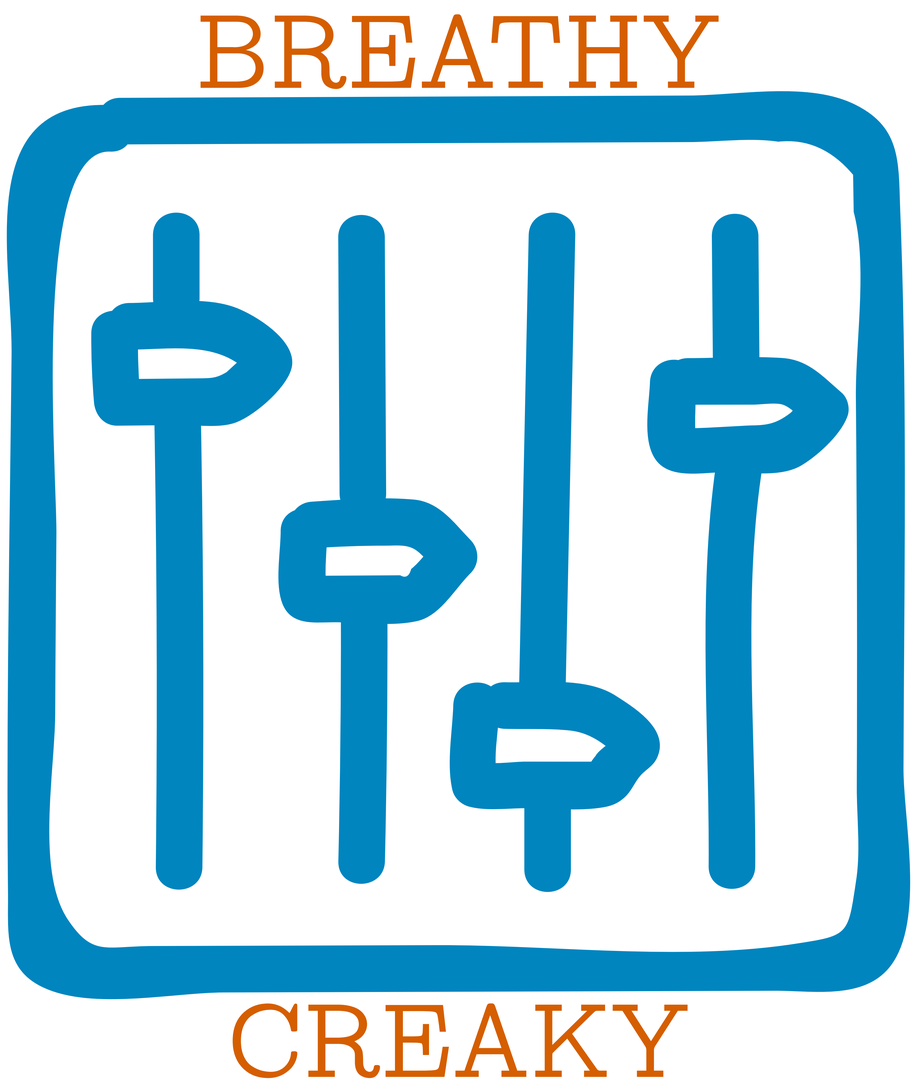
CHALLENGES FOR SYNTHESIS SOFTWARE

Kristine M. Yu
Department of Linguistics
University of Massachusetts Amherst

ICPhS2019
5-9 August 2019 | Melbourne Australia

INTRODUCTION

Understanding interaction of fundamental frequency and other aspects of voice quality in tone perception demands fine control of voice quality in stimulus design.



