Assessing tonal specifications through simulation and classification: The case of post-*wh* accent in Japanese

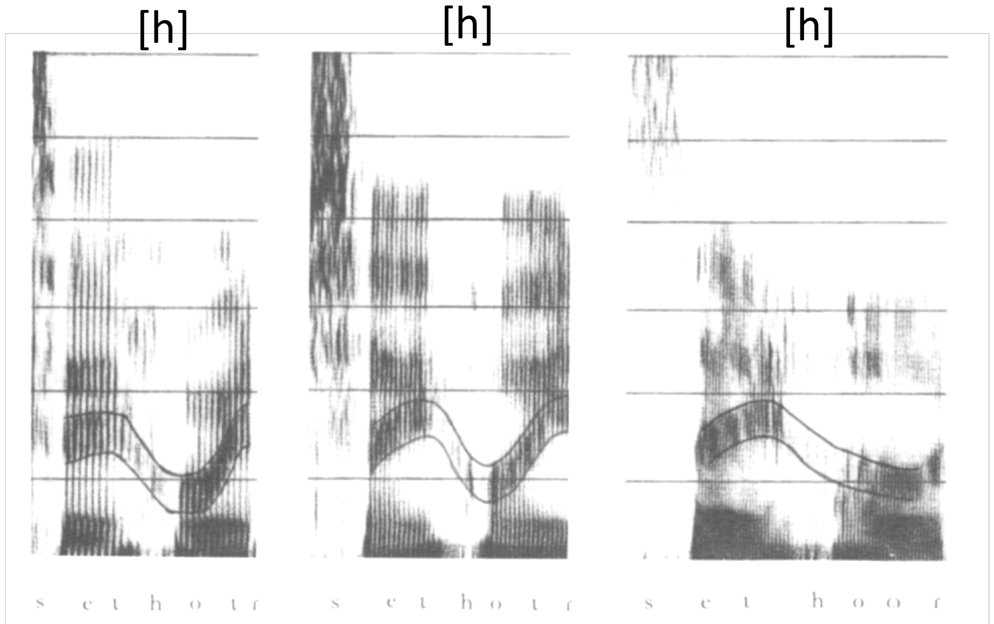
Shigeto Kawahara (Keio)
Jason Shaw (Yale)
Shinichiro Ishihara (Lund University)

May 2019
Seoul, HISPhonCog

Introduction

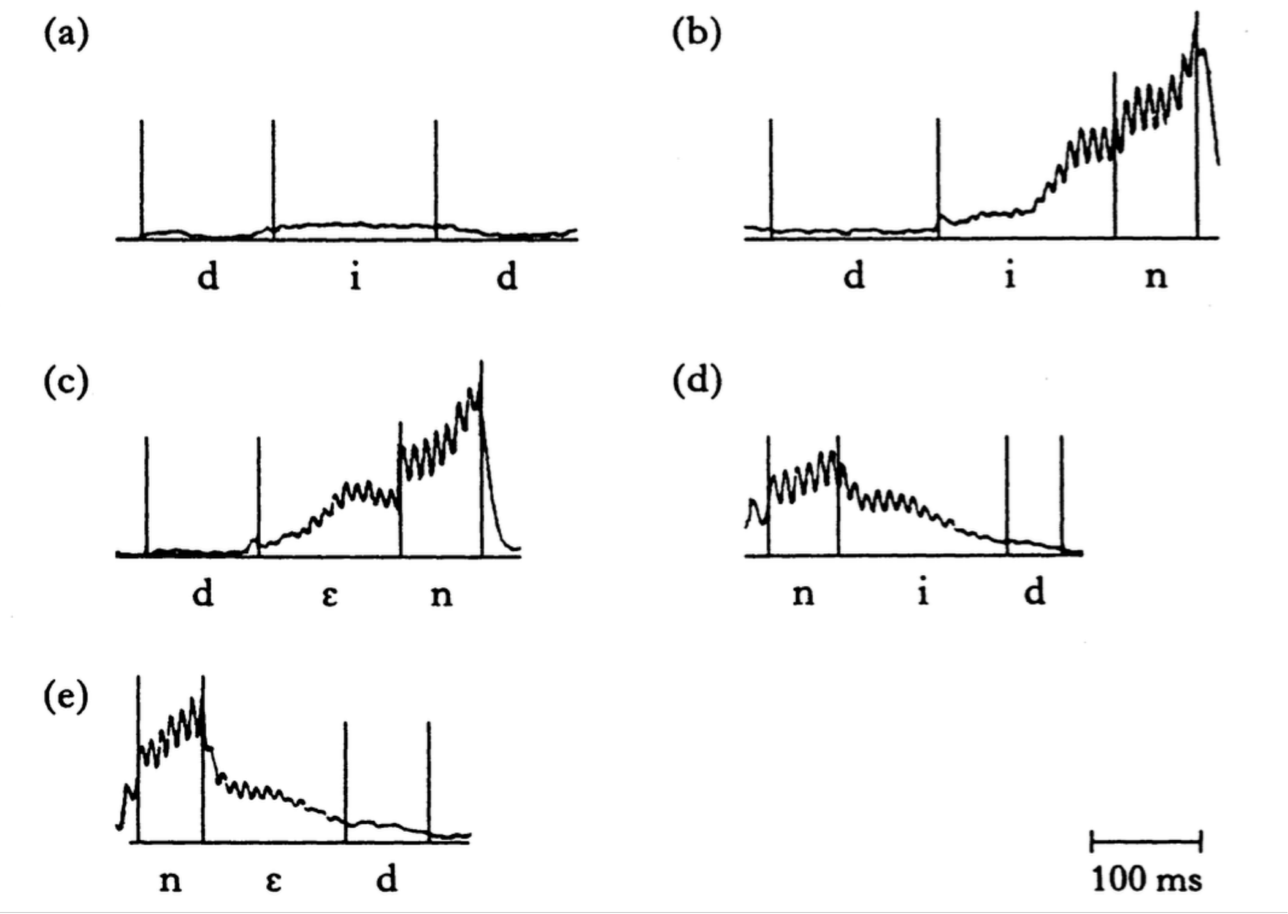
(A) very broad question

- How do we know whether a particular segment X has a phonetic target in some phonetic dimension?
- Assumption: contra SPE, not every segment has a phonetic target in every dimension.

