# Delayed-fall of pitch accents in Japanese Infant-Directed and Adult-Directed Speech

Mafuyu Kitahara[1], Ken'ya Nishikawa[2], Yosuke Igarashi[3], Takahito Shinya[4], and Reiko Mazuka[5]

[1]Waseda Univ, & RIKEN BSI, [2]Keio Univ & RIKEN BSI, [3]Hiroshima Univ & RIKEN BSI, [4]Otsuma College & RIKEN BSI, [5]RIKEN BSI & Duke Univ

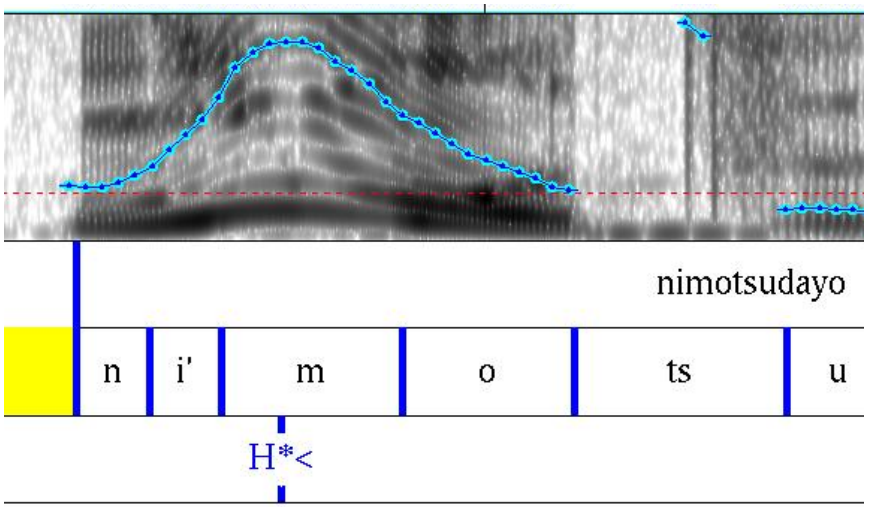
ICPP
2011-12-12
Kyoto

# Introduction

- Topic:
  - Pitch accents in Infant/Adult-Directed Speech (IDS/ADS)
  - Part of a larger research project at RIKEN on language acquisition: brain imaging, eye-tracking, longitudinal study...
- Structure:
  - Background and issues
  - Methods
  - Results [1] (2008, 2010)
  - Reanalysis
  - Results [2]
  - Discussion
  - Conclusion

# Background and issues (1): IDS and Late-fall

- IDS = higher mean pitch, wider pitch range: Ferguson (1964), Fernald&Simon(1984), Fernald&Kuhl(1987)
  - Common in English, Chinese, German, Thai, Japanese: Grieser&Kuhl(1988), Fernald et al.(1989)
  - Chinese tone distinctions are preserved: Liu et al.(2007)
- "Late-fall": pitch fall comes later than phonologically specified position: Sugito (1982), Ishihara 2006)



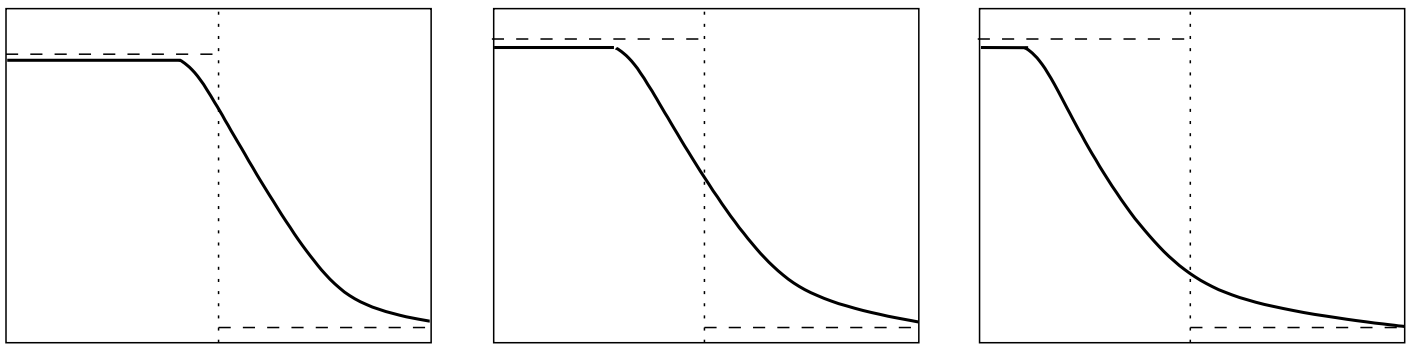

  - Frequent in female speech: Hasegawa& Hata (1988)
  - Synthesized late-fall speech gives female-impression: Hasegawa&Hata(1995)

# Background and issues (2): Previous studies

- Ishihara (2006)
  - Initial-accented words with CVCV-, CVN-, CVV- structures
  - Regularly "late" H* alignment at CVCV- and CVV-. V in CVN is longer.
  - Speech rate nor speaking style does not alter the alignment pattern.
- Cho (2010)
  - Non-initial accent, extensive analysis of rise with varied speech rate.
  - Against Ishihara (2006), target duration between tones gives a better fit.