Strategies for voice quality stimuli design

- **Use library of exemplars collected from natural speech**
- **Synthesize from scratch using voice quality synthesizer**

- **Synthesis from scratch enables isolation of parameters that can be confounded in naturally produced stimuli, e.g., types of creak** (Keating et al. 2015, Garellek 2019)
- **But relation of output to properties of natural speech *entirely dependent on underlying physical/mathematical model***

KlattGrid

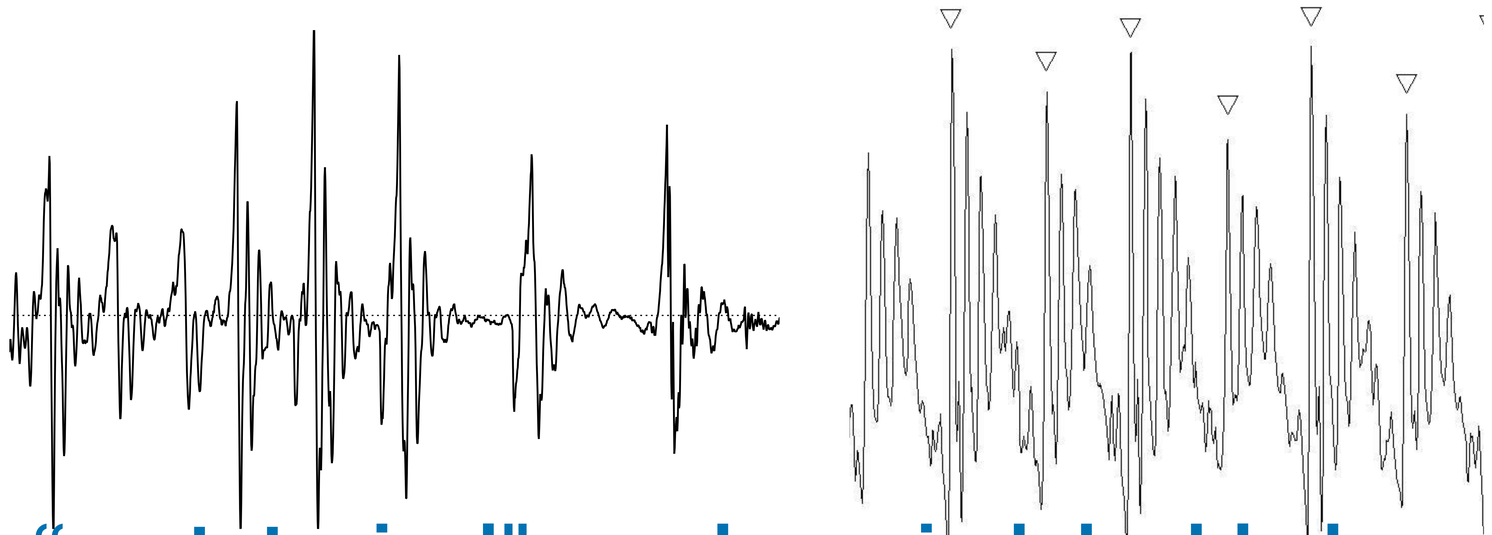
(Weenink 2009)

Brunelle 2012: Cham,
Zhang & Kirby 2018: Cantonese,
Huang 2019: Mandarin)

UCLA Voice Synthesizer

(Kreiman et al. 2010)

(Garellek et al. 2013: Hmong)



"prototypical" creak period doubled creak

How well does individually manipulating voice quality parameters in these tools produce stimuli with desired acoustic properties?

METHODS: parameters varied

KlattGrid parameters individually varied while others held constant at two default settings for [a]: man (M), woman (W)

PARAMETER	DESCRIPTION
TL (dB)	Spectral tilt
DI (%)	Double pulsing/period doubling
OQ (%)	Open phase/quotient
FL (%)	Flutter: fluctuations in f0
F0 (Hz)	Fundamental frequency
AH (dB SPL)	Breathiness amplitude

Just these shown in results

UCLA synthesizer

(stimuli from Garellek et al. 2013)