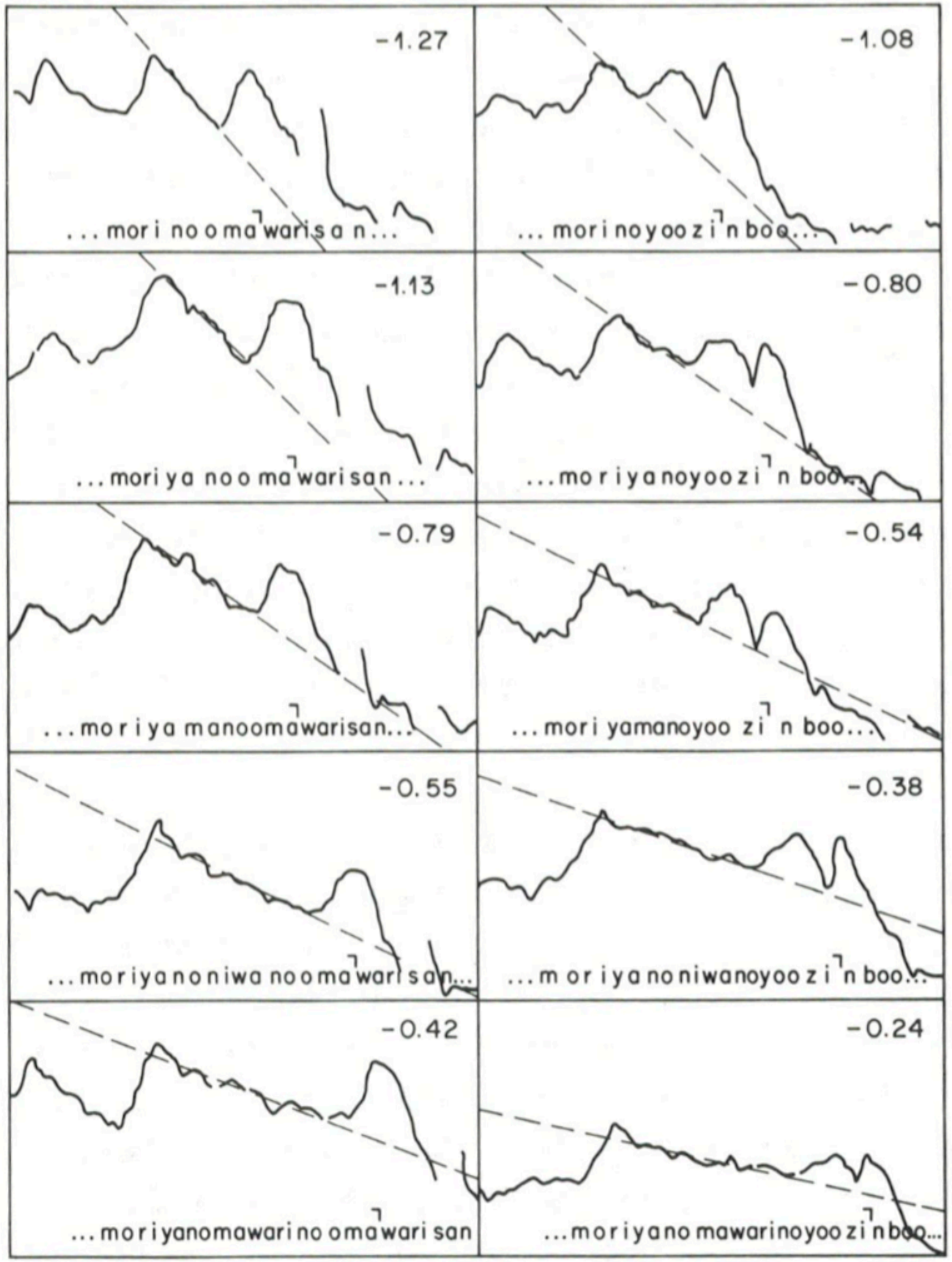


Keating (1988):
[h] may not have a F2 target of its own.

Another example of phonetic underspecification



Cohn (1993):
English vowels do not have a phonetic target in terms of its nasality.



Tonal underspecification

P & B (1988):
Japanese unaccented nouns
do not have a tonal target (except
for phrasal initial rise).

Cf. Haraguchi (1977) who posit H-tone
spreading rule

The challenge

- What looks like (linear) interpolation seems to be a good indication of targetlessness.
- But, a real phonetic trajectory is *always* noisy, and is never completely linear. **What is a realistic baseline for assessing targetlessness**?
- Shaw & Kawahara (2018) developed a computational toolkit to address this problem. S&K analyzes the tongue dorsum trajectory of devoiced vowels in Japanese. This work extends the same toolkit to intonational analyses.

The current case study

- Post-wh accent in Japanese is *eradicated* (Deguchi & Kitagawa 2002). Sample pitch tracks from Ishihara (2001).

declarative

wh-sentence

(28a): Non-interrogative sentence

(28b): *Wh*-question

wh

eradication?