# Background and issues (3): Weighted Constraints

- Carryover tonal coarticulation is predominant: Flemming (2008)



(a) Carryover (b) Split (c) Anticipatory

  - Realize tone in rhyme, Priority on H over L, Articulatory constraints on transition

- Cho (2010): Target Duration between L-H, Align(H), Delay(L)...from the onset

# Methods (1): R-JMICC (Mazuka et al. 2006)

- 22 mothers in Tokyo-metropolitan-area (Age: 25-43)
- IDS: Play with and/or showing picture books to infant (18-24mo), 3925 AP (accentual phrase)
- ADS: Free conversation with experimenter, 3679 AP
- Recording: 44.1kHz, 16bit DAT
- Morphological, segmental, and prosodic info added manually

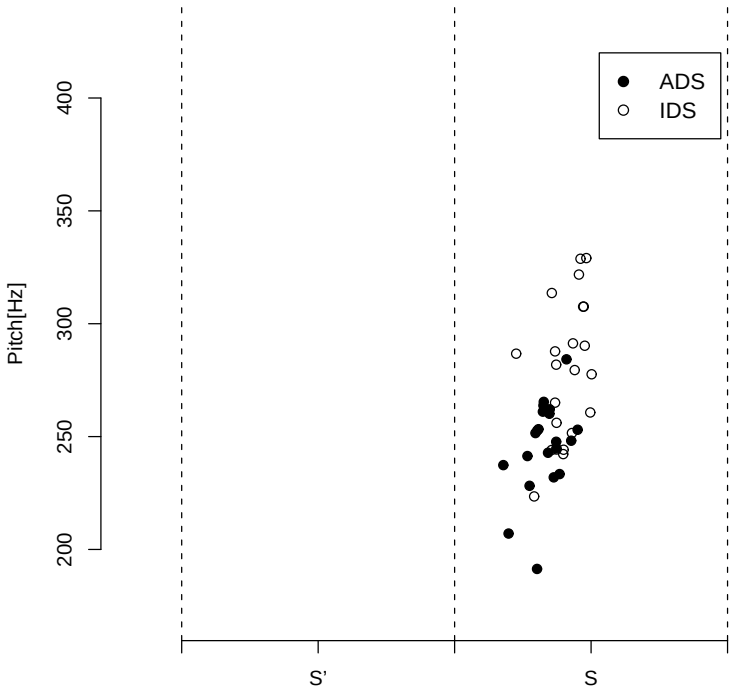
# Methods (2): Word-reading task

- Same 22 speakers, same setting
  - Divided into 2 groups, initial vs. penultimate
  - IDS: carrier sentence "____ dayo", 516 AP, read each word with a picture addressing to her own child
  - ADS: carrier sentence "____ desu", 572 AP, read the list addressing to the experimenter
  - Measurement: peak (max pitch around S') alignment and scale, range (max-min)
- Words

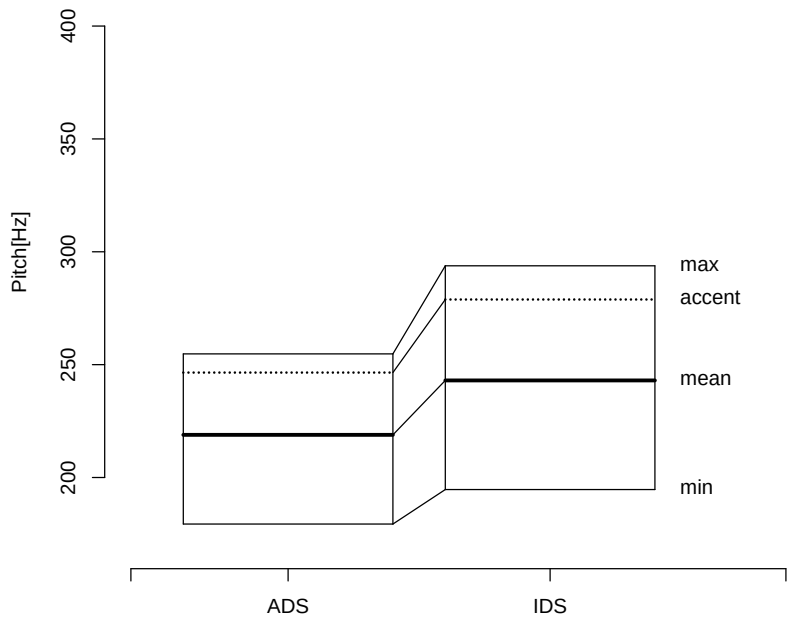
| Initial | | Penultimate | |
|---|---|---|---|
| doresu | banana | amagu | rokuji |
| pazuru | miruku | naname | hitori |
| megane | kimuchi | otsuyu | uchiwa |
| namida | rizumu | amido | tamago |
| midori | gorira | furoya | yumiya |
| burashi | nimotsu | ichibu | kagami |