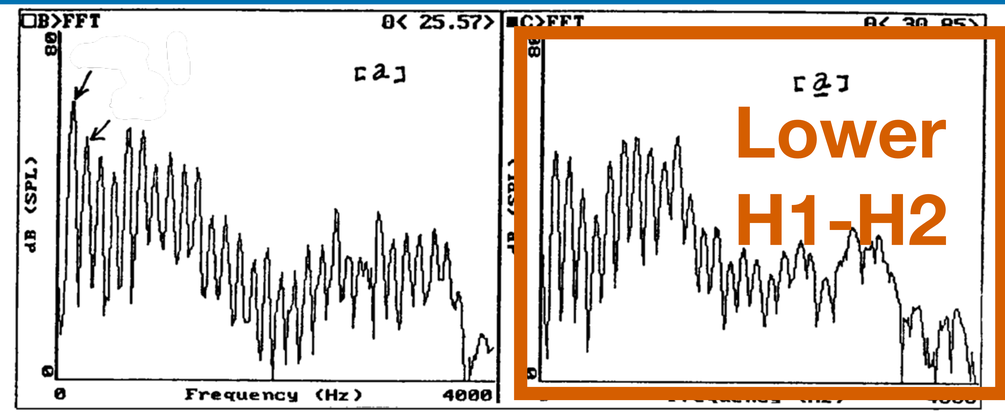
PARAMETER VARIED	MANIPULATION
H1	H1-H2 ↑
H2	H1-H2 ↑, H2-H4 ↓
H4	H2-H4 ↑, H4-2kHz ↓
2kHz	H4-2kHz ↑, 2kHz-5kHz ↓
5kHz	2kHz-5kHz ↑

METHODS: acoustic analysis

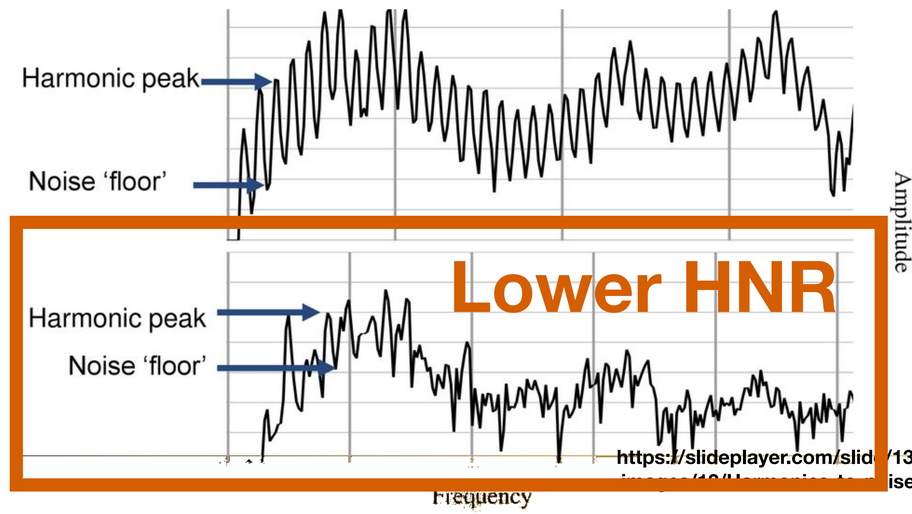
- Acoustic voice quality parameters measured using VoiceSauce (Shue et al. 2011)
- Regression analysis in R, with synthesis parameter as IV, acoustic parameter as DV

EXAMPLE

- Period-doubled creak has low H1-H2 (low spectral tilt) and high noise (low HNRs) (Keating et al. 2015)
- Therefore we'd like to see increase in KlattGrid DI lower HNR and H1-H2