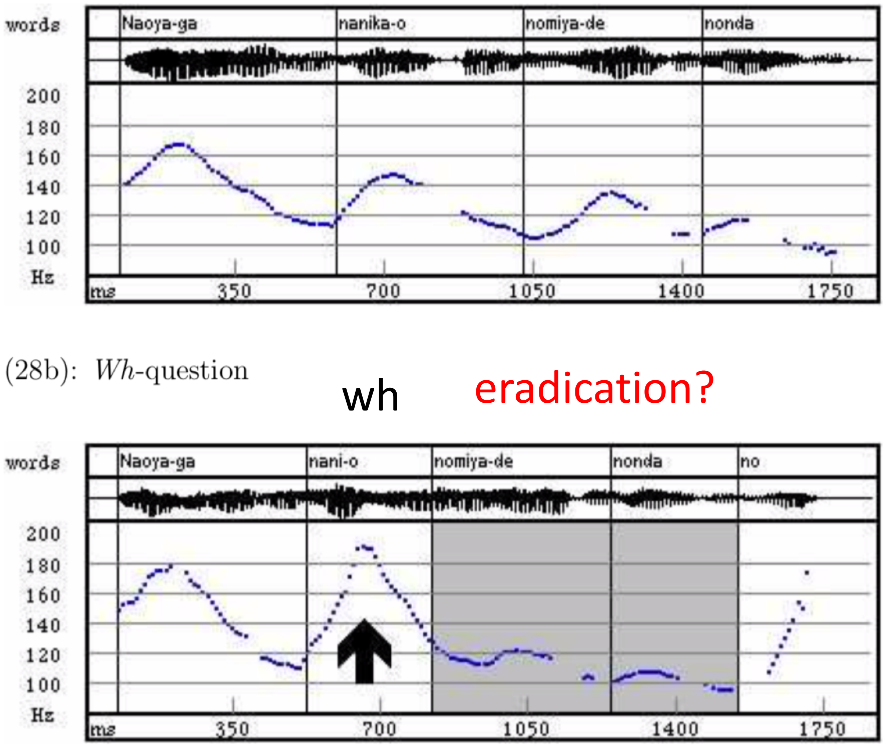
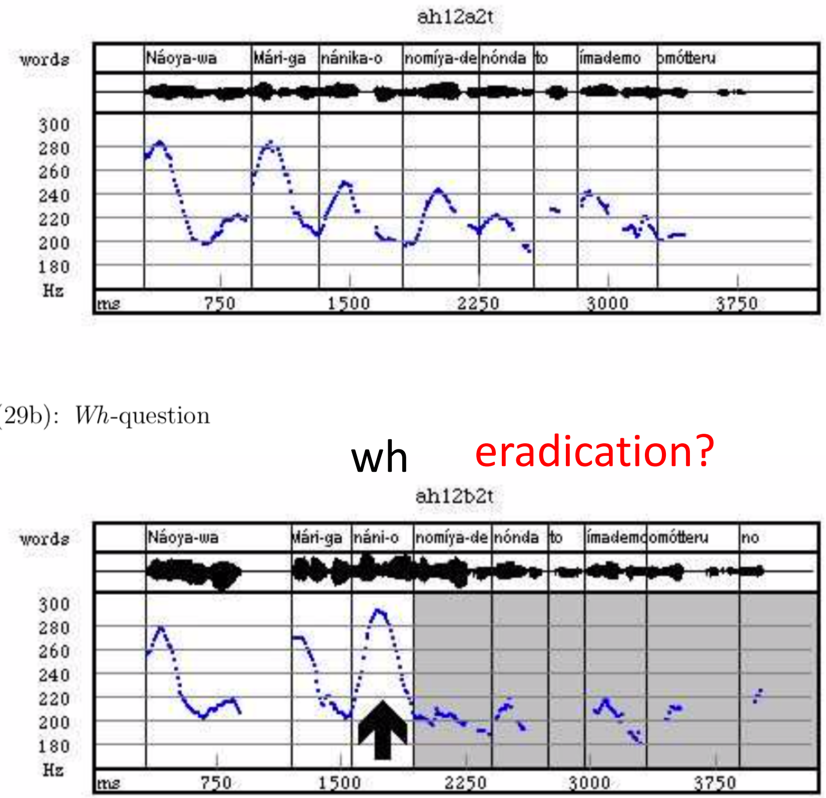


Figure 3-1: Single *wh*-question

(29b): *Wh*-question

wh

eradication?



Richards (2010)

- Why wh-elements in English move, and wh-elements do not move in Japanese?
- Strong/weak features (Chomsky 1995)?

- Richard's (better) answer: Japanese has a prosodic means to group *wh*-elements and a licenser, and hence no need for overt movement. English does not have this prosodic means, hence needs to resort to syntactic movement.

Is post-wh accent really eradicated?

- Subsequent studies cast doubt on the claim that post-wh elements' accent is really *eradicated* (Hirotani 2005; Ishihara 2011 – also Maekawa 1994).

1. Do naïve speakers (i.e. those who are not authors) show eradication? From D & K's Footnote 2.

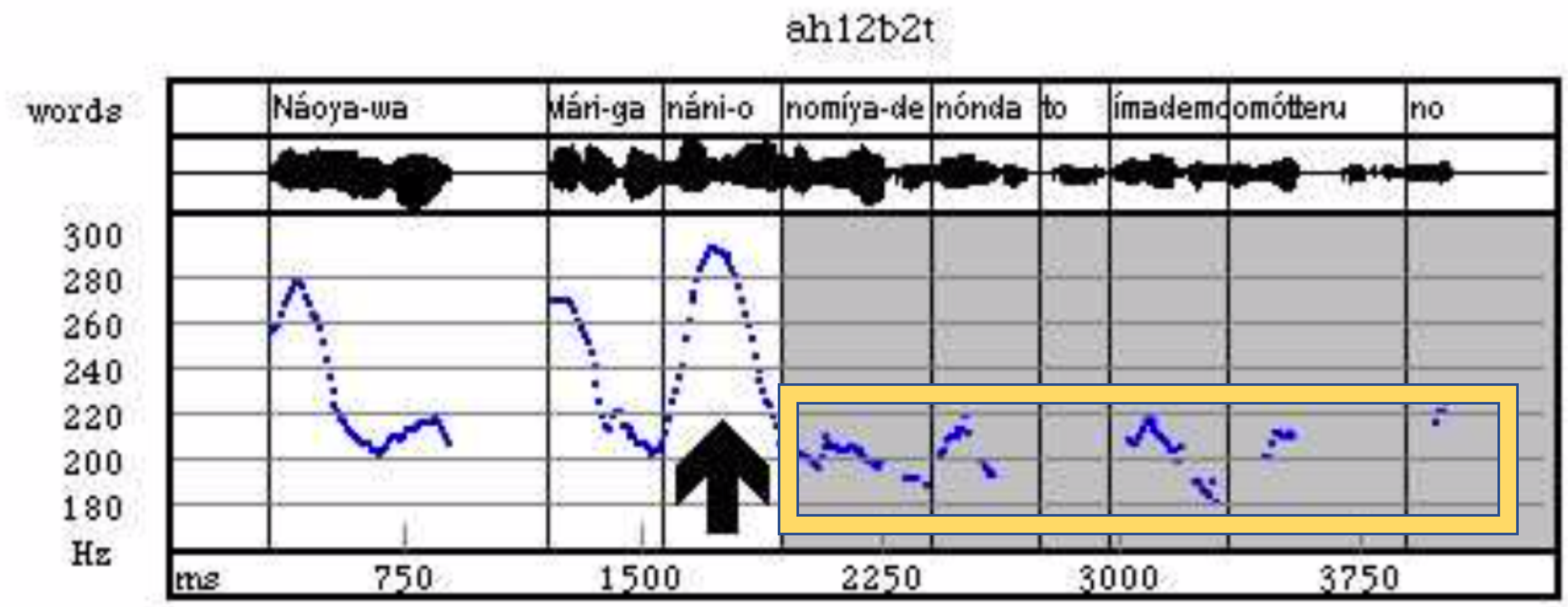
[2] Our pitch-tracking experiments involve our own recordings and are in many ways informal and insufficient. We are presenting them, however, because they seem to us to have turned out to be the faithful physical reflection of our intuition and they help us illustrate our points. In order to minimize our own biases, we have at least conducted some informal perception tests, presenting our recordings and/or our own utterances to over a dozen native speakers of Japanese. They have confirmed that the utterances we have presented to them are accompanied by "natural intonation" for the intended interpretations.