# Results (1): Corpus

Max pitch in time-normalized syllable



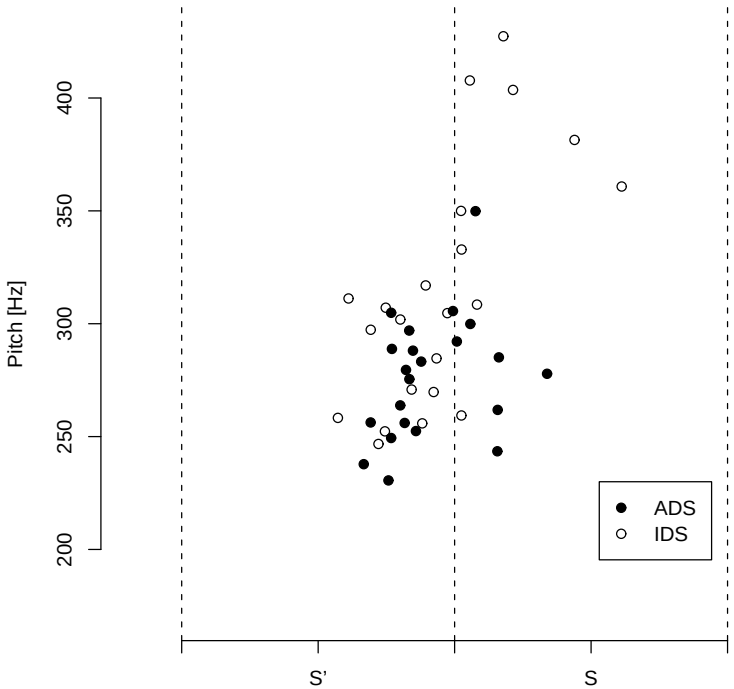

Pitch range and mean



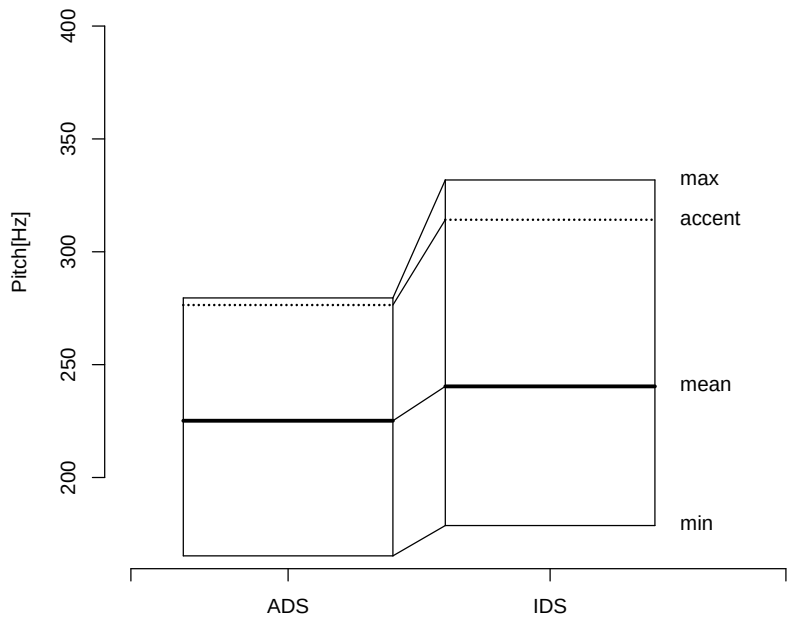

- Scale ($t(21) = 4.79, p < 0.001$), Alignment ($t(21) = 4.63, p < 0.001$)
- Max ($t(21) = 5.69, p < 0.001$), Min ($t(21) = 3.85, p < 0.001$), Mean ($t(21) = 4.47, p < 0.001$)

# Results (2): Word-task

Max pitch in time-normalized syllable




Pitch range and mean

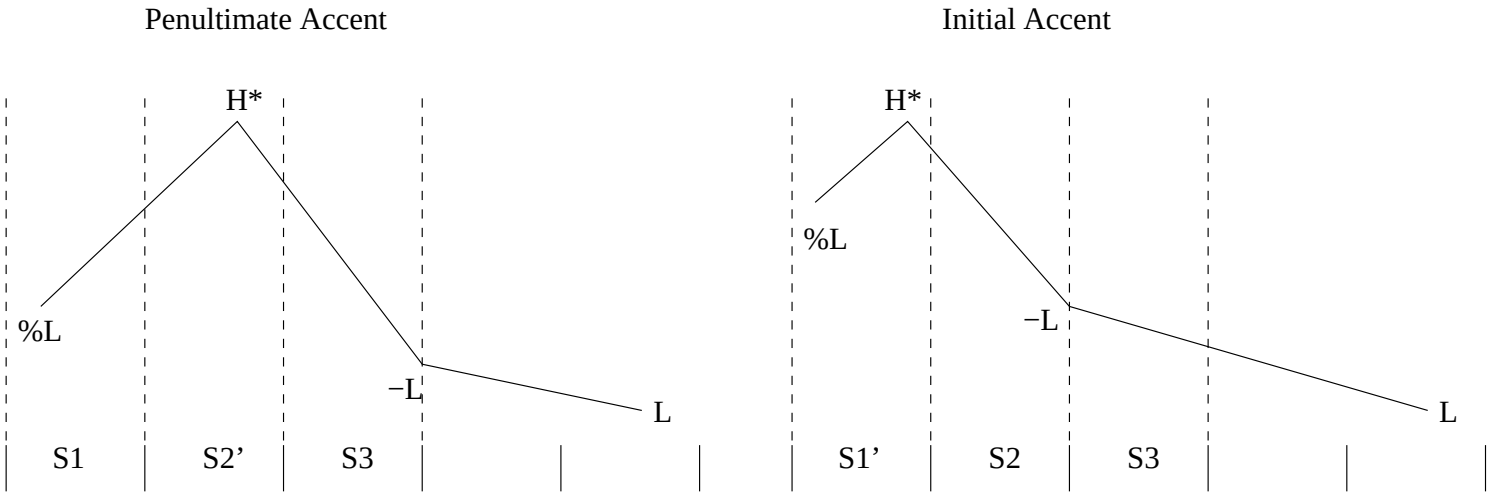



- Scale ($t(21) = 3.36, p < 0.01$), Alignment ($t(21) = 1.07, p = 0.295, NS$)
- Max ($t(21) = 3.76, p < 0.002$), Min ($t(21) = 2.65, p < 0.02$), Mean ($t(21) = 3.29, p < 0.01$)