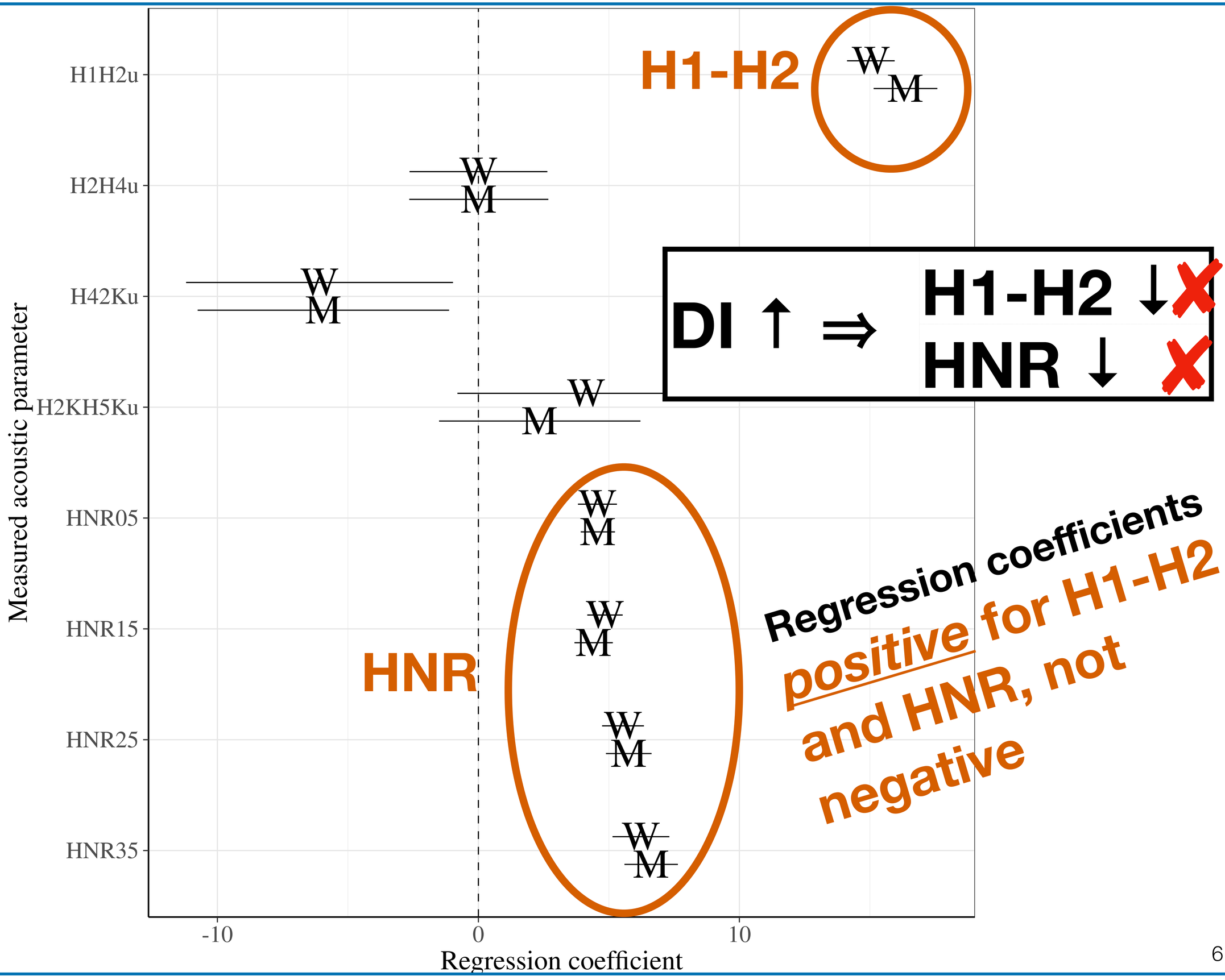


Kong (1996)