The other, more challenging, issue

2. Are they really eradicated instead of merely *reduced*?

Going back to the broad question, how do we know that post-*wh* items do not have a phonetic (f0) target?

(29b): *Wh*-question



Noisy null or reduced accent?

Method

Method

- Materials come from recordings of Ishihara (2011)
- Nine Tokyo Japanese speakers (4 female)

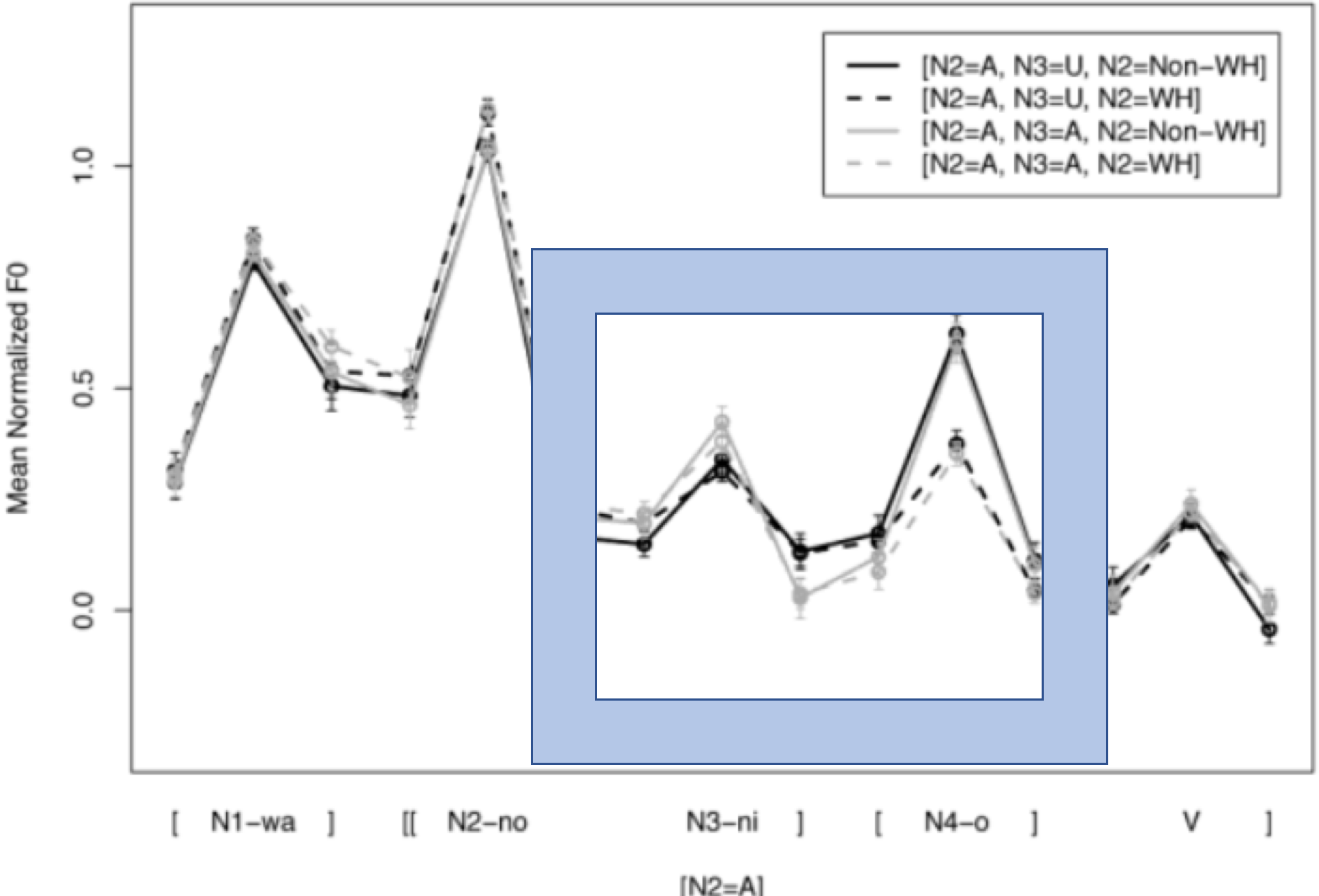
(1) Control sentences: Word$_1$ Word$_{2[-wh]}$ Word$_3$ Word$_4$ Verb

(2) Test sentences: Word$_1$ Word$_{2[+wh]}$ Word$_3$ Word$_4$ Verb

(1) 丸山は$_1$　エルメス$_{[-wh]2}$　襟巻きに$_3$　飲み物を$_4$　こぼしました。

(2)丸山は$_1$　どの人の$_{[+wh]\,2}$　襟巻きに$_3$　飲み物を$_4$　こぼしましたか？

Averaged contours from Ishihara (2011)



Step 1: Discrete Cosine Transform (DCT)

F0 trajectory represented as the sum of Cosines:

$$y(k) = w(k) \sum_{n=1}^{L} x(n) \cos\left(\frac{\pi(2n-1)(k-1)}{2L}\right)$$

$$k = 1,2,\ldots L$$

Where *L* is the number of data samples and ***x(n)* is the trajectory to be modelled** and:

$$w(k) = \begin{cases} \dfrac{1}{\sqrt{L}} & k = 1 \\ \sqrt{\dfrac{2}{L}} & 2 \le k \le L \end{cases}$$