# Discussion

- Peak is higher and later in IDS
- Higher, THUS later?
  - Scale-alignment interaction
    - Word-task IDS ($r = 0.66, p < 0.001$)
    - Corpus IDS ($r = 0.41, p = 0.055, NS$)
- Why late fall? $\rightarrow$ needs more detailed analysis
  - Word-task data reanalyzed
  - Not just peak but the whole contour
  - Shape of the contour in rise and fall: elbow analysis (Cho 2010)
  - Differences due to accent location

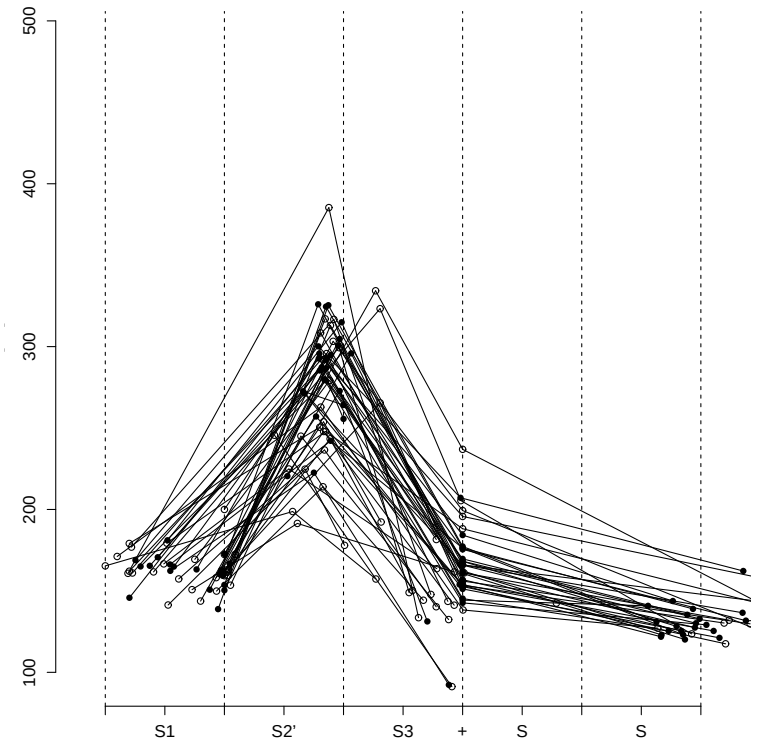
# Reanalysis



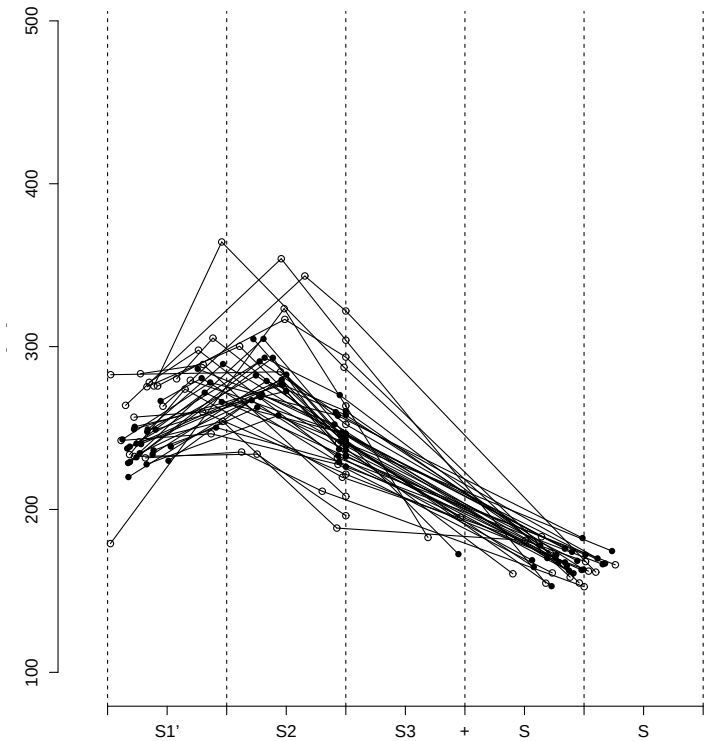

- %L: F0 minima before the accent (even in Initial Accent)
- H*: F0 maxima in or near the accented syllable (Sn')
- -L: F0 minima after the accent and before the end of Sn'+1
- L: F0 minima after -L (if any)
  - Issues
    - F0 perturbations around obstruents
    - Creaky voice in -L and L region