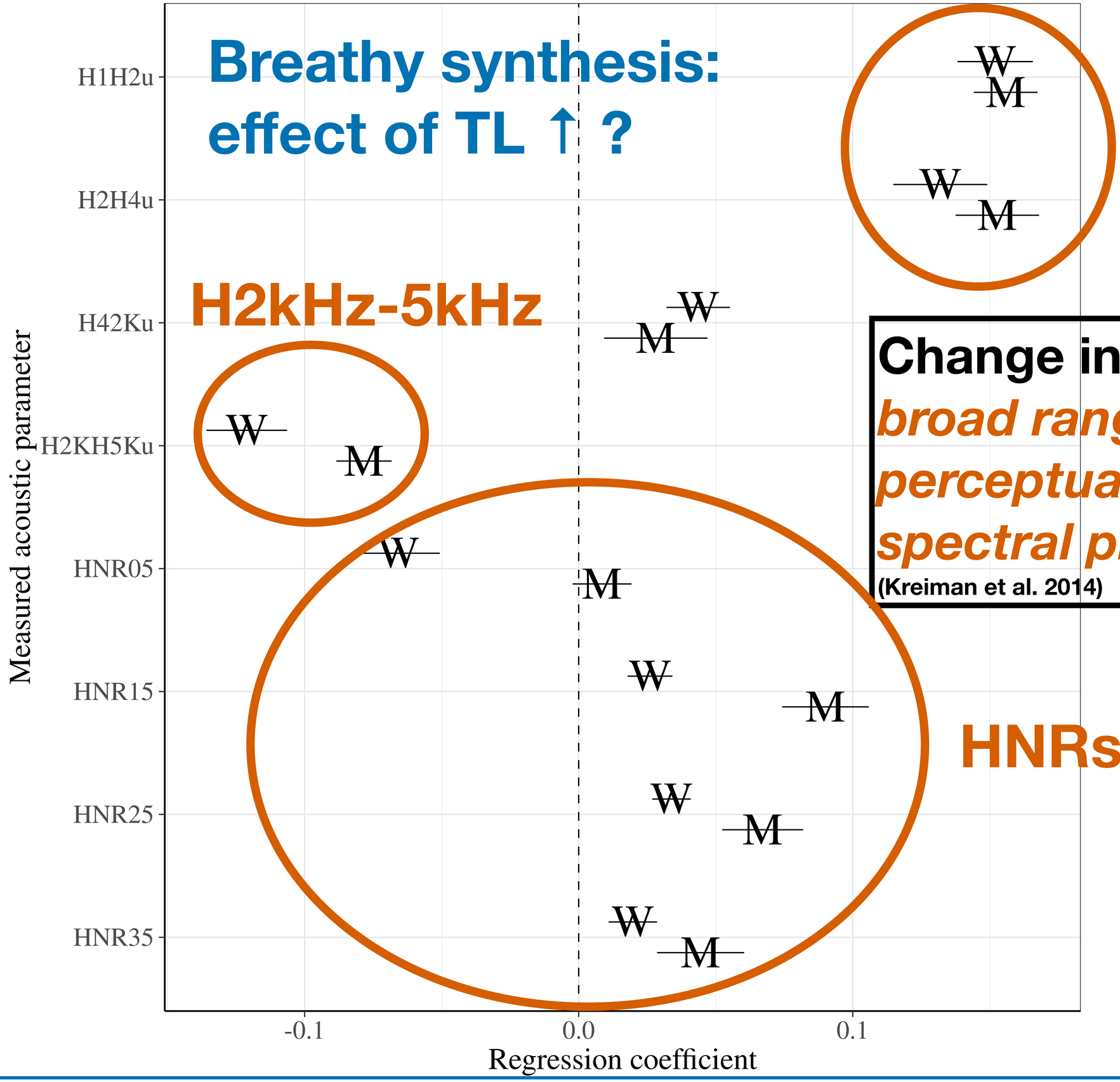
https://slideplayer.com/slid /13152153/79/ ise+Ratio.jpg

Klatt: DI (period doubling)