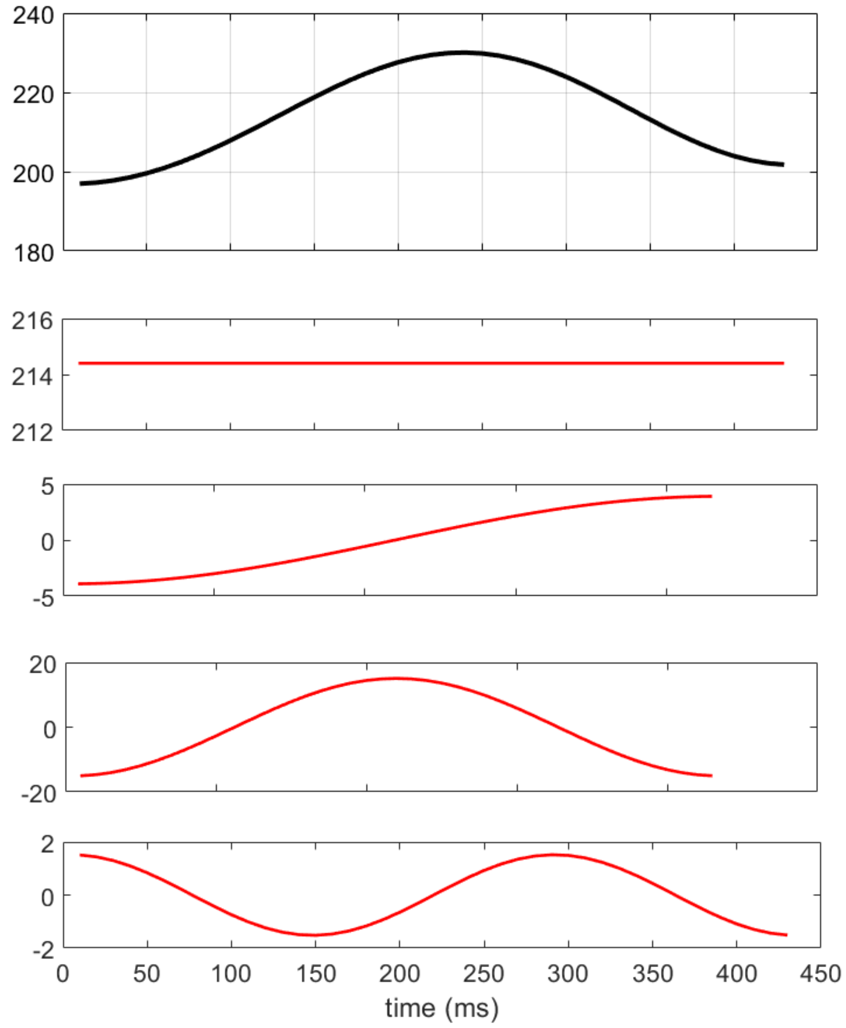
Signal – F0 (Hz)

1st component, *y*(1)

2nd component, *y*(2)

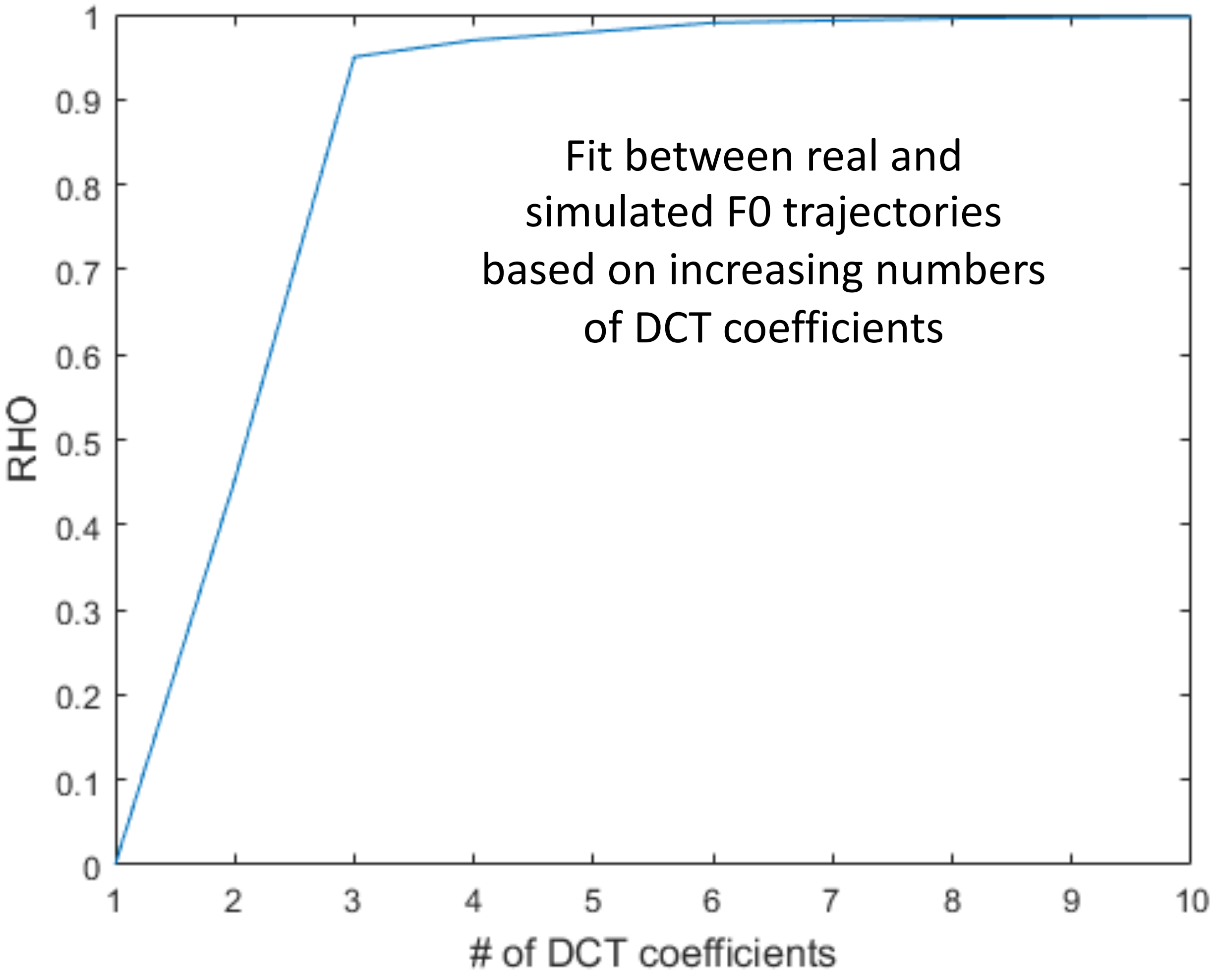
3rd component, *y*(3)

4th component, *y* (4)