# Results (1): Raw data sample IDS(open) ADS(filled)

Penultimate

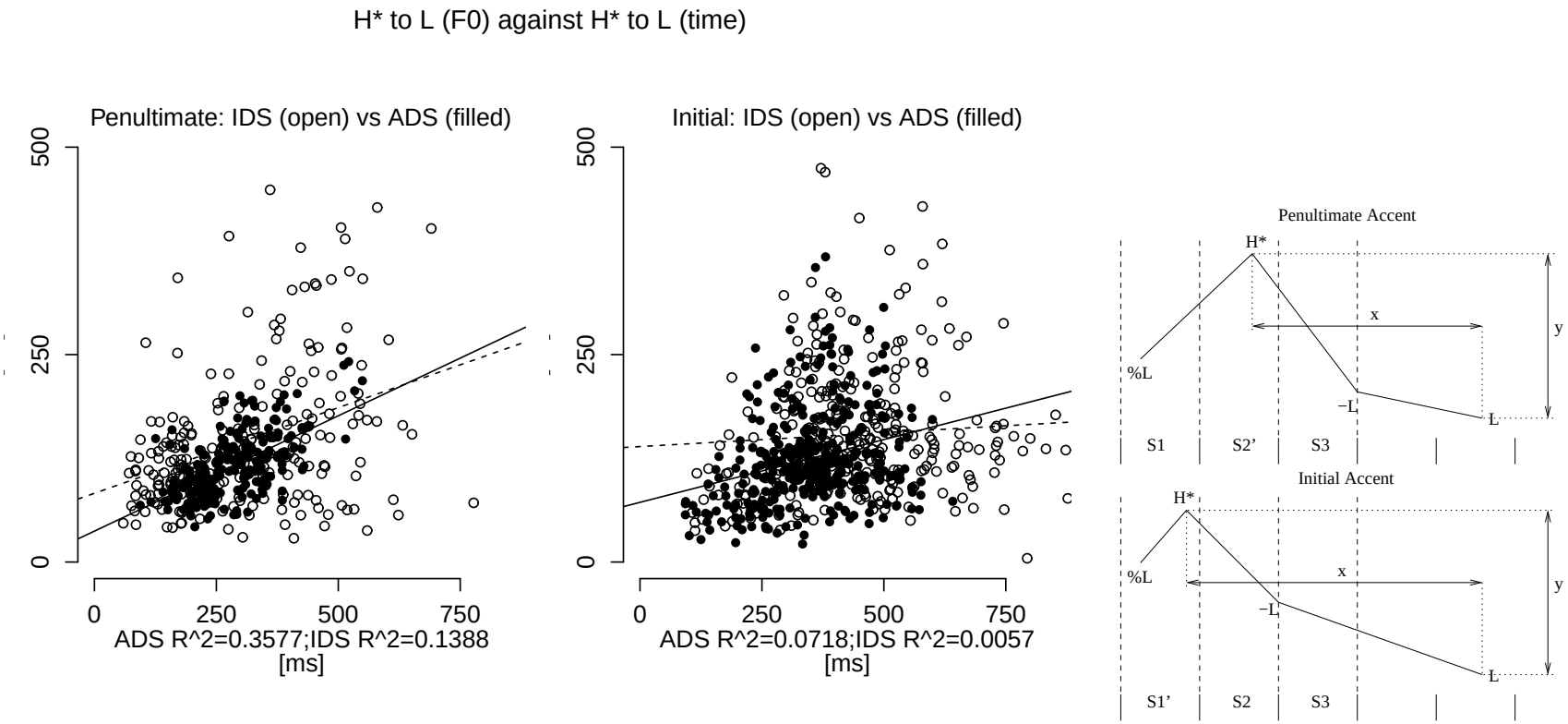



Initial




- Either -L or L unavailable in 294 (26.9%) out of 1094 tokens

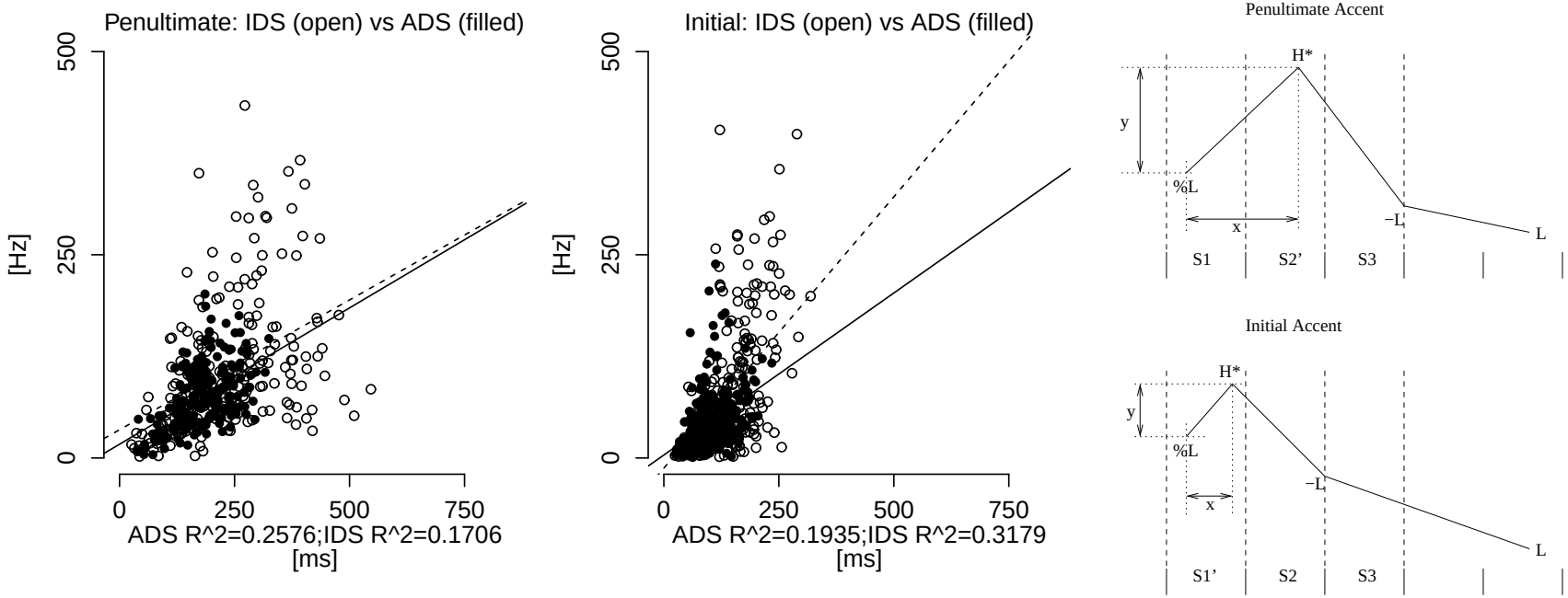
# Results (2): Overview of the fall




- "L" is the last tonal mark (either -L or L)
- IDS tend to have a larger accentual fall (up to 500ms) in both
- No correlation in initial $\rightarrow$ no constant fall to the end

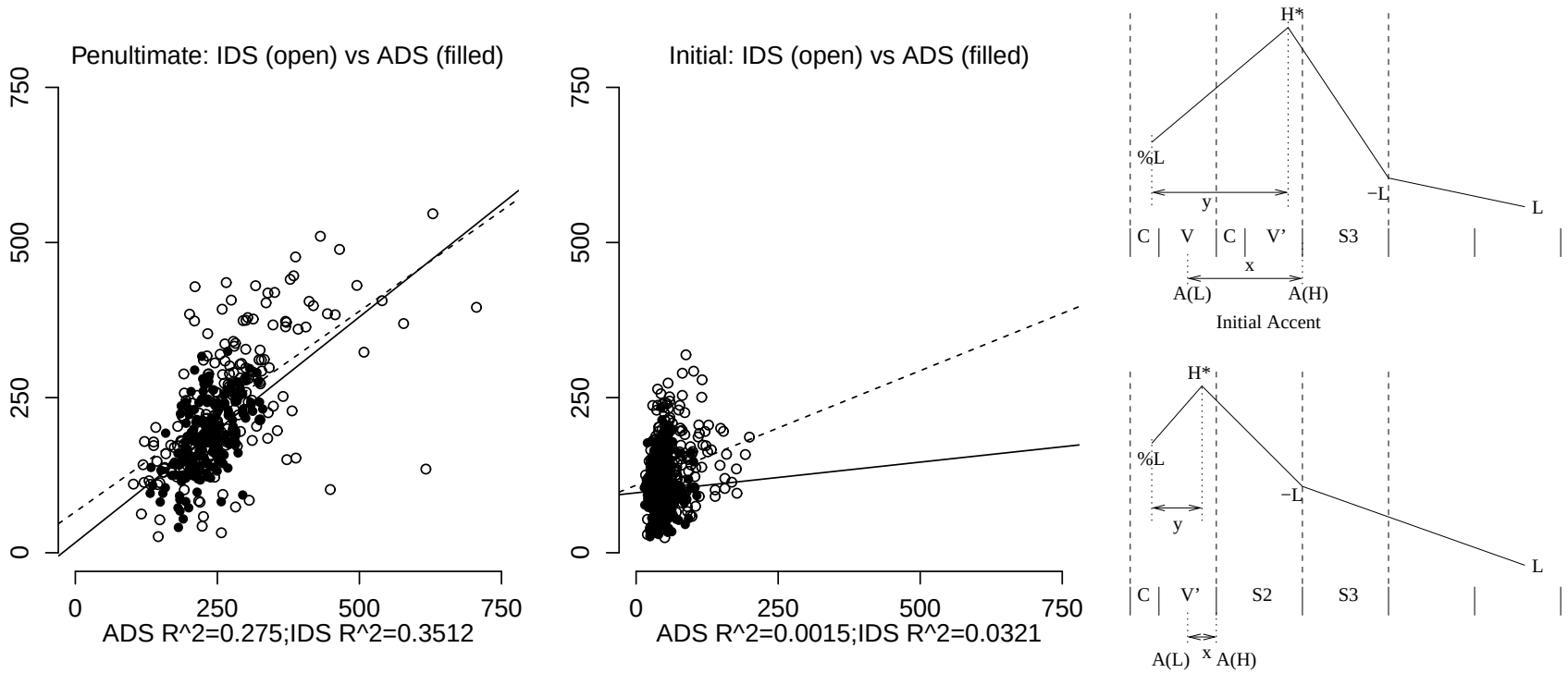
# Results (3): Overview of the rise

%L to H* (F0) against %L to H* (time)