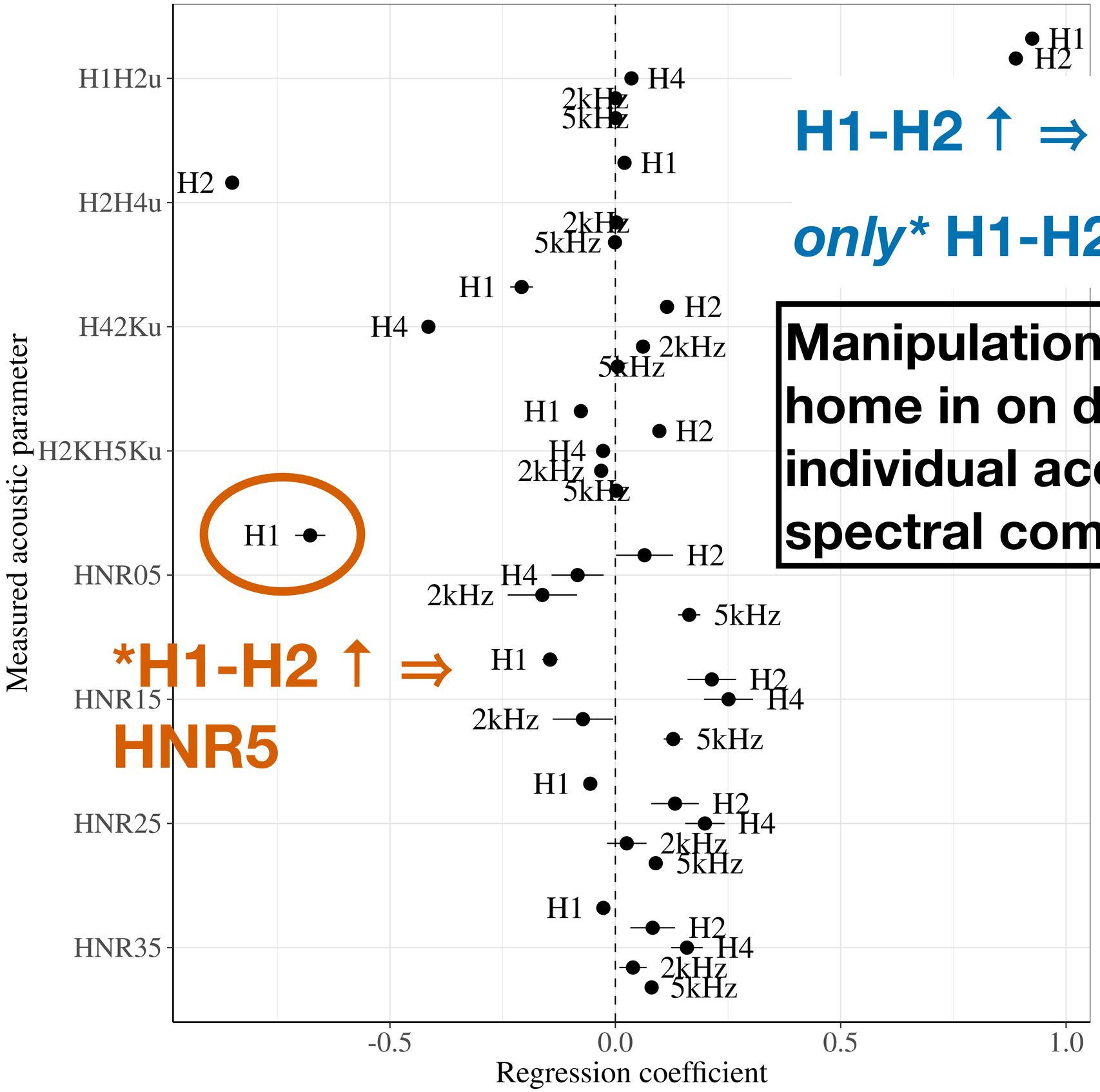
Klatt: TL (spectral tilt)

Breathy synthesis: effect of TL ↑ ?



Change in TL targets
broad range of perceptually distinct spectral properties
(Kreiman et al. 2014)