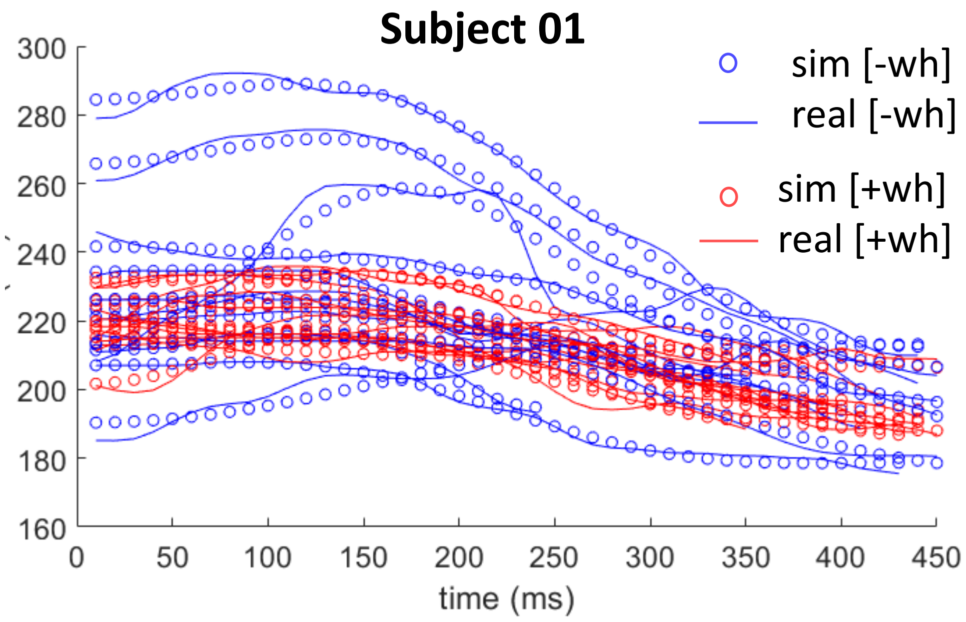


Shaw, J. A., & Kawahara, S. (2018). Assessing surface phonological specification through simulation and classification of phonetic trajectories. *Phonology, 35*(3), 481-522. doi:10.1017/S0952675718000131

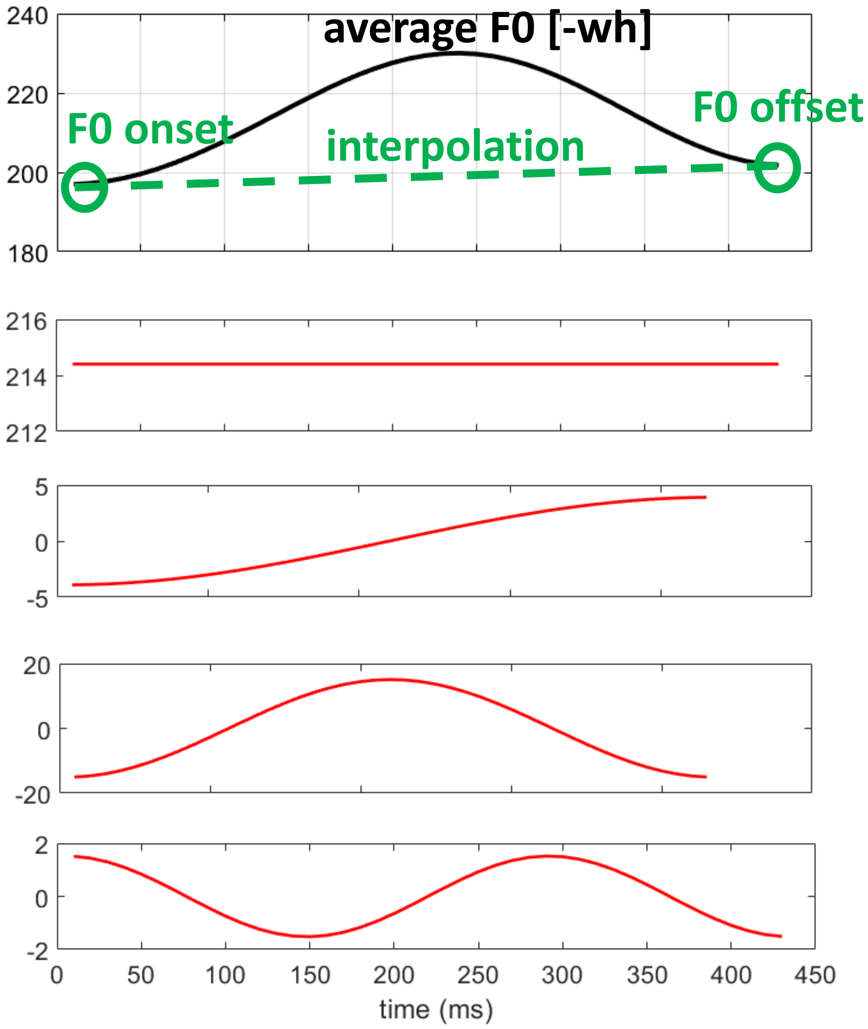
Fit between real and simulated F0 using iDCT



- Simulations from 4 DCT components explain > 90% of variance for all 9 speakers



Step 2: Micro-prosody of targetlessness



Simulate F0 trajectories from DCT components:

Interpolation trajectory	$\boldsymbol{y(k)} \sim \mathrm{N}(\boldsymbol{\mu(k)}, \boldsymbol{\sigma(k)})$
Target present [-Wh]	$\boldsymbol{y(k)} \sim \mathrm{N}(\boldsymbol{\mu(k)}, \boldsymbol{\sigma(k)})$

$$x(n) = \sum_{n=1}^{L} w(k)y(k)\cos\left(\frac{\pi(2n-1)(k-1)}{2L}\right)$$

$$n = 1,2,\ldots L$$

Where *L* is the number of data samples and *x(n)* the trajectory to be simulated and:

$$w(k) = \begin{cases} \dfrac{1}{\sqrt{L}} & k = 1 \\ \sqrt{\dfrac{2}{L}} & 2 \leq k \leq L \end{cases}$$