- Longer rise → higher peak (IDS+initial acc in particular)
- Still, not too high even you have enough time (IDS+penult acc)
- **Conjecture**: H* target is set higher in IDS at the expense of articulatory markedness

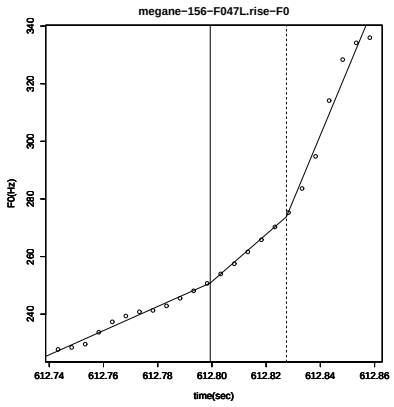
# Results (4): Alignment



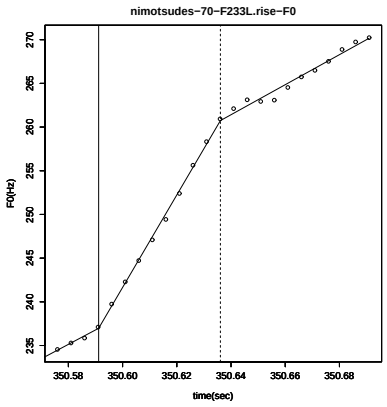

- Distance between anchors relatively stable in ADS and initial
- Initial = more horizontal → Target Duration as minimum requirement
- Penultimate = more slope → bigger weight on Align (less time pressure)