UCLA: spectral measures



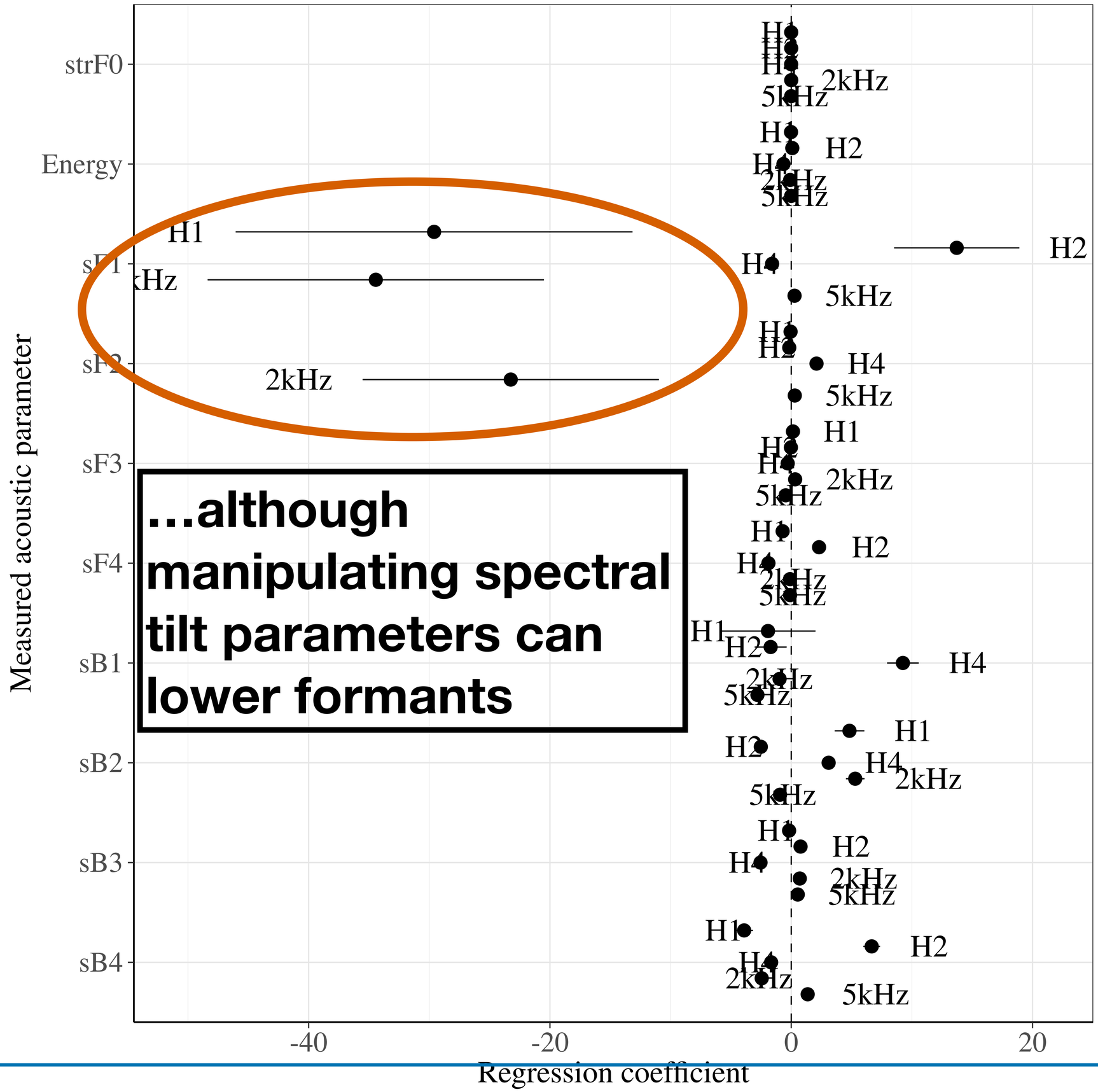
H1-H2 ↑ ⇒

only H1-H2 ↑*

Manipulations mainly home in on desired individual acoustic spectral components…

***H1-H2 ↑ ⇒ HNR5**

UCLA: f0, energy, formants



CONCLUSION

- **Not clear that KlattGrid synthesis of period-doubled creak by manipulating DI produces stimuli with the properties of naturally produced period-doubled creak.**
- **KlattGrid synthesis of breathiness using TL, OQ, AH doesn't home in on desired acoustic properties as well as UCLA Voice synthesizer**

Further work needed to define models that can synthesize voice quality for perception experiments, especially for creaky stimuli.

For further results and analyses, see OSF repository: https://bit.ly/2HPXmvd