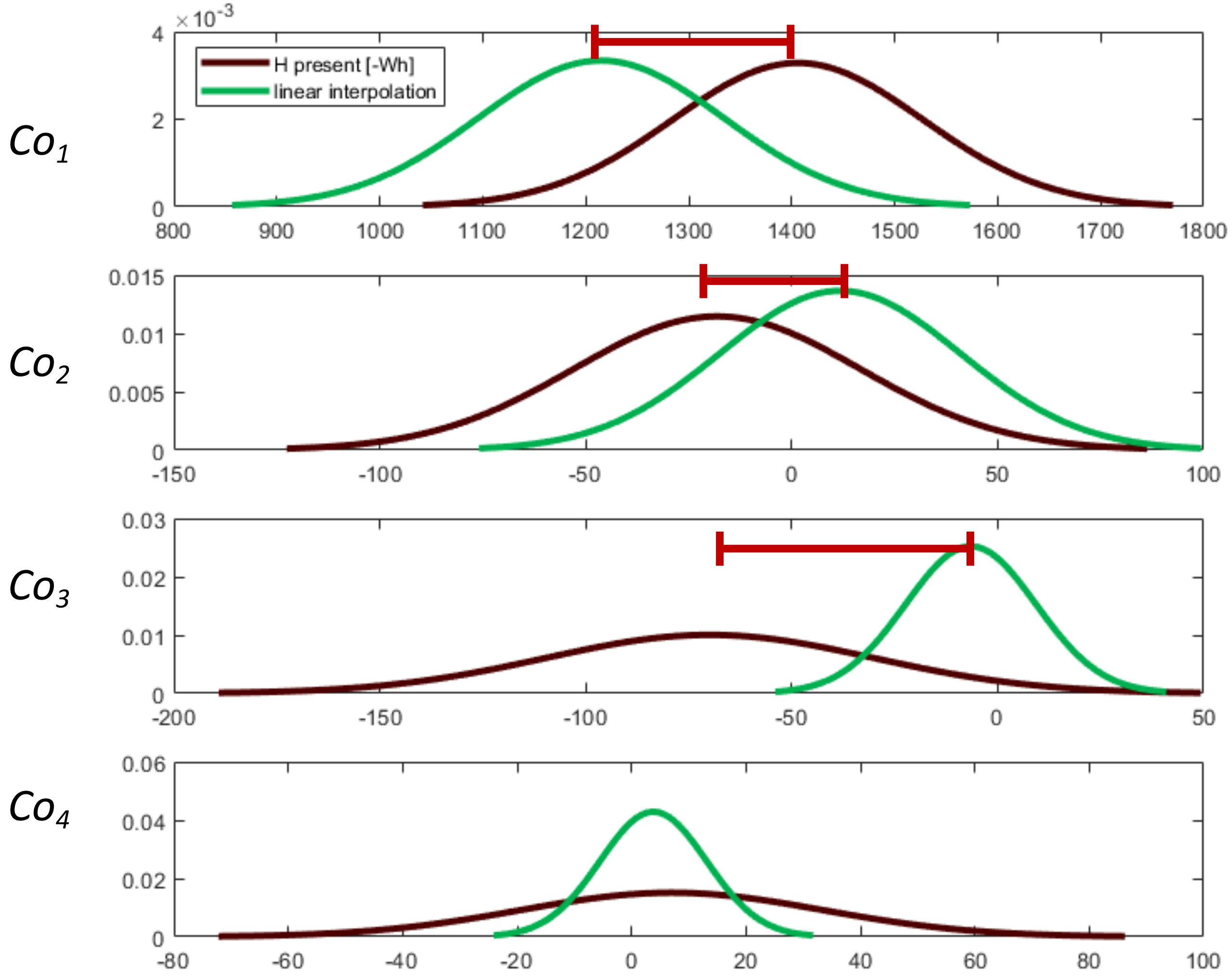
Step 3: Bayesian classifier

- Training data
  - ❖ Word3 Word4 with H accent [-Wh]
  - ❖ **Linear interpolation**
- Test data
  - ❖ Word3 and Word4 in [+Wh] context

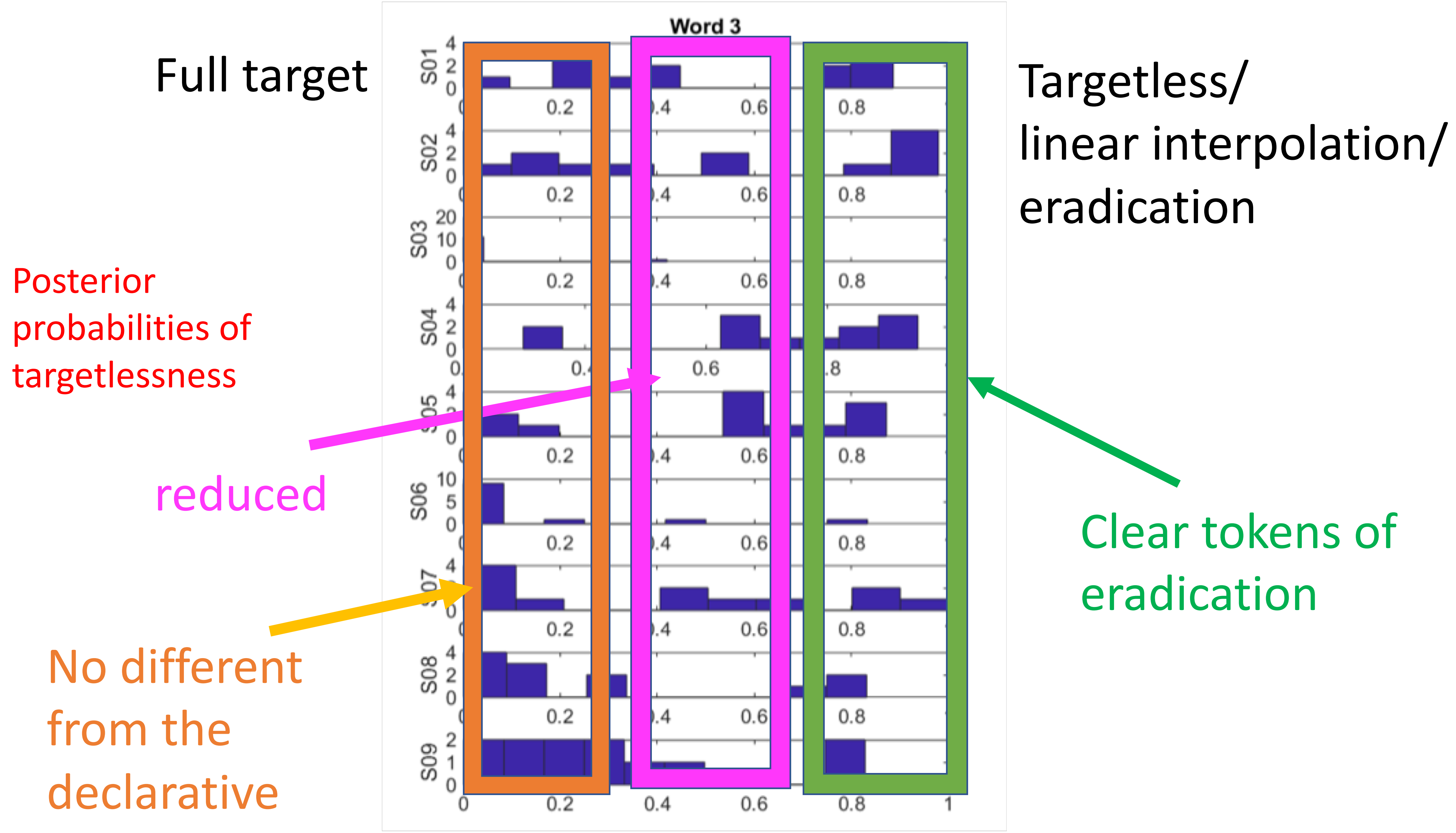
$$p(T \mid Co_i, \ldots, Co_n) = \frac{p(T) \times \Pi_{i=1}^{n} p(Co_i \mid T)}{\Pi_{i=1}^{n} p(Co_i)}$$