# Results (5): Elbow Analysis

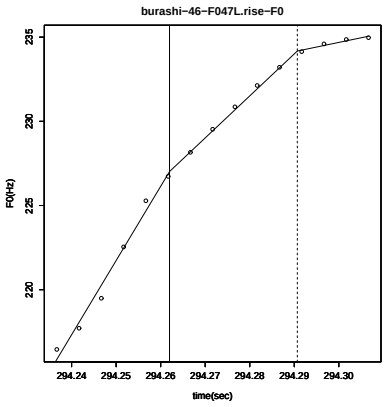
Scoop



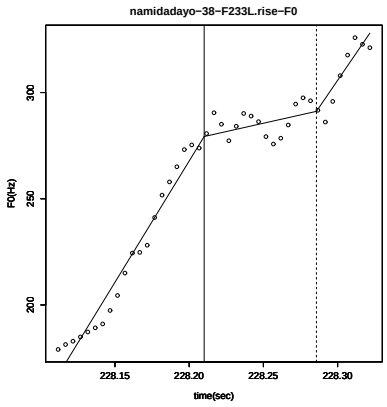

Sigmoid




Dome




Other

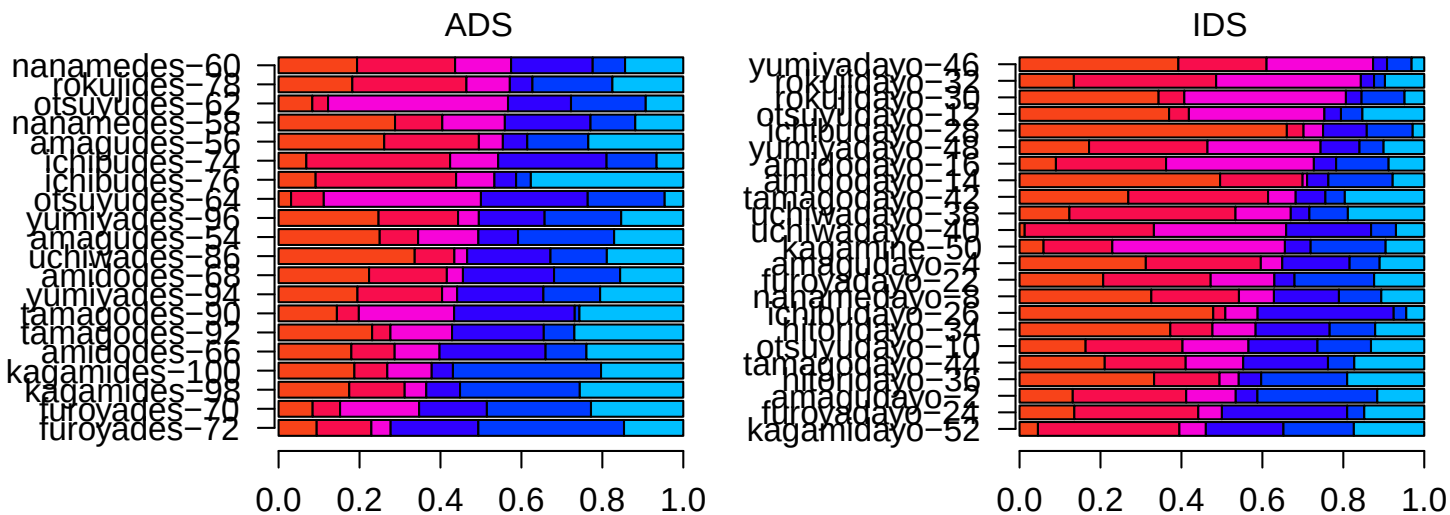



- 3-line fitting algorithm between %L to H* and H* to -L (no L)

| | Scoop | Sigmoid | Dome | Other |
|---|---|---|---|---|
| ADS-Rise | 14 | 218 | 109 | 41 |
| ADS-Fall | 26 | 196 | 148 | 113 |
| IDS-Rise | 28 | 228 | 76 | 43 |
| IDS-Fall | 18 | 167 | 143 | 79 |

# Results (6): Rise-Fall Elbow Analysis

Relative location of elbows and peaks within %L to -L region
— sample data from one speaker, penultimate accent —