where Co_i is the ith DCT coefficient

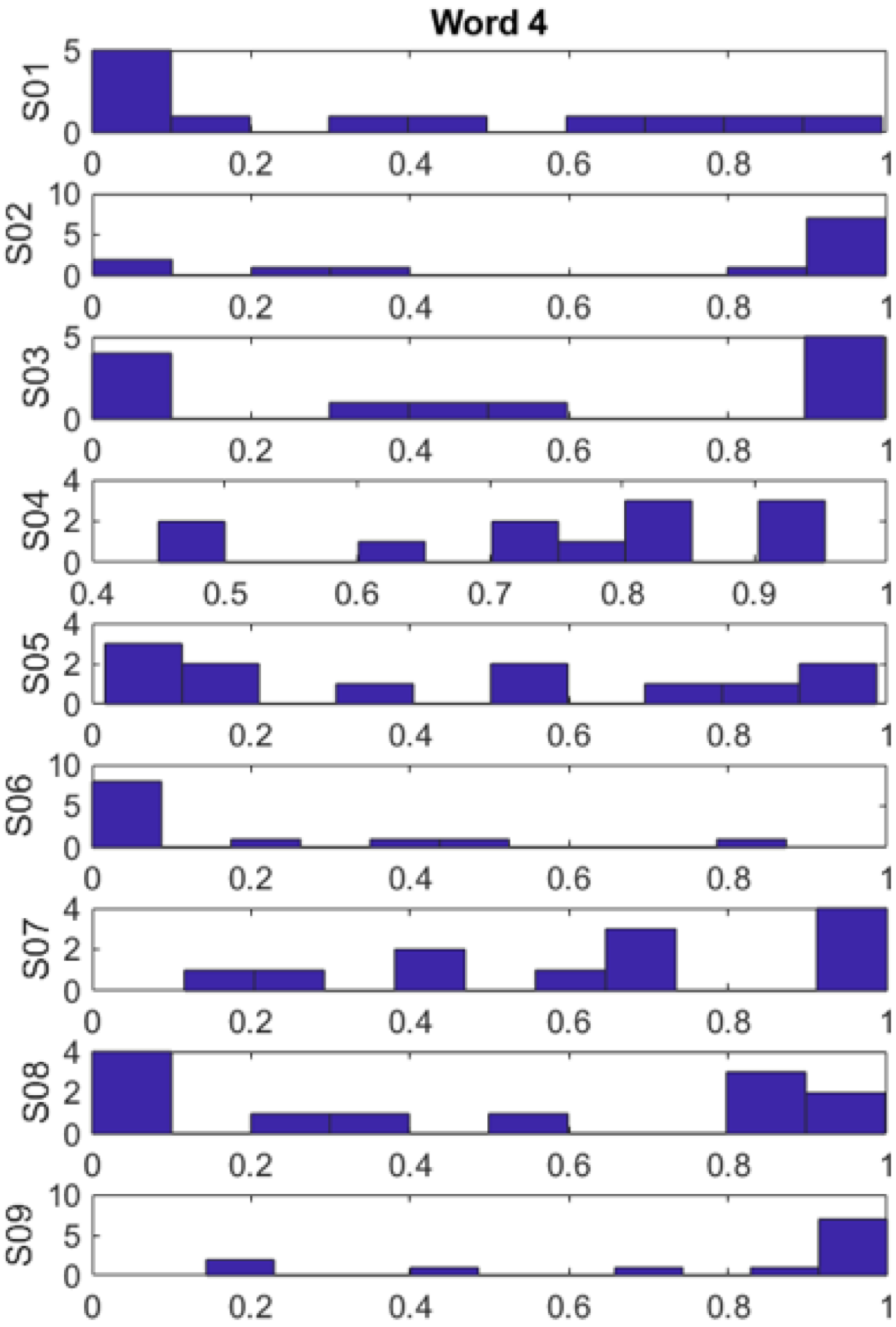
Parameters (4 DCT Coefficients)



Results


Word 3
Full target
Targetless/
linear interpolation/
eradication
Posterior
probabilities of
targetlessness
reduced
Clear tokens of
eradication
No different
from the
declarative
S01
S02
S03
S04
S05
S06
S07
S08
S09
0
0.2
0.4
0.6
0.8
1

Full target



linear interpolation

Discussion

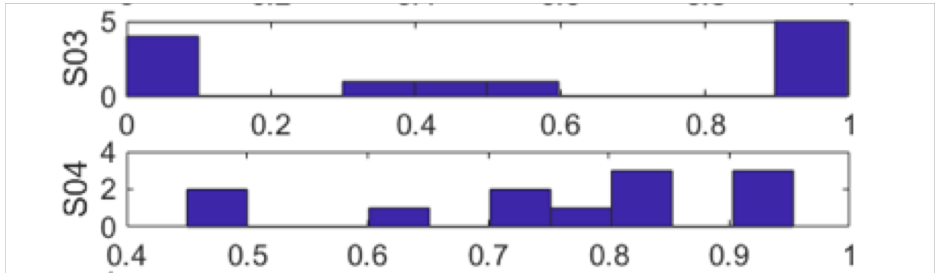
- Some (most) speakers show some tokens that are best characterized as eradicated.
- However, no speakers consistently show eradication.
- Deguchi & Kitagawa’s (2002) observation was correct at some level of analyses, but the current results pose an interesting challenge to Richard’s (2010) theory.

Implications for Richard’s (2010) theory

- If eradication is what allows Japanese *wh*-elements to stay in-situ, how come those tokens without eradication show no wh-movement?
- How come Speaker 6, who almost always showed high probability of full target, does not move *wh*-elements?
- If prosody is a driving force for overt wh-movement, then it must involve some kind of abstraction. But what exactly is that abstraction?

Importance of a token-by-token analysis

- Most previous studies analyze averaged contours, but analyzing only averaged contours can be misleading.
- Take the case of Word_4 for Speakers 3 and 4, for example. Speaker 4 shows reduction for all tokens; Speaker 3 on the other hand shows a bimodal distribution of full targets and eradication.



- If we were to be only looking at averages, we would have erroneously concluded that both speakers show reduction. This highlights the importance of analyzing each token separately.

Comparison with other approaches

- Maekawa (1994) addressed the question of whether “deaccented” phrases and unaccented phrases are different or not, by fitting linear regression lines (see also Pierrehumbert & Beckman 1988).

- The regression lines are different between deaccented phrases and unaccented phrases.

- One distinct advantage of our approach is that it does not have to assume linearity, as the first step of our analysis decomposes trajectories into a sum of cosine waves.

Overall

- Token-by-token analysis offers great promise for the study of intonational variation.

- We look forward to other researchers trying out our computational toolkit.

Thank you