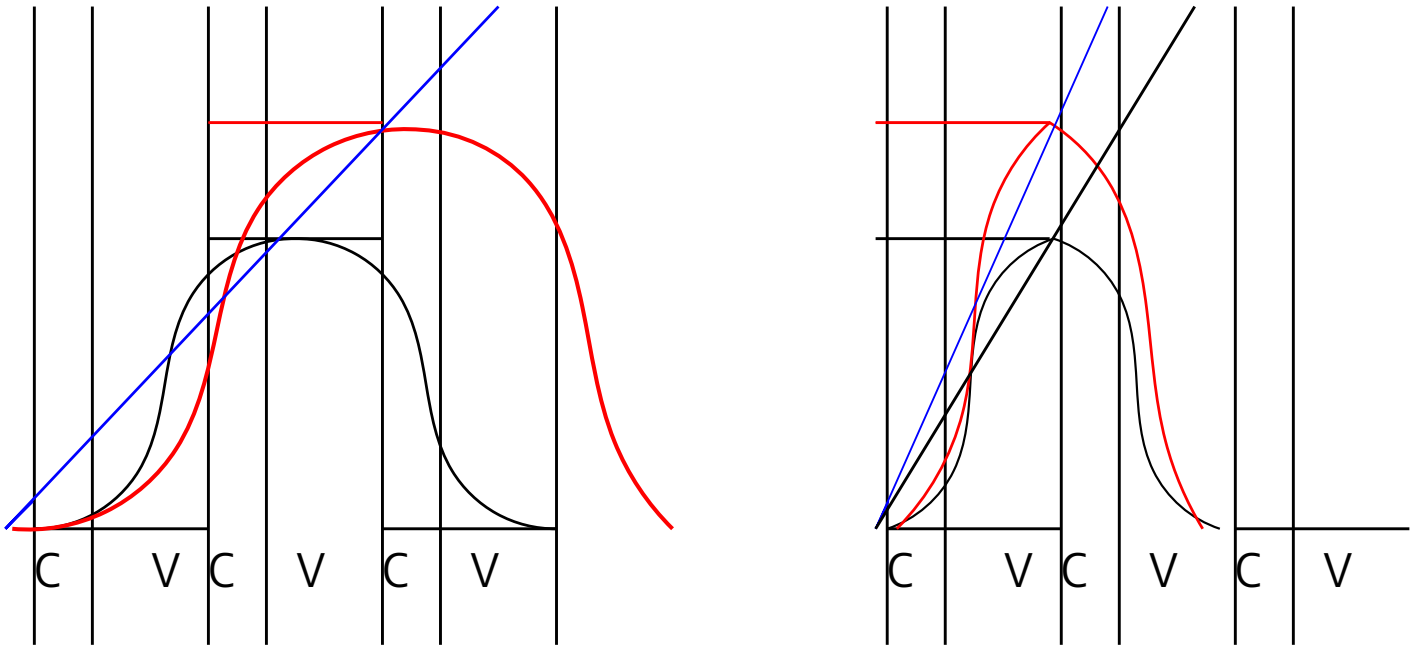



- Not just the peak but the whole contour shifted rightward
  - H* relative location in IDS vs. ADS: t = 5.2636, df = 827.021, p-value = 1.801e-07
  - 1st elbow in rise in IDS vs. ADS: t = 3.2433, df = 857.696, p-value = 0.001227
  - 1st elbow in fall in IDS vs. ADS: t = 4.0846, df = 900.663, p-value = 4.808e-05

# Discussion (1):

- Findings (2008):
  - Expansion of pitch range and mean-pitch rise in IDS
  - More expansion in the upper range than mean-pitch
  - Mixed results for alignment in word task
- Findings (2011):
  - Results 2: Higher and Later trend in both IDS and ADS,
  - Results 3: Even higher for IDS in initial accent
  - Results 4: Target duration effect between tones in initial accent
  - Results 5,6: Not just peak but the entire contour shifts later.

# Discussion (2):




- Assigning heavy weight on Target duration below some threshold (i.e., against tonal crowding). Otherwise articulatory markedness on rapid pitch movement.
- Needs more work for the simulation...

# References


Cho, H. (2010) *A Weighted-Constraint Model of F0 Movements* PhD thesis, MIT.

Ferguson, C. A. (1964) "Baby talk in six languages", American Anthropologist, **66**, pp. 103–114.

Fernald, A. and T. Simon (1984) "Expanded intonation contours in mothers' speech to newborns", Developmental Psychology, **20**, pp. 104–113.

Fernald, A. and P. Kuhl (1987) "Acoustic determinants of infant preference for motherese speech", Infant Behavior and Development, **10**, 3, pp. 279–293.

Fernald, A., T. Taeschner, J. Dunn, M. Papousek, B. deBoysson-Bardies and I. Fukui (1989) "A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants", Journal of Child Language, **16**, pp. 477–501.

Flemming, E. (2008) "The Grammar of Coarticulation", To appear in M. Embarki and C. Dodane (eds.) La Coarticulation: Indices, Direction et Representation.

Grieser, D. L. and P. Kuhl (1988) "Maternal speech to infants in a tonal language: Support for universal prosodic feature in motherese", Developmental Psychology, **24**, pp. 14–20.

Hasegawa, Y. and K. Hata (1988) "Delayed pitch fall in Japanese", JASA Suppl, **1.83**, S29.

Hasegawa, Y. and K. Hata (1995) "The function of f0-peak delay in Japanese", BLS, **21**, pp. 141–151.

Ishihara, T. (2006) *Tonal Alignment in Tokyo Japanese*, PhD thesis, University of Edinburgh.

Liu, H.-M., F.-M. Tsao and P. Kuhl (2007) "Acoustic analysis of lexical tone in mandarin infant-directed speech", Developmental Psychology, **43**, 4, pp. 912–917.

Mazuka, R., Igarashi, Y. and Nishikawa, K. (2006) "Input for learning Japanese: RIKEN Japanese Mother-Infant Conversation Corpus", Technical report, TL2006-16, 11-15.

Sugito, M. (1982) *Nihongo akusento no kenkyuu*, Sanseido, Tokyo.